INTERNATIONAL JOURNAL OF MODERN ENGINEERING RESEARCH (IJMER)

ISSN : 2249-6645

Volume 1 - Issue 2

Web : www.ijmer.com

Email : ijmer.editor@gmail.com

International Journal of Modern Engineering Research (IJMER)

Editorial Board

PHOTO-ELASTIC METHOD OF STRESS ANALYSIS FOR PANEL OF INFILL FRAME SUBJECTED TO RACKING

S.KANAKAMBARA RAO
Associate Professor

Department of Civil Engineering, K L University Vaddeswaram,
Guntur(dist.), AP, India

ABSTRACT: A two dimensional photo-elastic model analysis has been adopted to analyze the masonry infill behavior in an in-filled frame against racking load. The stress distribution is studied for the relative stiffness of the frame and infill, equal to 3.5.The model tested is composite, fabricated using Aluminum for frame and araldite AY103 with hardener HY 951 as the infill material. Observations of the model placed in the photo –elastic bench revealed the visible picture of fringes over the whole area of the infill from which the stress distribution is accurately readable at any point for both the direction and magnitude. The results illustrate that the photo-elastic method can be effectively utilized to study the elastic behavior of in-filled frames.
KEY WORDS: In-filled frames, failure modes, masonry panel, relative stiffness, epoxy resins, photo-elastic bench.

1. Introduction:

Analysis of in-filled frames for stresses has been arrived at by various analytical and experimental methods as documented in the published work. It has been established that the analysis of in-filled frame is a complicated one as it involves structural interaction and stress concentration. The mutual interaction of the frame and in-fill plays an important role in controlling the stiffness and strength of the in-filled frame and the problem is to be examined in terms of their relative properties. The relative property of in-fill and frame is controlled by the parameter λh, smaller the value of λh, stiffer the frame relative to the infill. In the present analysis the value of λh has been chosen to be 3.5.
As both structural interaction and stress concentration are involved, obviously experimental stress analysis was envisaged and photo-elastic method has been employed to study the complete stress distribution in the in-fill.

2. Methodology:

2.1 In-filled frame model:
Epoxy resin was chosen as one model material of the in-fill because of various advantages like ease with which it can be casted, optical sensitivity etc. Araldite AY-103 and harder HY-951 were used to cast an in-fill model of size 58×58×6 mm. For the frame aluminum plate of 6 mm thickness had to be machined to represent structural members to detailing. To relieve the stresses induced during cutting and machining, liquid paraffin bath technique has been adopted.
Calibration of the photo-elastic material i.e. the in-fill was carried out using a tension specimen. Young's modulus of the model material has been arrived at by loading at two points on a small strip of araldite specimen.

2.2 Testing arrangement of lateral load:
The models were made in a duplicate back to back arrangement to simulate rigid foundation for each half, and to allow for ease of testing. The testing arrangement for the model analysis is shown in Fig .1a and Fig. 1b

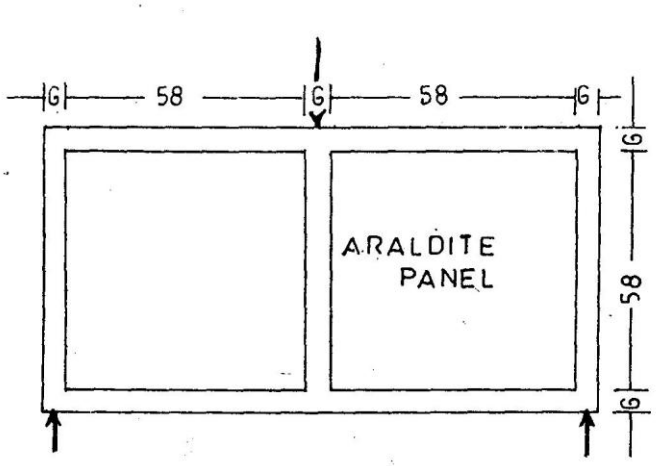


Fig.1a Testing arrangement (Dimension in mm)

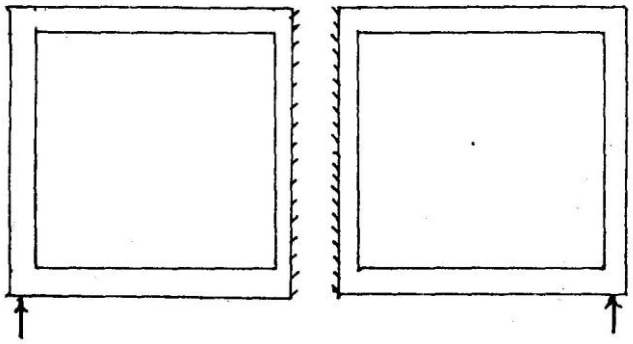


Fig. 1b Equivalent effect of testing arrangement

2.3 The parameter λh for the model:

The behavior of in-filled frames has been shown by Stafford smith to be partly related to the flexural stiffness of the frame relative to the in-plane diagonal of the in-fill defined by the parameter λh where

$$\lambda h = \sqrt[4]{\frac{E_1 t h^3}{4EI}}$$

In which, E_1, t and h are the modulus of elasticity, thickness and height of the in-fill frame respectively and E and I are the modulus of elasticity and second moment of area of the columns respectively. A ratio of E and E_1 considered to be a representative ratio of reinforced concrete frame with a masonry in-fill works out to be 18 for E_1 to be 4200MPa and E to be 75000MPa and for infill thickness of 6mm with a height of 58mm, and for a width of the frame 6mm and thickness of 6mm, λ h will be 3.5.

2.4 Measurements

For the measurements of magnitude and direction of principal stress a plane polarioscope is used with white light and with necessary calculations, various stresses are obtained and are discussed below.

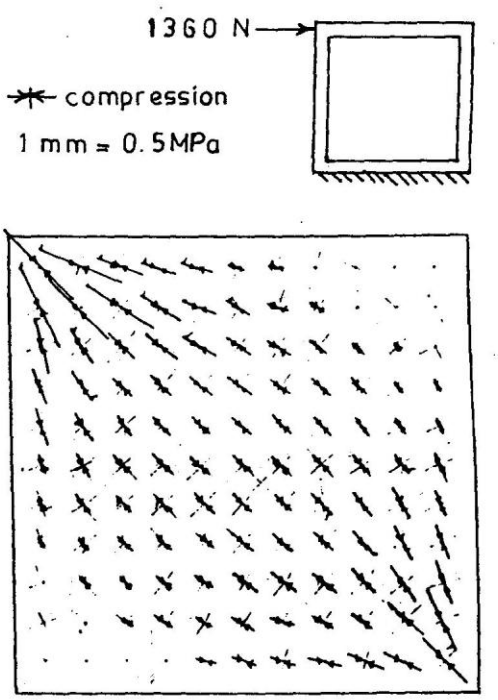


Fig 2 Flow of stresses in the panel of solid in-filled frame.

3. Analysis of result:

3.1 Principal stress distribution
Flow pattern of principal stresses in the infill for solid infill is shown in fig.2. The length of contact will be only over a small length at the loaded corners and as such the compressive stress concentration gets localized. Most of the tensile stresses at the corners of the tensile diagonal of the infill get relieved.

3.2 principal compressive stress contours
To study the gradient of compressive principal stress distribution in the infill panel, compressive stress contours are drawn and presented in fig.3 which shows that the contours are concentrated at the corners of the in-fill along the loaded diagonal and also most of the region of the infill panel is stressed by compressive stresses.

3.3 principal tensile stress contours
The principal tensile stress contours, revealed that the tensile stress contours run along the loaded diagonal and hence is susceptible for initiation of tensile cracking as the masonry is very weak in tension.

3.4 Vertical stress distribution
Fig.4 represents the vertical stress distribution in the infill panel. The stress concentration at the infill corners loaded increase substantially. At the base of the infill, the compressive stress gets concentrated over a small length of contact and similarly at the top edge also. Because of high stress concentration the failure of the infill initiates crushing at the infill corner.

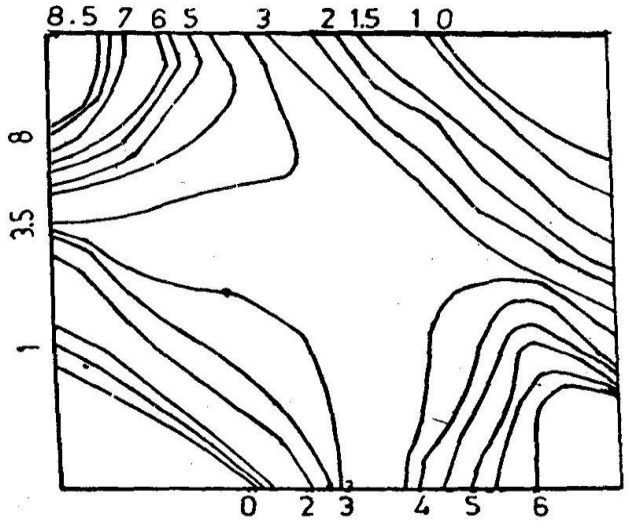


Fig 3 Principal compressive stress contours in the infill (stresses in MPa)

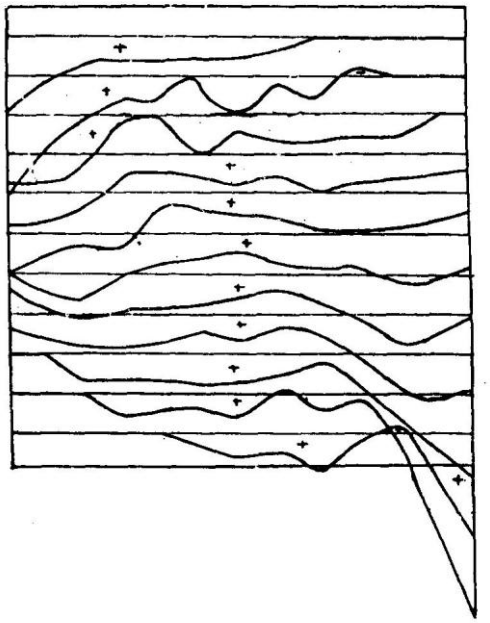


Fig 4 Vertical stress variation across the panel (1mm=0.12MPa)

3.5 Horizontal stress distributions

At the loaded points compressive stress concentration observed is more along the two vertical edges where there is no contact with the frame horizontal stresses are found to be zero.(fig.5)

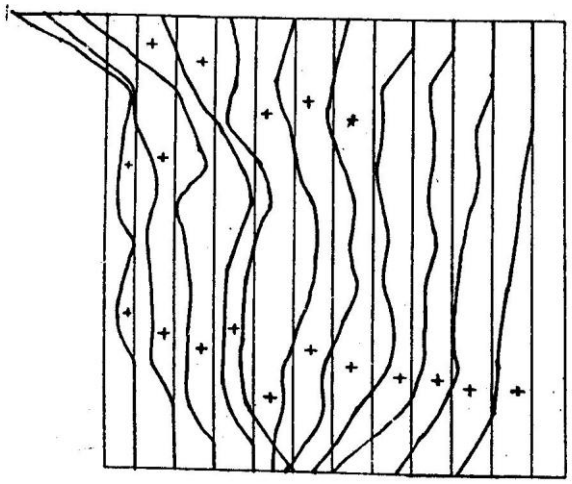


Fig 5 Horizontal stress variation across the panel (1mm=0.12MPa)

3.6 Shear stress distribution

The shear stress distribution shown in fig.6 depicts shear stress concentration at the corners of the support and the load. At top and bottom of the infill it is not uniform but varying in nature, the mid portion of the panel behaves as a rectangular beam. Along the vertical edges of the infill wherever separation takes place shear stress is zero.

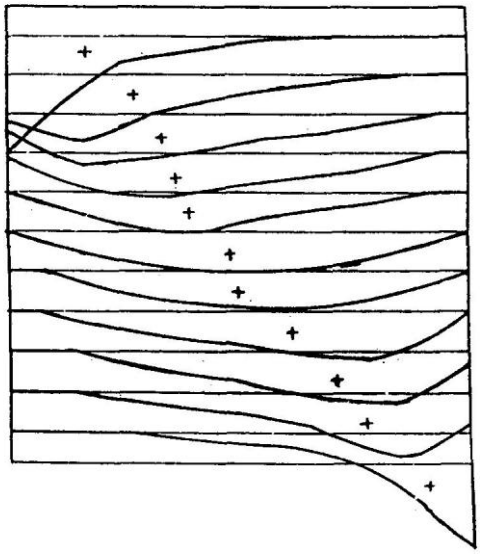


Fig 6 Shear stress variation across the panel (1mm=0.12MPa)

4. Conclusions

1. The photo-elastic method of analysis using transmission polariscope can be effectively used to study the elastic behavior of in-filled frames.

2. Full interaction at the panel frame interface contributes to the composite nature of the structure and the strength and also the establishment of full interaction at the interface.

3. When horizontal load is applied, the frame racks and bears against the infill panels across their loading diagonals, causing them to behave as diagonal compressive struts. As the horizontal loading is increased, the infill is liable to one or more of three modes of failure. The first possibility is a diagonal tension crack along the compression diagonal. This is induced by the outward curving components of the compressive stress trajectories from the infill compression diagonal. The second, a shear failure, follows an approximately diagonal path. The third is a crushing of a corner of the infill, at one end of the diagonal strut, where compressive stress concentration is very high.

5. References:

1.Liauw,T.C stress analysis for panel of in-filled frames building science,1973, 8, 105-112.

2.Jagadish,R studies on the behavior of in-filled frames with openings, ph.D thesis, Indian Institute of Technology,Madras,India,1981.

3.Chandrashekar,B. and Jacob, K.A. photo-elastic analysis of composite action of walls supported on beams, Building and Environment,1976, 11,139-144.

4.Riddington,J.R and STAFFORD SMITH,B. analysis of in-filled frame subjected to racking with design recommendation, structural Engineer, 1977,55,263-268.

5.Hendry.A.W, photo-elastic Analysis, 1966, pergamon, Oxford.

6.Frocht.M.M, photo-elasticity, vol 2, 1957,Wiley,Newyork.

7.Smith.B.S, the composite behavior of in-filled frames, J.struct, Div,ASCE,1966,92,381-403.

8. Luis Decanini, Fabrizio Mollaioli, Andrea Mura, Rodolfo Saragoni, seismic performance of masonry in-filled R/C frames, 13th World Conference on Earthquake Engineering,Vancouver, B.C., Canada, August 1-6, 2004, Paper No. 165.

9. Diptesh Das and C.V.R. Murty, brick masonry in-fills in RC framed buildings: part1 –cost implications, The Indian Concrete Journal, July 2004, 39-44

10. Goutam Mondal and Sudhir K. Jain, M., Lateral Stiffness of Masonry In-filled Reinforced Concrete (RC) Frames with Central Opening, *Earthquake Spectra,* Volume 24, No. 3, pages 701–723, August 2008; © 2008, Earthquake Engineering Research Institute.

11.Ghassan K. Al-Chaar and Gregory E. Lamb, Design of Fiber-Reinforced Polymer Materials for Seismic Rehabilitation of In-filled Concrete Structures, US Army Corps of Engineers, Engineer Research and Development Center, December 2002.

12. Kashif Mahmud1, Md. Rashadul Islam and Md. Al-Amin, Study the Reinforced Concrete Frame with Brick Masonry Infill due to Lateral Loads, International Journal of Civil & Environmental Engineering IJCEE-IJENS Vol: 10 No: 04 36

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue2, pp-242-246 ISSN: 2249-6645

Review and Diagnostics of noise and vibrations in automobiles

Prof. Deulgaonkar V.R[1], Prof.Dr.Kallurkar S.P[2,] Prof.Dr. Mattani A.G [3]

[1](Department of Mechanical Engineering, MMCOE, Pune, University of Pune,)
[2](A.G.Patil Institute of Technology, Solapur, University of Solapur India)
[3](Department of Mechanical Engineering, Govt.COE Amravati, Sant Gadge Baba Amravati University, India)

ABSTRACT

The present work describes various automotive noise & vibration sources and their contribution. Noise and vibration reduction technique is studied through energy flow path. Various international and Indian standards for vehicles consider two types of noise measurement i.e. pass by noise and stationary noise. This paper discusses the appropriateness of SN test for in use vehicle. A methodology for interior noise source identification and its analysis is descried. Two vehicles of same class but of different makes were compared and evaluated for interior noise and vibration levels. The effectiveness of the firewall, silencers and engine mounts are checked and compared. The correlation between pressure and vibration levels of different sources with acoustical and structure transfer path are studied. Basic causes, design guidelines and validation techniques using lab simulation and data acquisition are discussed. Application of damping technology using viscous materials to control noise and vibration in vehicles is described.

Keywords - E*nergy flow path*, Noise, *Stationary Noise*, Vibrations.

INTRODUCTION

Sound is a propagating type of energy traveling through a medium with particular velocity. The unwanted sound is noise. Vibration is the variation or displacement of a body with respect to specific reference position with time when displacement is alternatively greater or smaller than reference. Harshness is defined as vibration perceived actually and audibly produced by interaction of the tyre with road irregularities and vibrations of the structure and components. [1]. A significant part of the world energy consumption is related to transportation. The wide use of automobile vehicles causes detrimental effects on the surrounding environment. The 20-25% of the total greenhouse gas emission in industrialized countries is generated by transportation [2]. The transportation noise is one of the major sources of noise exposure in residential areas and causes substantial annoyance during night. Considering this, many countries have enocked legislation limiting the noise levels in residential areas. Various international and Indian Standards for vehicles consider two types of noise measurement viz. passby noise (PBN) and stationary Noise (SN).The oil thickness plays a major role in determining the engine's vibration characteristics [3-4].The acceptance criterion of any vehicle in terms of user comfort depends on the vehicle interior noise and vibration characteristics. The levels of sound energy and structural excitation inside the vehicle compartment measures the amount of annoyance in terms of quality and comfort. For vehicle interior noise identification and treatment, quantification of noise sources by determining the sound power contribution from each vehicle component, acoustic leakages inside the vehicle body panel, vibrations during gear shifting at lever and steering wheel vibrations needs to be identified, because interior noise in a vehicle has a major impact on customers perception of operation, performance and quality [5]. In the highly competitive global automotive market the need to develop high quality products and achieve product excellence in all areas to obtain market leadership is critical.

1. Sources of Noise & Vibrations in Automobile

Interior noise in any vehicle reduces the users ride comfort. For today's compact era the trend towards compact power units is substantially increased resulting in vehicles running at higher level of noise and vibrations.

1.1 Engine**:**

Vibrations in engine are generated due to the reciprocating mechanism used for converting the energy into rotary motion. The forces producing the engine vibrations are: Combustion, Reciprocating and Rotational Forces. A downward force is generated during combustion stroke on the piston which due to geometrical construction of connecting rod and crankshaft generates a torque around crankshaft axis. Torsional vibrations are generated due to

the torque variations. A multi-cylinder engine can be compared with a system of masses rotating on a single crankshaft in single and different planes. The primary & secondary forces as well as couples generate vibrations due to reciprocating unbalance. Significant inertia effects are generated due to small unbalance of rotating masses in high speed engines. Rotating unbalance generates unacceptable levels of vibrations and stresses in individual and supporting structures

1.2 Noise Sources**:**

Various noise sources in an automobile are induction noise, exhaust noise, noise from accessories, and noise radiation from engine sources. Induction noise is due to opening and closing of valves. In cylinder on opening the valve, the inlet air column is set into oscillation due to intense pressure thump. Closing of the inlet valve produces forced undamped vibrations. Exhaust noise exists when exhaust valve opens and releases gas into exhaust system. Various accessories used generate unwanted sound. In this category engine fan is the main source of noise. It is used in addition to radiator for cooling, and operated by air during ride. Pressure fluctuations result in generation of noise. Transient vibrations are induced by periodic and aperiodic distortion of engine due to combustion processes. Figure 2 shows Propagation of tyre noise of an automobile at frequency of 600 Hz. Alternating inertia loads and mechanical impacts of the engine mechanism produces noise. Often it is very difficult to sort out which force is the cause of excitation of engine structure. Table 1 shows the percentage contribution of sources of the total noise.

Table 1. Percent contribution of sources to total noise

Sr.No	Source	% Contribution
i.	Engine	22 to 30
ii.	Exhaust system	25 to 35
iii.	Intake system	05 to 15
iv.	Fan and cooling system	07 to 15
v.	Transmission	12 to 15
vi.	Tyres	09 to 15

1.3 Driveline Sources

Noise and vibration in driveline are a consequence power transmission from engine to wheels. Mechanical layout of front wheel drive and propeller shaft of rear wheel drive is the sources of noise and vibration in respective automobiles. The various sources are transmission gear noise, drive and propeller shaft, axle noise, tyre noise, aerodynamic noise, wind noise and interior noise. Generation of noise & vibrations from gears results due to improper bending dynamics of gear tooth and both torsional and bending characteristics of shafts. Propeller shaft generates excitation at elemental speeds. Due to large coupling angles, universal joints generate excitation. Also most of modern vehicles induct constant velocity coupling at the centre of two piece propeller shaft results into noise. Axle noise is due to response of rear axle to vibration generated by meshing action of the axle gear set. The so generated noise is annoying even at squat levels in passenger compartment of the vehicle. Tyre noise is due to tribology between tyre and road. Mechanics of tyre noise generation may be combination of squash vibration (primary noise source) exists due to rough road surface, tread squirm results lateral vibrations and generates noise spectra. Slick/aerodynamic noise is generated by chaotic flow of air around the tyre contributes to the tyre noise. Tyre is excited by several means, which include non-uniform wear, radial or lateral run-out, road roughness, road surface irregularities, road surface discontinuities that induces impacts, bumps etc, which contribute to noise and vibration of automobiles. Wind noise is superficial and is experienced at the interior of vehicle. Flow of air over the exterior of vehicle and the flow of air into and out of the cabin arising from imperfect sealing of door frames and glasses are the causes of wind noise generation. Ample number window and door seals ensure successful wind noise control. Fig. 1 shows various noise/vibration sources.

Interior noise is a prominent acceptance criterion of any vehicle in terms of comfort at the interior part. To identify interior sources of noise and diagnose them, the noise sources are quantified by determining sound power contribution from each vehicle component, panel acoustic leakages, panel vibrations gear shifting, and steering wheel vibrations. Engine being the main source of noise, the noise from the engine is transmitted in two ways viz. direct infiltration & structural vibrations. Improper sealing, holes in lower dashboards, complicated geometry, worn out engine mounts leads noise from engine to reach directly into the cabin. Structural vibrations are due to rings in exhaust systems. These vibrations are transferred from engine to body through drive shafts supported on bearings, rear axle etc. Table 2 depicts engine noise, vibration phenomenon and sources. Table 3 presents the permissible noise levels according to EU directive 96/20 EC

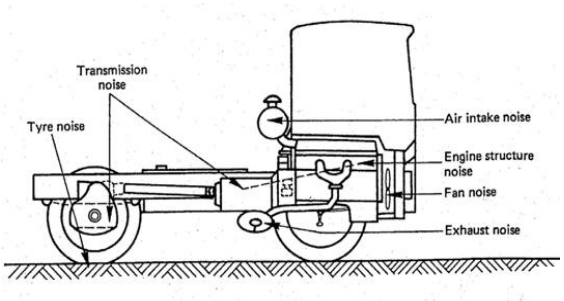


Figure 1 Various vehicle noise /vibration sources

www.ijmer.com Vol.1, Issue2, pp-242-246 ISSN: 2249-6645

Figure2. Propagation of tyre noise of an automobile at frequency of 600 Hz

2. Noise control/reduction techniques

The following techniques are proposed as under

2.1 Energy Flow Path technique

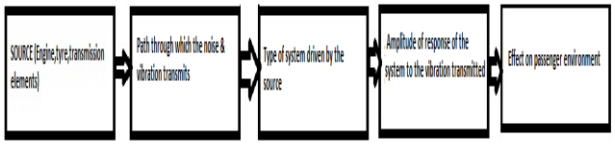


Figure3. Energy flow path diagram showing the propagation of energy (noise/vibration) from source to rider/passenger

Fig.3 shown above depicts the flow of noise and vibration through various stages from source to destination. The transmission path properties are determined by the vibration modes of the structure. Outer surface properties also influence the sound propagation. The ways in which the final engine noise radiation may be influenced or controlled are reduction at the source of combustion forces and mechanical forces, reduction of vibration transmission between the source and the outer surface, reduction of the sound radiation of the outer surface, control or reduction of combustion pressures, reduction of piston slap by redesign of the piston and cylinder or by oil film injection, gear and bearing noise are reduced by improved design e.g. gear tooth profiles and bearing clearances, more advanced redesigns can be made involving extensive simulation the dynamics using finite element modeling. Fig. 4 shows the noise and vibration reduction technique.

Table 2 Engine noise, vibration phenomenon and sources

Sr.No	Phenomenon	Source
i.	Noise during idling	High compression and cylinder pressure.
ii.	Thriving Noise	Low order harmonics of inertia forces in multi-cylinder engines
iii.	Engine component reverberation	Harmonics of gas and inertia forces during respective compression and power strokes.
iv.	Vehicle component reverberation	Harmonics of gas and inertia forces.
v.	Airborne sound of engine	Mechanical impacts, combustion noise.

Table 3 Allowable sound level for road vehicles according to EU directive 96/20 EC

Sr.No	Type of Vehicle	Sound level dB (A)
i.	Personal car	74
ii.	Bus and truck weighing between3.5 to 2 tones and below	76
iii.	Bus with total weight above 3.5 ton and engine power below 15kW	78
iv.	Bus and truck weighing in between 2 to 3.5 to ton.	77
v.	For engine power 150kW or above	80

Figure 4 Noise and vibration reduction technique through flow path.

2.2 Exhaust and Intake Noise Control

Exhaust and intake system noise originates, from the pressure fluctuations of the engine and additional flow generated noise. Control of noise generation at the source involves making changes to the combustion process, which influences engine performance and exhaust gas emissions. So mufflers or silencers were used placed in a flow duct to prevent sound from reaching the openings of the duct. Reactive silencers do this by reflecting sound back towards the source while absorptive silencers attenuate sound using absorbing material Basic requirements for a modern exhaust systems; compact outer geometry, sufficient attenuation and low pressure drop.

2.3 Vibration Damping

Use of viscoelastic materials enhance the damping in a structure in three different ways viz. free layer damping treatment, constrained layer damping treatment, tuned viscoelastic damping treatment. The damping material is either sprayed on the structure or bonded using a pressure-sensitive adhesive in free layer damping. An interesting feature of the free-layer treatment is that the damping performance is independent of the mode shape of vibration. Constrained-layer damping includes a sandwich of two outer elastic layers with a viscoelastic material as the core. This damping is more effective than the free-layer design as more energy is consumed and dissipated into heat in the work done by the shearing mode within the viscoelastic layer. The TVDs are applicable to reduce vibration/noise associated with a single frequency or a narrow band of frequencies. Properly tuned TVDs eliminate an unwanted resonance by splitting the original peak into two, one below

and one above the resonance frequency of the original system.

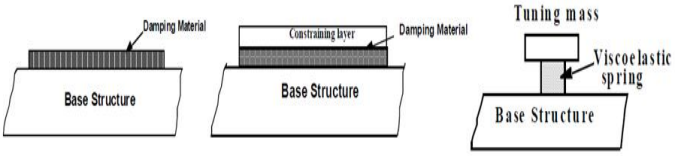


Figure5. Free Layer Damping, Constrained Layer Damping and tuned viscoelastic damping treatments

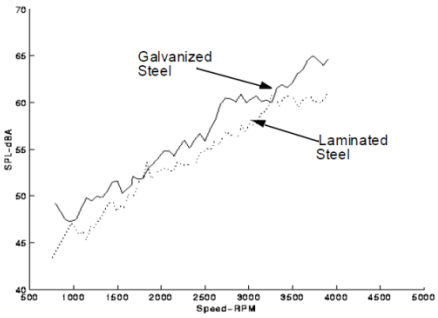


Figure6. Distinction of sound pressure level at driver's ear for a car equipped with regular and damped oil pans

CONCLUSION

Various sources of vibration from engine to driveline are identified and a detailed analysis of the cause is carried to influence the vibration characteristics of automobile. Engine vibration sources resulting from various forces viz: Combustion, reciprocating and rotational is reduced by using in-plane and two plane balancing methods. Method to reduce the rotating and reciprocating unbalance is gives the designer a means to influence the noise and vibration characteristics. The contribution of each vibration source and its reduction technique is focused here. The noise sources; their contribution and the engine related noise and vibration phenomenon in the vehicle is tabulated. An attempt is made to co-relate the noise and vibration sources and further methods to reduce the same are provided which reduces the noise and vibration and improve vehicle ride comfort characteristics. The energy flow method's four stages are used as a benchmark to reduce any vibration problems in automobiles arising from the random as well linear vibrations. It also helps in identification of the proper flow of vibrations and noise.

Noise one of the major concerns is taken care of here by identifying various noise sources and techniques to reduce the same are discussed which improves the vehicle ride quality and comfort level of passenger Interior noise is the major concern and is required to be taken care at the design stage itself. To reduce the vehicle interior noise manufactures prefer placement of sound absorbing materials, proper sealing of all openings and cavities present in the vehicle. Use of viscoelastic materials for damping vibrations is given, it is to be remembered that the use of these viscoelastic materials improves the damping properties of damper and are used in structures that are not primary load carrying members. For load carrying members, the design should first satisfy the strength and stiffness requirements over damping benefits. The noise limits for in-use vehicles are legislated as; highest noise levels permitted are 103 dB (A) as compared to the noise level of 96db (A) observed in new vehicles. Countries which have implemented similar requirements are Japan, New Zealand, California as well as USA under federal Motor Carrier Safety, Administration (FMCSA). [3].A detailed investigation regarding noise, vibration and their sources is very much essential to gratify the market and be a pioneer in automotive world. Further step is of identification of BSR problems and their reduction techniques.

REFERENCES

1. Raju.S, ARAI Pune, Workshop on Noise, vibration and harshness for automotive engineering, 2004, 123- *139.*
2. Herman V Auleraer, Noise and vibration characteristics of low emission vehicles. Keynote paper, *51-62.*
3. Karanath N.V. and Raju .S, Investigation of relation between stationary and pass by noise for new in use vehicle, *SAE paper No. 2005-26-051. ARAI Pune, 623-629.*
4.Nayak N. Reddy P.V, Aghav Y, Navtej Singh Sohi and A.D. Dani A.D.,Study of Engine vibration due to piston slop on single cylinder High powered Engine,Kirloskar Oil Engines Ltd. Pune, India, *SAE paper 2005-26-046, pp. 581-588.*
5.Mahale P.S, KalsuleD.J, A. Muthukumar A and Raju S, Vehicle interior noise source identification and analysis for benchmarking,ARAI Pune. *SAE paper 2005-26-048,592-603*
6. Gosavi S.S, Automotive Buzz, Squeak and Rattle (BSR) detection and prevention,TATA technologies Ltd, *ARAI Pune, 661-667.*
7. Rao M.D, Dept. of Mech. Engg. Mechanics, Michigan Technological University, Houghton, Michigan & 9931 USA 2001. Recent applications of viscoelastic damping for noise control in automobiles and commercial airplanes.
8.Gabiniemic J, Gatt J,Cerrato G. – Jay (Tecumesh products research laboratory) Automatic detection of BSR events. (Magna Automotive Testing).
9. Nashif A.D, Jones D.I.G and Henderson J.P *Vibration Damping, New York: 1985 John Wiley and Sons.*
10.Soovere J., Drake M.L and Miller V.R vibration Damping workshop Proceedings *1984* ,AFWAL-TR-84-3064 publications by Air force Wright aeronautical Laboratories, *Wright-Patterson Air Force Base, Ohio,* VV-

1-VV-10, a design guide for damping of aerospace structures.
11.Kung.S.W., and Singh R., Development of approximate methods for analysis of patch damping and design concepts. *Journal of Sound and vibration 219, 1999 785-812*

www.ijmer.com Vol.1, Issue2, pp-242-246 ISSN: 2249-6645

12. Vydra E.J and Shorgen J.P. Noise and noise reducing materials. *Society of Automotive Engineers, 1993 Paper No 931267.*
13.Kerwin E.M.Damping of flexural waves by a constrained viscoelastic layer. *Journal of the Acoustical Society of America 31 1959, 952-962.*
14. Lilley K.M, Fasse M.J and Weber P.E 2001. "A comparison of NVH treatments for vehicle floorplan applications". SAE paper No.2001-01-1464.
15. Hussaini A,Designing an interior waterborne coating for use in automotive paint shops to replace sound deadening pads. *SAE paper no. 2001-01-1391. 2001*

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-247-251 ISSN: 2249-6645

Performance Evaluation of Filter-bank based Spectrum Estimator

M.Venakatanarayana[1], Dr.T.Jayachandra Prasad [2]

[1]*Assoc . Prof , Dept. of ECE, KSRM College of Engg., Kadapa.*
[2]*Professor, Dept. of ECE, RGMCET, Nandyal,*

Abstract—
In this paper on attempt has been made to study the performance of Filter-bank based nonparametric spectral estimation. Several methods are available to estimate non parametric power spectrum. The band pass filter, which sweeps through the frequency interval of interest, is main element in power spectrum estimation setup. The filter-bank based spectrum estimation is developed and is applied to multi tone signal. The spectrum estimated based on filter-bank approach has been compared with conventional nonparametric spectrum estimation techniques such as Periodogram, Welch and Blackman-Tukey. It is observed that the filter-bank method gives better frequency resolution and low statistical variability. It is also found there is a tradeoff between resolution and statistical variability.

Index Terms— **Correlogram, Filter-bank, Frequency resolution, Periodogram, Spectral leakage etc.,**

I. INTRODUCTION

In Signal processing, the nonparametric spectrum estimation plays an important role in determining perioditicity in random signals and thus a comprehensive elaboration of filter-bank based spectrum estimation techniques has been presented. In general, spectrum estimation can be categorized into direct and indirect methods. In direct method (usually recognized as frequency domain approach), the power spectrum is estimated directly from signal being estimated $x(n)$. On the other hand, in indirect method, also known as time domain approach, the autocorrelation function of the signal being estimated $R_{xx}(k)$ is calculated. From this autocorrelation value, the power spectrum density can be found by applying the Discrete Fourier Transform on$R_{xx}(k)$. Another way to categorize spectrum estimation methods is by classifying them into parametric or non-parametric methods. Parametric method is basically model based approach [10]. In this method, a signal is modeled by Auto Regressive (AR), Moving Average (MA) or Auto Regressive Moving Average (ARMA) process. Once the signal is modeled, all parameters of the underlying model can be estimated from the observed signal. Estimator based on parametric method provides higher degree of detail.

The disadvantage of parametric method is that if the signal is not sufficiently and accurately described by the model, the result is less meaningful. Non Parametric methods, on the other hand, do not have any assumption about the shape of the power spectrum and try to find acceptable estimate of the power spectrum without prior knowledge about the underlying stochastic approach. The following sub-sections give review on some of the spectrum estimation methods.

A. Periodogram

The most commonly known spectrum estimation technique is periodogram, which is classified as a non parametric estimator. The procedure starts by calculating the Discrete Fourier Transform (DFT) of the random signal being estimated, followed by taking the square of it and then dividing the result with the number of samples *N*.

There are five common nonparametric PSE available in the literature: the periodogram, the modified periodogram, Bartlett's method, Blackman–Tukey, and Welch's method. However, all these nonparametric PSE, s are modifications of the classical periodogram method introduced by Schuster.

The periodogram is defined as [2]

$$\hat{S}_{xx}(k)=\frac{1}{N}\left|\sum_{n=0}^{N-1}x(n)e^{-\frac{2\Pi jkn}{N}}\right|^2 \qquad (1)$$

It is known that the periodogram is asymptotically unbiased but inconsistent because the variance does not tend to zero for large record lengths. One can show [10], that the Variance on the periodogram $\hat{S}_{xx}(k)$ of an ergodic weakly stationary signal $x(n)$ for $n=0:N-1$is asymptotically proportional to $S^2{}_{xx}(k)$, the square of the true power at frequency bin. The periodogram uses a rectangular time-window, a weighting function to restrict the infinite time signal to a finite time horizon, the modified periodogram uses a nonrectangular time window [3].

A way to enforce a decrease of the variance is averaging. Bartlett's method divides the signal of length

N into K segments of length $L = \frac{N}{K}$ each. The periodogram method is then applied to each of the K segments. The average of the resulting estimated power spectra is taken as the estimated power spectrum. One can show that the variance is reduced by a factor K, but the spectral resolution is also decreased by a factor K, [2]. The Welch method eliminates the tradeoff between spectral resolution and variance in the Bartlett method by allowing the segments to overlap [4].

B. Blackman-Tukey method (Windowed Correlogram)

Blackman-Tukey method is a variant of correlogram that computes the approximated autocorrelation $R_{xx}(k)$ and later applies a suitable window function $w[k]$. The power spectra density is then obtained by computing the Fourier Transform of windowed auto-correlation sequence [10].

Blackman-Tukey method is generally described as follows

$$S_{xx}^{BT}(e^{jw}) = \sum_{k=-L}^{L} \hat{R}_{xx}(k)\, w(k) \exp(-jwk) \quad (2)$$

And its frequency domain representation is given by

$$S_{xx}^{BT}(e^{jw}) = \frac{1}{2\pi} W(e^{jw}) S_{xx}^{c}(e^{jw}) \quad (3)$$

from equation (3) the Blackman-Tukey method can actually be viewed as a process of smoothing the correlogram by convolving the correlogram with the kernel of selected window. This smoothing process plays an important role to reduce the bias of estimated PSD but this convolution process would reduce the frequency resolution. The amount of frequency resolution reduction is strongly related to the size of the main lobe of the window kernel. Section II talks about filter bank approach to spectrum estimation technique. Section III focuses the performance analysis of the proposed techniques in comparison with the conventional techniques.

II. Spectrum Estimation as a Filter Bank Analysis

From the perspective of spectrum estimation, a filter bank can be considered as an array of band pass filters that separates the input signal into several frequency components, each one carrying a single frequency sub-band [6]. The filter banks are usually implemented based on single prototype filter, which is a low pass filter. This low pass filter is normally used to realize the zero-th band of the filter bank while filters in the other bands are formed through the modulation of the prototype filter [9]. Figure 2.6 illustrates the main idea of filter bank concept. This section basically tries to explore the filter bank paradigm in spectrum estimation.

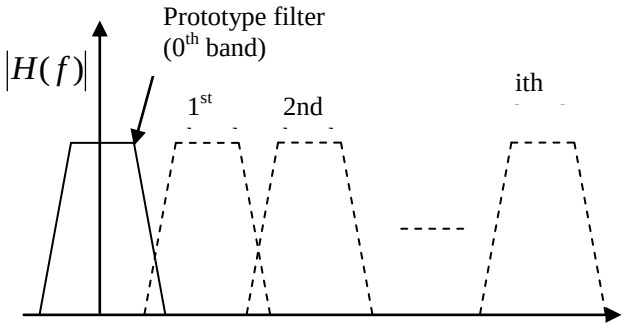


Figure.1 The filter bank concept.

A. Periodogram spectral estimator realization through filter banks

Spectrum estimation is about finding the power spectrum density (PSD) of a finite sample set $\{x(n), n = 1,2,\ldots\ldots N\}$ for frequency $|\omega| \le \pi$. The classical approach to spectrum estimation is to use Fourier transforms to obtain a Periodogram, given as [17]:

$$S_{xx}^{p}(e^{j2\pi f}) = \frac{1}{N}\left|\sum_{n=1}^{N} x(n) e^{-j2\pi f n}\right|^2 \quad (4)$$

for any given frequency f_i, (4) is written as:

$$\begin{aligned} S_{xx}^{p}(e^{j2\pi f_i}) &= \frac{1}{N}\left|\sum_{n=1}^{N} x(n) e^{-j2\pi f_i n}\right|^2 \\ &= \frac{1}{N}\left|\sum_{n=1}^{N} x(n) e^{j2\pi f_i (N-n)}\right|^2 \end{aligned} \quad (5)$$

it should be noted that (5) is possible since $\left|e^{(j2\pi f_i N)}\right| = 1$. By introducing new variable $k = N - n$, equation (5) is written as:

$$\begin{aligned} S_{xx}^{p}(e^{j2\pi f_i}) &= \frac{1}{N}\left|\sum_{k=0}^{N-1} x(N-k) e^{j2\pi f_i n}\right|^2 \\ &= \frac{1}{N}\left|\sum_{k=0}^{N-1} h_i(k) x(N-k)\right|^2 \end{aligned} \quad (6)$$

where $h_i(k) = \frac{1}{\sqrt{N}} e^{(j2\pi f_i k)}$ for $k = 0,1,2,\ldots\ldots N-1$

by observing the summation within the magnitude operation in (6) and the summation expressed as:

$$y(N) = \sum_{k=0}^{N-1} h_i(k) x(N-k) \quad (7)$$

(7) is actually the truncated convolution sum at particular point N, which is again written as general convolution sum at the same point associated with a linear causal system by padding

$h_i(k)$ with zeros[43]. Then (7) is rewritten as

$$y_a(N) = \sum_{k=0}^{\infty} h_i(k)x(N-k) \tag{8}$$

with

$$h_i(k) = \begin{cases} w(k)e^{(j2\pi f_i k)} & for k = 0,1,2,\ldots\ldots\ldots N-1 \\ 0 & otherwise \end{cases} \tag{9}$$

and window function $w(k) = 1/\sqrt{N}$. It is clear that (8) represents as passing N samples through a filter having impulse response $h_i(k)$ and then taking only single sample of the filtered signal at point N. Based on this perspective, the frequency response of the linear filter having impulse response $h_i(k)$ is

$$H_i(\omega) = \sum_{k=0}^{\infty} h_i(k)e^{-j\omega k} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{j(\omega_i - \omega)} = \frac{1}{\sqrt{N}} \frac{e^{j(\omega_i-\omega)N} - 1}{e^{j(\omega_i-\omega)} - 1} \tag{10}$$

finally gives

$$H_i(\omega) = \frac{\sin[N(\omega_i - \omega)/2]}{\sqrt{N}\sin[\omega_i - \omega)/2]} \exp[\, j(\frac{N-1}{2})(\omega_i - \omega)] \tag{11}$$

If $w(k)$ in (9) is taken to be a prototype FIR (Finite Impulse Response) low pass filter, then $h_i(k)$'s will constitute a bank of band pass filters centered at frequencies f_i s. This filter bank is constructed by modulating the prototype filter. By considering (4)-(11), the periodogram estimate at particular frequency point f_i is obtained by passing the received samples through the band pass filter centered at f_i. The power calculation of this estimate is performed based only on a single sample of the output of the filter from (8). The entire periodogram estimates can then be related to the output of several filters in the filter bank constructed by modulating a single prototype filter $w(k)$. For the case of simple periodogram, the window function $w(k)$ is rectangular with $w(k) = 1/\sqrt{N}$. As it is clear from (11), the frequency response of filter based on prototype filter having rectangular

window as its impulse response would have significant level of side lobes. This is actually the main reason why the periodogram estimates have high side lobe or large leakages. This problem can be alleviated by replacing the rectangular window with a window function with a taper that smoothly decays on both sides to obtain a prototype filter with much smaller side lobes. A few popular windows are Hanning, Kaiser and Blackman [8]

.

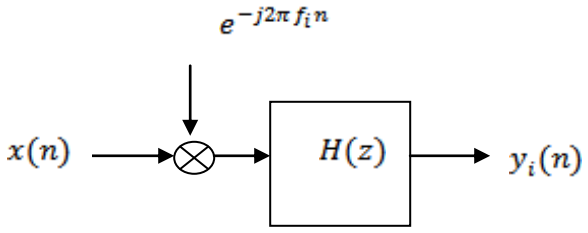


Figure.2 The demodulation of received signal

From the above, the implementation of a spectrum estimator using filter bank for signal analysis is clear, namely by passing an input signal through a bank of filters. The output power of each filter is a measure of the estimated power over the corresponding sub-band. Hence the power spectral density (PSD) estimate of *i*-th sub band of the filter bank is represented as [16]:

$$\hat{S}\left(\frac{i}{N}\right) = avg[|y_i(n)|^2] \tag{12}$$

In (12), avg [] describes time average operator while $y_i(n)$ is the output signal of i^{th} sub band filter.

III. Results and Analysis

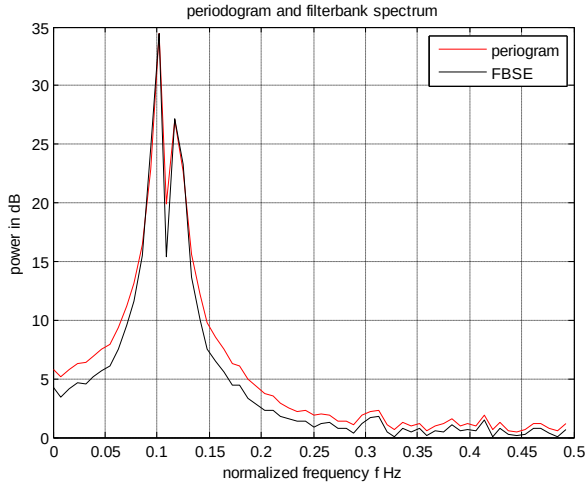
Consider a signal having two closed frequencies embedded in white noise i.e. $x(n) = \sin(2\pi f_1 t) + \sin(2\pi f_2 t) + e(n)$, where $f_1 = 0.1, f_2 = 0.12$ and $e(n)$ is white noise having zero mean and unit variance. The performance of the estimation techniques is mainly evaluated with respect to three different parameters:

- Frequency resolution
- Variance of the estimated power spectrum density (PSD)

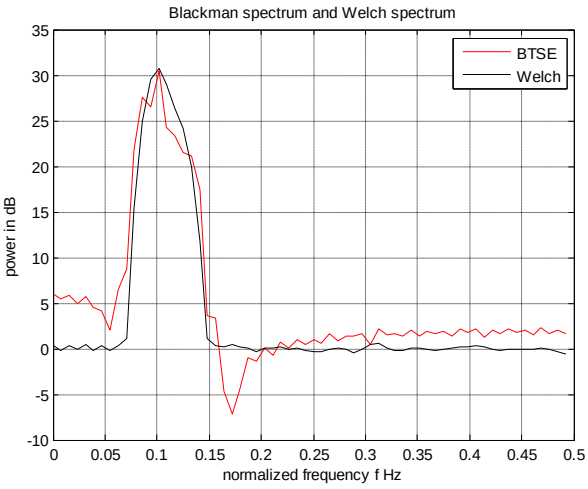
Fig. 6 depicts Periodogram, Blackman-Tukey, and Welch approach as well as filter-bank based estimates for the case of a signal having two closed frequencies. In these figures, the number of samples used in the experiment is N=128,256 and 512. For the purpose of this experiment, the Welch approach divides the received samples into M=4 segments of K=N/M samples. Two consecutive segments overlap to one another by 50%. Before performing the averaging process, Hamming window is applied on each segment. As in the case of Blackman-Tukey approach, a triangular window is used having its length K=N/2. The number of bandpass filter in filter-bank implementation is denoted by K.

In Fig.6, assumed K=1, where both the filter-bank based spectrum estimate and periodogram based spectrum estimate have good frequency resolution with high statistical variability. In Fig. 7 and Fig.8, assumed K=4, where the filter-bank spectrum estimate is much better than periodogram having low variability while maintaining acceptable frequency resolution. It is also observed that the level of estimated power in unoccupied band for Welch approach is higher than for simple periodogram meaning that the Welch approach offers poorer rejection in the unoccupied band. This is understandable since

Welch approach divides the received samples into several segments with lower number of samples before estimating each segment. Hence Welch approach offers very low variability and poor frequency resolution. The Blackman-Tukey method is also good estimates to power spectrum having low variability and moderate frequency resolution. Finally, we noticed that when K=4 and N=256 and 512, with respect to variance and frequency resolution, the filter bank approach most preferable to estimate power spectrum among the above techniques.

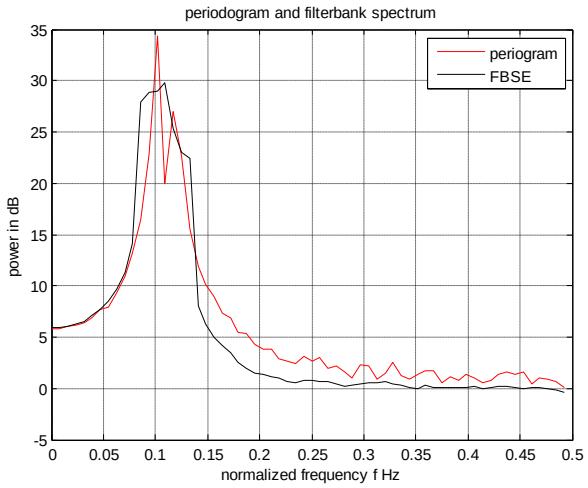


(a)

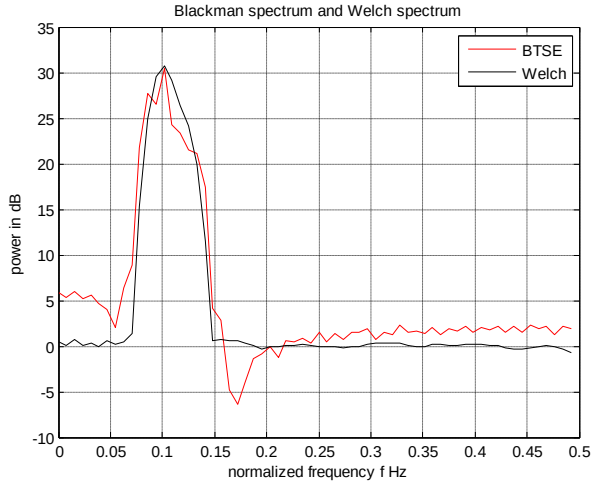


(b)

Figure. 6 (a) Periodogram and filterbank based spectrum (b) blackman tukey an Welch based spectrum (K=1, N=128)

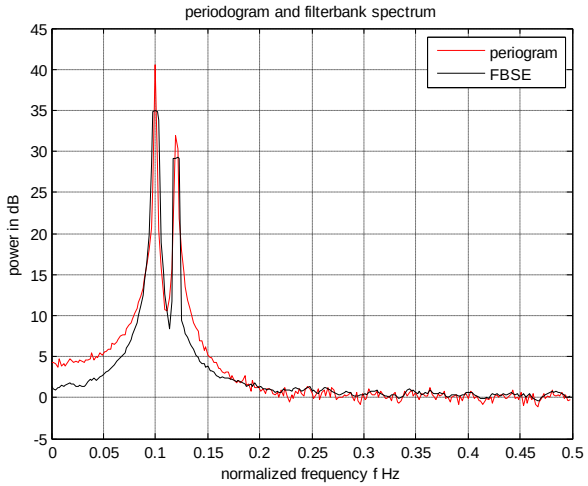


(a)

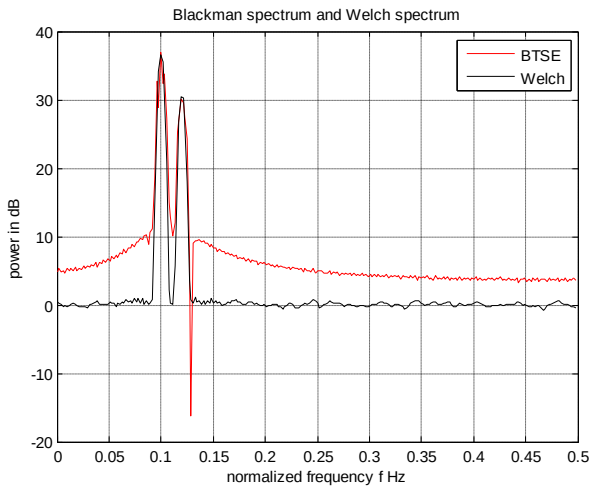


(b)

Figure.7. (a) Periodogram and filterbank based spectrum (b) blackman tukey an Welch based spectrum (K=4, N=128)



(a)



(b)

Figure.8. (a) Periodogram and filterbank based spectrum (b) blackman tukey an Welch based spectrum (K=4, N=512)

Table I: the number of bandpass filters K=1, In Welch approach (hamming window, 50% overlap and N/4 segments) and in Black-man Tukey (hamming window of length N/2)

Variance	N=128	N=256	N=512
Periodogram	1.2156e+005	1.3194 e+005	5.2874e+005
Filter-bank approach	1.2409e+005	1.5029e+005	5.2829e+005
Black-man tukey	0.3110e+005	0.7978e+005	1.5834e+005
Welch	0.4704e+005	0.8525e+005	1.7550e+005

Table II: the details are same as in Table I, except change in the number of bandpass filters K=4

Variance	N=128	N=256	N=512
Periodogram	1.2056e+005	1.3177e+005	5.3098e+005
Filter-bank approach	0.3772e+004	0.7844e+005	1.4257e+005
Black-man tukey	0.3047e+004	0.7946e+005	1.5844e+005
Welch	0.4631e+004	0.8513e+004	1.7584e+005

The variance analysis among these techniques as follows

- From Table I and Table II, for length of samples N=128,256 and 512, the Blackman-Tukey approach offers low variability than other methods.
- From Table II, for a record length of N=256 and 512, the FBA offers low variability without scarifying frequency resolution than other methods.

The frequency resolution among these techniques as follows

- From Figures, for record length of N=256, the Welch approach does not resolve two closed frequencies. Its resolving capability increases with data record length.
- From Figures, when K=4, N=256 and N=512, the Filter-bank approach resolve two closed frequencies.

IV. CONCLUSION

In this paper on attempt has been made to develop and implement filter bank based nonparametric spectral estimation technique. The proposed technique has been subjected to multi tone signal, and estimated the spectral components. The performance of proposed technique has been compared with the conventional methods of nonparametric spectrum estimation such as periodogram, Welch and Blackman-Tueky. It is observed that the FBA method produce spectral estimates with high resolution and low statistical variability at expense of increased the number of bandpass filters. Hence, there is tradeoff between resolution and statistical variability. The studies show that the Filter bank based spectrum estimation is simple, offers great flexibility, reconfigurability and adaptability.

REFERENCES

[1] S. Haykin, "Cognitive Radio: Brain-empowered Wireless Communications", IEEE

[2] J.G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Fourth Edition. Upper Saddle River, NJ: Prentice Hall, Inc, 2007.

[3] M.S. Bartlett, "Smoothing Periodograms from Time Series with Continuous Spectra", *Nature* (London), Vol.161, May 1948.

[4] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", *IEEETransactions on Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 70-73, June 1967.

[5] F. J. Harris, *Multirate Signal Processing for Communication Systems*, New Jersey: Prentice-Hall PTR, 2004.

[6] B. Farhang-Boroujeny, "Filter Banks Spectrum Sensing for Cognitive Radios", *IEEE Transaction on Signal Processing*, Vol. 56, pp. 1801-1811, May 2008.

[7] J. Lim and A.V. Oppenheim, *Advanced Topics in Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1998.

[8] D. J. Thomson,"Spectrum Estimation and Harmonic Analysis", *Proceeding of IEEE*, vol. 70, no. 9, pp. 1055-1096, September 1982

[9] B. Farhang-Boroujeny, "A Square-Root Nyquist (M) Filter Design for Digital Communication Systems", *IEEE Transaction on Signal Processing*, vol. 56, no. 5, pp. 2127-2132, May 2008.

[10] P. Stoica and R.L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, Inc, 1997.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-252-260 ISSN: 2249-6645

On some commercial and technical aspects of multimedia educational systems

Sk.Khadar Babu & M.N.Srinivas

School of Advanced Sciences,
VIT University, Vellore,
Tamil nadu,India.

Abstract:

This article was made to identify some technical aspects of multimedia in education and it focuses on the issues related to identification of technical aspects of the multimedia educational systems. It also explains the different tools and his operating procedures of the technical components. Some modules are created consisting of interactive learning sessions which high light the features offered by the technology. It includes several short presentations on assorted topics from subjects such as physics, mathematics and chemistry. These modules are demonstrated to students and teachers belonging to different educational institutions.

---

1. Introduction:

"The characteristics of multimedia elements are so diverse in nature. It is highly impossible to manage them through a particular hardware or software tool". The audio and video elements have complex spatial and temporal characteristics which require special hardware and software techniques to generate and play them.

The multimedia titles strive to create a personalized learning environment through an intelligent combination of user interface mechanisms. These multimedia educational systems involve the user in every stage of the learning process such as creating awareness, understanding the concept and developing the skills to utilize the knowledge acquired on a given topic.

We choose edutainment as our application in which education is imported in an entertaining manner through interactive multimedia titles. Multimedia technology allows computers to handle different media elements like text, graphics, animations, audio and video. It is essentially a combination of the television and computer technology. The communicative power of audio visual medium coupled with the interactive capabilities of the computer systems.

According to Hertfordshire (1995), the main features of multimedia educational systems are provision of individual attention to learners, self-paced and user controlled learning environment convert the dull and passive learning in to a dynamic one.

A demo module is created consisting of interactive learning sessions which high light the features offered by this technology in T. Vaughan (2000). The demo module includes several short presentations on assorted topics from subjects such as physics, math's and chemistry. These modules are demonstrated to students and teachers belonging to different schools and collected the feed back them in the form of quest ionized. The system development concepts are discussed and identified by the Bohdan.o (1995) and B.M. Panday (1995).

The way we learn

It is argued that the depth of knowledge required in secondary schools makes it necessary for students to have many teachers who are subject experts. The development of multimedia technologies could soon after access to knowledge far superior than that of most subject teachers. This may result in teachers needing to develop new learning facilitation skills and reduce the emphasis on knowledge .The know-how is likely to become more valuable than the know what. Teachers primarily require access to learning resources, which can support concept development by learners in a variety of ways in order to meet individual learning needs.

2. Multimedia in Education

Over a past twenty years, education delivery systems are being constantly redefined and the computer has come to play a vital role in implementing new teaching –learning strategies. A fundamental basic difference between multimedia based learning and conventional system of learning is that in the conventional system, the book is basic material which follows its own step-by-step structure and the contents are accordingly structured. In multimedia, on the other hand, the content structuring has to be altered so as to incorporate the self-paced and non-linear, interactive exposition possibility, besides, the fact that the audio and visual material plays a greater role in multimedia than in a book.

In future the "classroom of tomorrow ''as a multimedia class room. The benefits of multimedia-bases classroom instruction in education are a stimulating teaching and learning environment and the encouragement of student ownership and self expression in their learning. Hypermedia materials are more engaging than the traditional print materials .The computer with a graphical user interface is an integral component of this multimedia-based teaching. Multimedia can allow students in a variety of disciplines to explore concepts that are typically unavailable to them: they might be too dangerous, impossible to explore, or just out of the budget of the institution. This might includes the examination of microscopic environments like journeys through the various organs of the body or the observation of violent chemical reactions in D.P. Mukherji (2000). For students, multimedia may allow them to perform analysis with out acquiring powerful, prohibitively expensive, and overly complicated software and hardware.

2.1 Modes of Multimedia Educational systems use in Earning

There are at least five modes in which multimedia resources can be used in teaching and learning of these are

1. Support
2. Exploration & control
3. Tutorial
4. Resource
5. link

2.1.1 The Support Mode

In the support mode, a student uses the computer to enhance the presentation of work. The computer can help by increasing the accuracy that the student might otherwise be able to achieve. There are a great number of tools that are currently used in some institutions.

2.1.2 The Exploration &Control mode

In exploration and control mode, the student is able to examine and build situations. This kind of work is often associated with curriculum materials. In this mode, students might create applications which will subsequently be used in any of the other modes. This are gives students the important chance to examine the social effects of multimedia technology.

2.1.3 The Tutorial mode

In the tutorial mode, the student will expect to learn new knowledge or skills. This will also give students opportunities to develop at their own pace and to receive feedback upon their progress .It is important to include the possibility of using the tutorial mode for assessment only.

2.1.4 The Resource mode

In the resource mode, the multimedia system is used to access information and other resources. When using the resource mode, students are developing questioning skills. They are solving problems by stating them and re-shaping them to fit different resource frameworks.

2.1.5 The Link mode

The link mode is typified by the computer being used for the communication between individuals. This represents an important role for computers in for coming century. This mode comes to the fore in projects such as the classroom without walls or the global school house.

3. Technical aspects of multimedia in education

This article begins with a discussion on the design constraints which is followed by an overview of the life cycle of the multimedia educational CD development.

3.1 Design constraints:

We have two choices for the development of such educational modules, usage of high level languages such as c++ or VB to create educational modules or utilize the existing s\w tools and create educational CDs. In the first case, it is not sufficient to design the contents of the presentation but it is also necessary to develop the required s\w tools to create and run the presentation. The media elements have their own timings. Constraints which make programming these devices a highly complex process. These are explained in Andrew.e (1) and Bates a.w (2). In order to create a full fledged CD through this approach, we need a team of experienced professionals with hardware and software backgrounds and a lengthily development time.

In second case, by making use of the existing s\w tools it is possible to produce such CD with ease and in a short time as the authors are free from specifying the intricate timing details and the control of several devices associated with the CD which is taken care by these tools. This allows concentrating on the content design rather than on the presentation issues.

In view of R.Hone & M.Kuntz (7), Most of the creators adopt the second approach as, the volume of the contents is very large and it may well become unprofitable and time consuming if we choose the first option. However, certain requirements of CD making not supported by the authoring tools were implemented using the first approach. This is possible by creating customized routines using any high level language and integrating them with the presentation using the authoring tool. Author ware supports customized functions and routines in the form of DLLs.

3.2 Multimedia CD development

The development of educational CD figure is shown below. In addition to the normal process of analysis, design, implement and test, the development of educational cds involves other tasks such as creation of media elements and integrating them to form a meaning presentation.

The media elements were created using appropriate s\w tools and also obtained from scanners, video cameras, tapes and other sources .The authoring tool uses to integrate the media elements according to create executable files. In preparation of CD s, the authoring tool creates the executable modules which

could be run under windows environments. These modules are copied on to a CD-ROM and tested on variety of platforms for finding out the performance of the educational CD.

3.3 Components of the title

The component of the multimedia presentation have been classified according to the way users access the information .The various components view and process the contents in different manner and display the information to the users . The classification is shown in figure is basic and common to similar multimedia applications. The collection of media elements describing the content information forms the major component in any multimedia applications. The media elements were created using appropriate s/w tools.

3.4 Navigation and browse

The key to the success of multimedia educational systems lies in providing the control of learning /viewing process to the users. These are several traditional techniques which are text intensive such as contents search, history list, and context sensitive help. The navigation features provided in multimedia product is extending to other media elements since the browse mechanism should suit the topics described through any media other than text.

3.4.1 Contents

Manufacturers or creators must design the contents menu to appear in the menu box. This can be done by adding an interaction icon titled contents to the flow line as seen in the design window of the figure. The list of the content is treating by attaching different map icons to the content icon. Each bearing the name of the topic .The interaction type of these map icons is set to filename type. So, that once the user clicks on the contents, a list of the available items displayed. The users have to branch to the chosen type, once they click on a particular item .Authors should take care of erasing the screen when the presentation starts a new topic. This can be done by the erase icon at the beginning of the new module or at the end of current module. Thus, the creator can create any number of menu options on the menu bar .The option of selecting either returnable or non-returnable jumps which provide great flexibility and choices in the design.

3.4.2 Menu

In this section, we discuss the menu options available on area other than the menu bar of the presentation windows. Here the interaction type is set to click/touch option. Hot-spots define unique clickable areas on the screen for each of the menu item. Programmers can display any text or image with in such hot-spot areas which serve as an indication of the contents and when user clicks on these areas, control branches are appropriate points through jump instructions.

3.4.3 Media gallery

This feature allows one to browse through the contents by specific categories based on the media elements. For example, one can view only the video clips of the equipments, select only the animations, image or textual narrations among available material in title by clicking on the appropriate menu item.

3.4.4 History lists

It provides the users with a list of pages or modules they have browsed since the start of the session. So that they may repeat any of the contents, if necessary. The references to the pages and modules are maintained through headers by the system.

3.4.5 Miscellaneous

There are several unique browse requirements for audio and video elements such as pause, stop, repeat and other features as represented by VCR metaphors. These can be built by using push buttons with appropriate text or image.

3.5 Tools and utilities

There are several instances where a user should be able to perform other tasks while remaining inside the presentation. The utilities shown in figure have been made available in multimedia CDs under the icon tools.

3.5.1 Bookmark

We can create a custom routine which saves this variable under the book mark category in the user records directory along with many runtime records such as username, log-in time, values of variables etc., and these variables have been used to keep track of run-time parameters of the CD. When user starts a new session, they are provided with an option of starting the presentation from the saved bookmark. These bookmarks can be accessed at any point of time with in the multimedia CD and can be used similar to that of history list.

3.5.2 Options

Under this heading provides facilities for the user to customize or control the run-time parameters of the application such as volume control of the audio element ,background/foreground colours , printing the screen content etc.,

3.5.3 Help

This feature provides in a similar style as that of contents button discussed in navigation sub section. The same structure which has been built for navigation has been utilized for the implementation of this feature. Creators have provides several hypertext links in the text

3.6 User interface

Presently, two major input devices available for the user to interact with the system are keyboard and the mouse. The user interface design is uniquely depends on the content in multimedia CD and users interest alive. The user interface is a challenging area of any CD which is limited only by the imagination of the creator in creating an appropriate and effective interface for the topic.

In any form if users interface, users have to be prompted on how to interact with the system. This becomes more important in novel user interface designs. Once the interacts. The system checks whether the user has responded correctly. To do this, the system expects the students create all the possible ways the users may respond to the given interaction.

3.6.1 Text interaction

This type of response is used mostly to obtain textual information from the users such as user names, and annotations. Normally programmers use this type of response in quiz sessions. The system waits for the user to key in an answer and this string is checked with the correct answer indicated by the programmer during the design stage. If it matches then the score is increased. On the other hand, if it does not match, an appropriate feedback message is given and control branches the next question.

3.6.2 Drag and drop interface

Here, it is possible for the user to move the object in many in correct ways. In those cases programmers provides appropriate message and the object back to its original position and the user to try

again. This is done by creating a map icon and selecting all the area on the screen other than one marked for correct response as the destination for the movable object.

3.6.3 Click-ons

In this interaction icons or images are displayed to the user and clicking a particular image. Once the user clicks on any one of the alphabet, an audio clip which pronounces the alphabet and an animation clip illustrated the procedure of that alphabets are displayed.

3.7 Performance evaluation

It is always essential to evaluate the user after the learning session as it reinforces the information in their minds .It is meaningful only if the acquired information is put to use in real life and tests and evaluation procedures ensure that the information has indeed comprehended by the users. Hence this component plays a vital role in determining the usefulness of the product.

Most of the programmer's uses mechanisms to text the user's comprehension through quiz and drill sessions. In the quiz session, solutions to the questions are not provided, and score is maintained and indication the number of correct answers given by the user. In drill sessions, users are allowed to repeat the questions till they get the correct answer. The prime idea here is make the users recollect the information learnt.

3.7.1 Object type questions

This type of question –and-answer session is most important in present day competitive exams and provides an exhaustive bank of questions of this day.

3.7.2 Match the correct answers

This type provides through drag and drop interface where the users are requested to drag the correct and drop them in appropriate slots. The students are requested to form the proper arrangement of sentences which describes the experiments correctly.

3.8 User data

The personal data about the users such as names ,roll names, class, section and the performance related data such as levels , scores, number of tries ,time taken to answer re logged in separate files. By this user data we can develop customized routines such as average, standard deviation and display the results through bar charts or other graphical mechanisms.

3.9 Development of S/W tools

In this section, the details of utilities developed for the creating of multimedia CDs.

3.9.1 FEATURES OF THE WINDOWS ENVIRONMENT UTILISED FOR THE MULTIMEDIA CD CREATION

Windows provides several high level multimedia services in the form of MCIs (media control interface) and APIS (Application Programmers Interface) which are utilizes by the high level language such as VB and C++. There are two types of interfaces: the string and command interface.

String interface consists of text based commands which are parsed and compared with a string look up table and translated in to device-specific instructions by the driver's .THIS Method is suitable for script-based languages where a sequence of action can be performed by sending a string of MCI commands to MCI drivers.

Message/command interfaces programmers are required to fill an appropriate structure which consists of commands, flags. These structures are sent to devices directly as they map on the command set of the specific devices. Hence execution may be faster in this mechanism due to the absence of look-up table comparison. The information returned from the devices is also available in similar structure from the MCI drivers which the application can make use of in getting the status of the command sent to the drivers.

3.9.2 VB SUPPORT TOOLS FOR MULTIMEDIA APPLICATION DEVELOPMENT

Most of the programmers develops customizes routines through VB for the development of performance evaluation of performance evaluation modules and some unique navigational and browse mechanisms of the CD.VB offers several techniques for multimedia application develops such as dynamic link libraries (DLLs),OLE controls,OLE automation servers and VB custom controls (VBXs)which are discussed in the following paragraphs .

A DLL is a library routine which can be called from a program and associated code is linked with the application only at the runtime and not during the design/compile time. More over the DLLs can be shared by programs written in any language since windows provides a standard language definition for a DLL and ensures that parameters are passed in a consistent manner.

Object linking and embedding (OLE)is a feature unique to the windows environment which enables communication between application currently running under windows. OLE techniques define an interface with which application can be made programmable regardless of the programming language environment. An object consists of properties and a method represents the data variables that the container application manipulates methods constitute a specified action performed on these objects.

There are two types of OLE automation servers in process servers and local server.

These techniques are utilized by the programmers in allowing branching between applications under the windows environment.

3.9.3 INTEGRATION OF CUSTOM –BUILT ROUTINES

Each DLL has several routines with a fixed procedure for parameter passing. Which is described by windows .VB allows compilation of source code in to an executable or DLL file and programmers used the DLL option and compiled the source code in to DLLs. The DLL is converted in to VCD file, which is a header.

Information read by authoring tool, this header consists of functions available in the DLL. The required parameter and definitions, the VCD files and thru DLLs are distributed along with the multimedia application, and thus any function called by the title is linked with the appropriate library routine during run-time and executed by windows.

3.10 PERFORMANCE EVALUATION MODULE

The performance evaluation module can be utilized by teachers to conduct tests. The system supports the teachers to type-in the questions, create multiple choices for the solutions, indicate the correct choice scoring procedure, an time constraints in answering the questions. The text material which is input y the teacher is converted in to appropriate authoring structure acceptable to author ware. Once this done, teachers can preview and edit the quiz.

Once the quiz session is launched the students have to be prevented from logging in to others account so as to prevent any malpractices. This is taken care by allowing the students to open the quiz session only once and prevent them from opening any other tasks during the session.

VB provides elegant mechanisms to implement the above procedure through the OLE techniques where the system from which the quiz was launched becomes the server and the other systems where the students take the tests act as clients.

4. CONCLUSIONS

However, the most difficult part of the system to put in place will be an appropriate educational infrastructure to support the kind of learning needed in the 21 st century. The provision of appropriate education and training services to run on the information highway is critical: there is no automatic guarantee that people will use the information highway to an extent that justifies the cost of investment. If services are not provided that meet peoples needs. Unfortunately, existing educational institutions were created to meet the needs of a society that are fast disappearing. We need new educational organizations that can exploit the information highway to meet the needs of the 21st century. Economic development will depend on much on the success of creating and supporting such organizations, as an establishing the technological infrastructure .It is critical to get this right because those countries that harness the power of multimedia communications for education and training purposes will be the economic powerhouses of the 21st century.

The future of interactive media in education ,when it is separated from the issues of technology that tend to mystify discussion and place it firmly in the sole grasp of those that are highly technologically capable ,is that of communication tool. Its dimensions and capabilities will evolve and expand at the same time as the potential to author becomes more widely accessible .The potential for students of all ages to author as part of a creative educational programme that is based on achievement of goals and competencies rather than time served will assist educators to shift from teacher to facilitator and mentor. Interactive communication tools will transform our capability to embrace an educational paradigm that deals with learning as a vital, fulfilling, and continuing part of life at home and in the workplace as well as within educational institutions.

The results of the above are summarized in the following points.

1. The number of educational institutions adopting education through multimedia is on the rise and there are not many organizations to meet the demands of the institutions.
2. The prices of the systems are becoming affordable to many and we force a trend of multimedia computers becoming a consumer product in the next five years.
3. The existing products have been developed abroad and they do not target the domestic audience to a great extent.
4. The market is opening up with more and more organizations becoming aware of the potential of such products.

REFERENCES

1.Andrew E.(1995)"computers in schools – a frame work for development , Fluck of Claremont College, Tasmania. Web address is http://www.clare.tased.edu.au/acspaper/compsch1.html.

2. Bates, A.W.1993)'educational aspects of the telecommunications revolution'in Davies, G.andSamways, B. Teleteaching / newyork/Amsterdam: North Holland

3. BM. Pande &Abhay Mohanlal,(1995)"Multimedia in education: Rock art , International conference on Multimedia for humanities,New Delhi.

4. Bohdan .O. Szuprowicz Multimedia Technology: Combining sound, text Computing graphics & Video, Computer technology Research Corporation, South Carolina, 1995.

5. Paul P.L Regtien (et al) (2007): Comet: A multimedia internet based platform for education in measurement, volume 40, issue2,pages,287-292.

6 Judith jeffcoate Multimedia in practice: Technology and application. Prentice hall International (UK), Hertfordshire, 1995.

7. Robert Hone and Margy Kuntz: "making moves with your pc", Pustak Mohal, Jan1995.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-261-269 ISSN: 2249-6645

An Optimized Algorithm For Ringing Region Detection In Compressed Images

Mahalakshmi vandana , R. Srinivas

(Student,M.Tech,Department of computer science and engineering, AITAM College, India
(Associate Professor,Department of computer science and engineering, AITAM College, India

ABSTRACT

In current visual communication systems, the most essential task is to fit a large amount of visual information into the narrow bandwidth of transmission channels or into a limited storage space, while maintaining the best possible perceived quality for the viewer this is called compression, compression is useful because it helps reduce the consumption of expensive resources, such as hard disk space or transmission bandwidth. The occurrence of the compression induced artifacts depends on the data source, target bit rate, and underlying compression scheme, and their visibility can range from imperceptible to very annoying, thus affecting perceived quality In the last decades, a considerable amount of research has been devoted to the development of a blockiness metric, which has been already implemented for the optimization of image quality. Another common distortion type, namely ringing, intrinsically results from loss in the high-frequency component of the video signal due to coarse quantization. In the video chain of a current television set, e.g., various video enhancement algorithms, such as deblocking, deringing, and deblur, are typically employed to reduce compression artifacts prior to display An efficient approach toward a no-reference ringing metric intrinsically exists of two steps: first detecting regions in an image where ringing might occur, and second quantifying the ringing annoyance in these regions. In this direction an efficient algorithm for automatic detection of regions visually impaired by ringing artifacts in compressed images is presented. The proposed system will be implemented in MATLAB for its realization.

Keywords —**Luminance masking, perceptual edge, ringing metrics, texture masking .**

1. INTRODUCTION

Until recently, only a limited amount of research was devoted to perceived ringing. The methods in and both simply assume that ringing occurs unconditionally in regions surrounding strong edges in an image. This, however, does not always reflect human visual perception of ringing, because of the absence of spatial masking as typically present in the HVS. This issue is taken into account by incorporating properties of the HVS into the detection method. The approach in is based on the global edge map of an image, where binary morphological operators are used to generate a mask to expose regions that are likely to be contaminated with visible ringing artifacts. This procedure involves the identification of regions around all detected edges, and a further evaluation of these regions based on visual masking. A different way of including HVS masking properties is employed. This method classifies the potential smooth regions (i.e., regions in an image other than edges and their surroundings) into different objects based on their color similarity and texture features. The resulting objects are assigned as background around potential ringing regions. Texture masking is implemented by evaluating the contrast in activity between the potential ringing region and its assigned background (e.g., the higher the contrast in activity, the more visible ringing is assumed to be). Additionally, also luminance masking is implemented to further determine ringing visibility. There are two main concerns with the methods existing in literature. First of all, the edge detection methods employed to capture strong edges using an ordinary edge detector, such as a Sobel operator, where a certain threshold is applied to the gradient magnitudes to remove noise and insignificant edges. Depending on the choice of the threshold, these methods run the risk of omitting obvious ringing regions near nondetected

edges (in case of a high threshold) or of increasing the computational power by modeling the HVS near irrelevant edges (in case of a low threshold). Fig. 3 illustrates the effect of the threshold value of a Sobel operator. The edge map in Fig. 3(c), resulting from a high threshold value, largely removes noisy edges while eliminating a number of important edges, at which ringing obviously exists [see Fig. 3(b)]. This may heavily degrade the accuracy of the prediction of perceived ringing. By lowering the threshold [as in Fig. 3(d)], all strong edges are maintained in the edge map, but it also contains more texture edges, which are nonrelevant to ringing detection, and consequently, result in a large number of unnecessary computations for ringing visibility. The second concern with the existing methods is related to the models of the HVS used, which are computationally very expensive. The HVS model involves a parameter estimation mechanism, which requires a number of calculations to achieve an optimal selection. The major cost of the HVS model is introduced by its clustering scheme embedded, which contains color clustering and texture clustering. Obviously, the optimal performance in terms of reducing the number of required computations, while maintaining the reliable detection of perceived ringing, can be achieved by optimizing two aspects: 1) the detection accuracy of relevant edges; and 2) the reduction in complexity of the HVS model itself. Hence, what is needed is an edge detector that only extracts edges most closely related to the occurrence of ringing, and a HVS model that is simpler (and thus more applicable for realtime implementation) than the approaches existing in literature. In this work, both aspects needed to efficiently detect regions with visible ringing are discussed. Our method mainly consists of two parts: 1) extraction of edges relevant for ringing, and 2) detection of visibility of ringing in the edge regions.

2. BACKGROUND

2.1. Perceived Ringing Artifacts

2.1.1 *Physical Structure*

Current image and video coding techniques are based on lossy data compression, which contains an inherent irreversible information loss. This loss is due to coarse quantization of the image's representation in the frequency domain. The loss within a certain spectral band of the signal in the transform domain reveals itself most prominently at those spatial locations where the contribution from this spectral band to the overall signal power is significant . Since the high-frequency components play a significant role in the representation of an edge, coarse quantization in this frequency range (i.e., truncation
of the high-frequency transform coefficients) consequently results in apparent irregularities around edges in the spatial domain, which are usually referred to as ringing artifacts. More specifically, ringing artifacts manifest themselves in the form of ripples or oscillations around high-contrast edges in compressed images. They can range from imperceptible to very annoying, depending on the data source, target bit rate, or underlying compression scheme .As an example, Fig. 1 illustrates ringing artifacts induced by JPEG compression on a natural image.

The occurrence of ringing spreads out to a finite region surrounding the edges, depending on the specific implementation of the coding technique. For example, in discrete cosine transform (DCT) coding ringing appears outwards from the edge up to the encompassing block's boundary . An

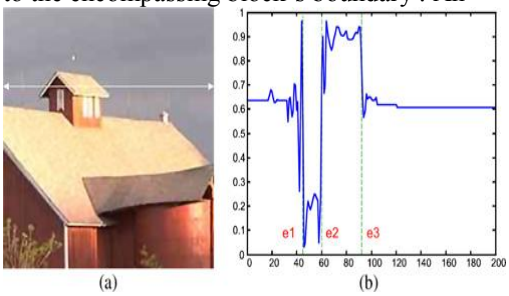


Fig. 1. Illustration of ringing artifacts.

(a) Natural image compressed with JPEG (MATLAB's imwrite function with "quality" of 30). (b) Gray-scale intensity profile along one row of the compressed image [indicated by the solid double arrowhead line in (a)]. Dashed lines "e1," "e2," and "e3" indicate the position of the sharp intensity transitions (i.e., edges) along that arrow. Ringing can be perceived as fluctuations in the gray-scale values around the edges at "e1," "e2," and "e3," while the image content here should be uniform.

example of how to calculate the extent of the ringing region in a particular codecs is given in . In addition to the edge location dependency, the behavior of ringing also depends on the strength of the edges. It is

found in and that, over a wide range of compression ratios, the variance of the ringing artifacts is proportional to the contrast of the associated edge. These important findings have great potential in the design of a reliable ringing metric, and therefore, are explicitly adopted in our algorithm.

2.1.2 Masking of the HVS

Taking into account the way the HVS perceives artifacts, while removing perceptual redundancies, can be greatly beneficial for matching objective artifact measurement to the human perception of artifacts. Masking designates the reduction in the visibility of one stimulus due to the simultaneous presence of another, and it is strongest when both stimuli have the same or similar frequency, orientation, and location . It is basically due to the limitations in sensitivity of a certain cell or neuron at the retina in relation to the activity of its surrounding cells and neurons. There are two fundamental visual masking effects highly relevant to the perception of ringing artifacts . The first one is luminance masking, which refers to the effect that the visibility of a distortion (such as ringing) is maximum for medium background intensity, and it is reduced when the distortion occurs against a very low or very high intensity background . This masking phenomenon happens because of the brightness sensitivity of the HVS, where the average brightness of the surrounding background alters the visibility threshold of a distortion . The second masking effect is texture masking, which refers to the observation that a distortion (such as ringing) is more visible in homogenous areas than in textured or detailed areas . In textured image regions, small variations in the texture are masked by the macro properties of genuine high-frequency details, and therefore, are not perceived by the HVS. The effect of luminance and texture masking on ringing artifacts is illustrated in Figs. 2 and 3, respectively.

2.2. Existing Ringing Metrics

Until recently, only a limited amount of research effort was devoted to the development of a ringing metric. Some of these

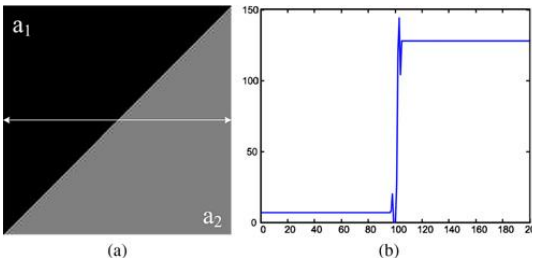


Fig. 2. Example of luminance masking on ringing artifacts.

(a) Image patch compressed with JPEG (MATLAB's imwrite function with "quality" of 30). (b) Pixel intensity profile along one row of the compressed image patch [indicated by the solid double arrowhead line in (a)]. Original image includes two adjacent parts with different gray-scale levels (i.e., 5 for "a1" and 127 for "a2"). Note that although both sides of a step edge exhibit ringing artifacts, the visibility of ringing differs.

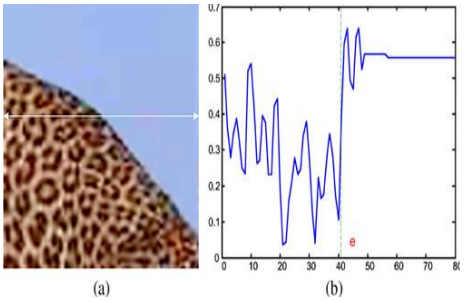


Fig. 3. Example of texture masking on ringing artifacts. (a) Image patch extracted from a JPEG compressed image of bit rate 0.59 bits per pixel (b/p). (b) Pixel intensity profile along one row of the compressed image patch [indicated by the solid double arrowhead line in (a)]. Dashed line "e" indicates the object boundary edge. Note that although both sides of the edge at "e" exhibit ringing artifacts, the visibility of ringing differs. metrics are FR, others NR. A FR approach presented in starts from finding important edges in the original image (noise and insignificant edges are removed by applying a threshold to the Sobel gradient image), and then measures ringing around each edge by calculating the difference between the processed image and the reference. Since this metric needs the original image, it has its limitations, e.g., for the application in a TV chain. The NR ringing metric, proposed in , performs a anisotropic diffusion on the image and measures the noise spectrum filtered out by the anisotropic diffusion process. The basic idea behind this metric is that due to the effectiveness of anisotropic diffusion on deringing, the artifacts would be

mostly assimilated into the spectrum of the filtered noise. The NR ringing metric described in indentifies the ringing regions around strong edges in the compressed image, and defines ringing as the ratio of the activity in middle low over middle high frequencies in these ringing regions. An obvious shortcoming of the metrics defined in and is the absence of masking, typically occurring in the HVS, with the consequence that these metrics do not always reflect perceived ringing. Typical masking characteristics, such as luminance and texture masking, are explicitly considered in the metrics defined in [and, in which ringing regions are no longer simply assumed to surround all strong edges in an image, but are determined by a model of the HVS. Including a HVS model in an objective metric might improve its accuracy, but often is computationally intensive for real-time applications. For example, the HVS model used in the metric presented in largely depends on a parameter estimation procedure, which requires a number of calculations to achieve an optimal selection. The model described in is based on a computationally heavy clustering scheme, including both color clustering and texture clustering. From a practical point of view, it is highly desirable to reduce the complexity of the HVS-based metric without compromising its overall performance.

The essential idea behind most of the existing metrics mentioned so far is that they consist of a two-step approach. The first step identifies the spatial location, where perceived ringing occurs, and the second step quantifies the visibility or annoyance of ringing in the detected regions. This approach intrinsically avoids the estimation of ringing in irrelevant regions in an image, thus making the quantification of ringing annoyance more reliable, and the calculation more efficient. Additionally, a local determination of the artifact metric provides a spatially varying quality degradation profile within an image, which is useful in, e.g., video chain optimization as mentioned in Section I. Since ringing occurs near sharp edges, where it is not visually masked by local texture or luminance, the detection of ringing regions largely relies on an edge detection method followed by a HVS model. Existing methods usually employ an ordinary edge detector, where a threshold is applied to the gradient image to capture strong edges. Depending on the choice of the threshold, this runs the risk of omitting obvious ringing regions near nondetected edges (e.g., in case of a high threshold) or of increasing the computational cost by modeling the rather complex HVS near irrelevant edges (e.g., in case of a low threshold). This implies that to ensure a reliable detection of perceived ringing while maintaining low complexity for real-time applications, an efficient approach for both detecting relevant edges and modeling the HVS is needed. Quantification of the annoyance of ringing in the detected areas can be easily achieved by calculating the signal difference between the ringing regions and their corresponding reference, as used in the FR approach described in. However, for a NR ringing metric, the quantification of ringing becomes more challenging mainly due to the lack of a reference. Metrics in literature estimate the visibility of ringing artifacts from the local variance in intensity around each pixel with in the detected ringing regions, and average these local variances over all ringing regions to obtain an overall annoyance score. This approach, however, has limited reliability, since it does not include background texture in the ringing regions, which might affect ringing visibility. To validate the performance of a ringing metric, its predicted quality degradation should be evaluated against subjectively perceived image quality. To prove whether a ringing metric is robust against different compression levels and different image content, the correlation between its objective predictions and subjective ringing ratings must be calculated. Unfortunately,

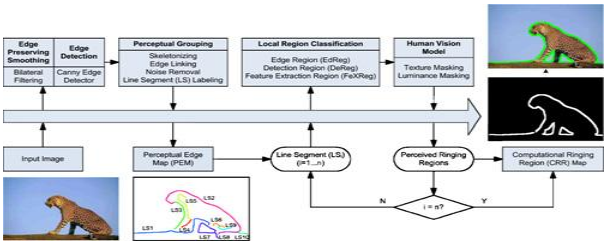


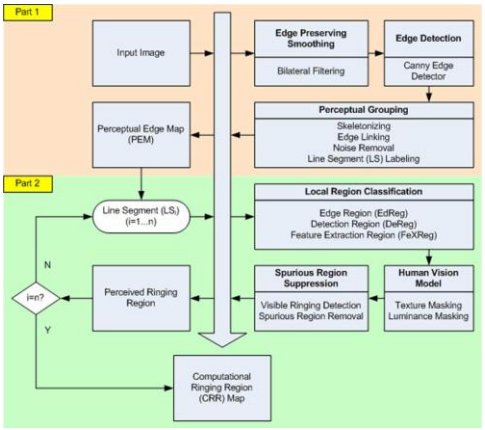
Fig. 4. Schematic overview of the proposed ringing region detection method.

In PEM, each perceptually relevant LS is labeled in a different color. In the CRR map, the white areas indicate the detected perceived ringing regions, and the spatial location of these regions is illustrated in a separate image by green areas. only the performance of the metric reported in is evaluated against subjective data of perceived ringing. For all other metrics nothing can be concluded with respect to their performance in predicting perceived ringing. Since we had no access to the data used in for our metric evaluation, we performed our own subjective experiment.1

In this paper, we propose a NR ringing metric based on the same two- step approach mentioned above. For the first step, we rely on our ringing

region detection method , the performance of which in terms of extracting regions with perceived ringing has been shown to be promising . Therefore, we consider this part of the metric readily applicable for the second step, in which the ringing annoyance is quantified. To quantify ringing annoyance, we consider each detected ringing region as a perceptual element, in which the local visibility of ringing artifacts is estimated. The contrast in activity between each ringing region and its corresponding background is calculated as the local annoyance score, which is then averaged over all ringing regions to yield an overall ringing annoyance score. It should be noted that the proposed metric is built upon the luminance component of images only in order to reduce the computational load. The performance of the NR metric is evaluated against subjective ringing annoyance in JPEG compression.

3. PROPOSED ALGORITHM



Our method mainly consists of two parts: 1) extraction of edges relevant for ringing, and 2) detection of visibility of ringing in the edge regions.

3.1 PERCEPTUAL EDGE EXTRACTION

3.1.1 Edge Preserving Smoothing and Canny Edge Detection

When interpreting the surrounding world, humans tend to respond to differences between homogeneous regions rather than to structure within these homogeneous regions. Hence, finding perceptually strong edges mainly implies that texture existing in homogenous regions can be neglected as if viewed from a long distance. This can be implemented by smoothing the image progressively until textual details are significantly reduced, and then applying an edge detector. Traditional low-pass linear filtering (e.g., Gaussian filtering) smoothens out noise and texture, but also blurs edges, and consequently, changes their spatial location. Since ringing detection intrinsically requires accurate spatial localization of the edges, edge-preserving smoothing is needed. Bilateral filtering was introduced in as a simple and fast scheme for edge-preserving smoothing. . The advantage of using bilateral filtering instead of Gaussian filtering for the localization specific detection of perceptually strong edges. Canny edge detector is applied to the bilaterally filtered image to obtain the perceptually more meaningful edges. Since the input image is already filtered, the subsequent Canny algorithm is implemented without its inherent smoothing step, while keeping the other processing steps unchanged. The Canny edge detector uses two thresholds to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. Their values is automatically set, depending on the image content .

3.1.2 Perceptual Edge Map Formation

Since the HVS does not perceive luminance variations at pixel level, the detected edge pixels are necessarily combined into perceptually salient elements, facilitating further analysis and processing. These perceptual elements, which we refer to as line segments (LS), are constructed over the Canny edge map and will be used as the basis for ringing region detection. The four steps are implemented to define the LS in the PEM.

Skeletonizing, Edge Linking , Noise Removal and Line Segment Labeling.

3.2. Ringing Region Detection

Each LS of the PEM is examined individually on the occurrence of visible ringing artifacts in their direct neighborhood, taking into account luminance and texture masking.

3.2.1 Local Region Classification

In order to characterize the visibility of ringing around a LS, its surrounding is classified into three different zones. Edge Region , Detection Region and Feature Extraction Region. These regions are defined by thickening the LS with a different size for the structuring element of a dilation operation.

3.2.1.1 Human vision model

Whether ringing is actually visible in the DeReg strongly depends (because of masking in the HVS) on the content of the original background, here represented by the FeXReg. Hence, the visibility of ringing is evaluated for each LS by applying a model for texture and luminance masking, using the texture and luminance characteristics of the FeXReg. As a

result, DeReg regions, in which ringing is visually masked are eliminated, and only the perceptually prominent DeReg ringing regions remain.

3.2.1.2 Texture Masking

The visibility of ringing is significantly affected by the spatial activity in the local background, i.e. ringing is masked when located in a textured region, while it is most visible against a smooth background . Texture masking is modeled classifying the FeXReg of each detected edge segment into "smooth" and "textured" parts. The DeReg is segmented accordingly, and only the regions of which the corresponding FeXReg is clustered as "smooth" are retained. removing the corresponding texture regions in DeReg. Hence, the remaining regions of DeReg are only smooth regions around the detected strong edges.

3.2.1.3 Luminance Masking

The visibility of variations in luminance depends on the local mean luminance. As a result, the visibility of ringing is largely reduced in extremely dark or bright surroundings. The implementation of luminance masking is the same as for texture masking, but to guarantee efficiency, it is only applied to those regions of the DeReg remaining after the application of texture masking. Classifying the "smooth objects" of FeXReg further into "visible objects" and "invisible objects" depending on the invisible components. Removing the DeReg that correspond to "invisible objects," i.e., where ringing is not supposed to be visible against a very low or very high intensity background. Ultimately, only the regions of DeReg that yield visible ringing remain.

3.2.1.4 Spurious Ringing Region Suppression

The ringing region detection method described so far only exposes regions in an image which are likely to be impaired corresponding to (b) a JPEG compressed image.

by visible ringing artifacts. The resulting CRR map, however, still includes obvious spurious ringing regions, containing either "unimpaired" or "noisy" pixels misinterpreted as ringing pixels.

The occurrence of "unimpaired" pixels is in an uncompressed image. The ringing region detection algorithm described so far will find the regions that might be impaired with visible ringing, independent of the compression level. But in an uncompressed image, these regions do not contain visible ringing, and hence, should be removed from the CRR map. Note that without removal of these regions the overall objective ringing metric including the step of quantification of ringing annoyance would not be less accurate, but less efficient.

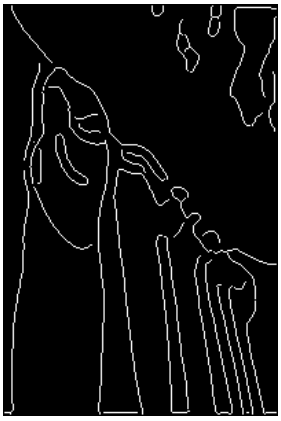
"Noisy pixels" are pixels in the detected regions of the CRR map, that actually belong to an edge or texture. They are accidentally misclassified to a ringing region as a consequence of the dilation operation used in the human vision model. To remove the spurious ringing regions, each detected ringing region (RR) is further examined by calculating its amount of visible ringing pixels. Those RRs with their number of visible ringing pixels below a certain threshold are considered as spurious, and consequently removed from the CRR map

4. RESULTS AND OBSERVATIONS

Original Image

Bilateral Filtered Image

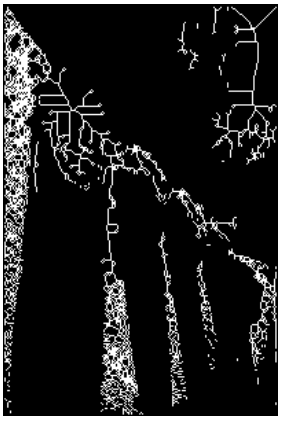
Bilateral - Canny Edge



Noise Removal

Skeletonizing

Edge Linking



5. CONCLUSION

A novel approach toward the detection of perceived ringing regions in compressed images is presented. The algorithm relies on the compressed image only, which is promising for its applicability in a real-time video chain, e.g., to enhance the quality of artifact impaired video. It adopts a perceptually more meaningful edge detection method for the purpose of ringing region location. This intrinsically avoids the drawback of applying an ordinary edge detector, which has the risk of omitting obvious ringing artifacts near non detected edges or of increasing the computational cost by measuring ringing visibility near irrelevant edges. The objective detection in agreement with human visual perception of ringing artifacts is ensured by taking into account typical properties of the human visual system, such as texture masking and luminance masking. The human vision model is implemented, based on the local image characteristics around detected edges, to expose only the perceptually prominent ringing regions in an image. The proposed detection method is validated with respect to ringing regions resulting from a psychovisual experiment, and shows to be highly consistent with subjective data. The proposed ringing region detection method is meanwhile extended with a ringing annoyance metric that can quantify perceived ringing annoyance of compressed images.

6. FUTURESCOPE

In most visual surveillance systems, stationary cameras are typically used. However, because of inherent changes in the background itself, such as fluctuations in monitors and fluorescent lights, waving flags and trees, water surfaces, etc. the background of the video may not be completely stationary. In these types of backgrounds, referred to as quasi-stationary, a single background frame is not useful to detect moving regions.

Detecting regions of interest in video sequences is one of the most important tasks in many high level video-processing applications. In the future scope of this work we have to design a system, which detects foreground regions in videos with quasi-stationary backgrounds. The main contribution should be the novelty detection approach, which automatically segments video frames into background/foreground regions. By using support vector data description for each pixel, the decision boundary for the background class is modeled without the need to statistically model its probability density function. The proposed method is able to achieve very accurate foreground region detection rates even in very low contrast video sequences, and in the presence of quasi-stationary backgrounds.

7. REFERENCES

[1] Z. Wang and A. C. Bovik, Modern Image Quality Assessment. Synthesis
Lectures on Image. Video & Multimedia Processing. San Rafael, CA: Morgan and Claypool, 2006.

[2] M. Ghanbari, Standard Codecs: Image Compression to Advanced Video Coding. London, U.K.: IEE Press, 2003.

[3] I. Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia. Hoboken, NJ: Wiley , 2003.

[4] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," Signal Process., vol. 70, no. 3, pp. 247–278, 1998.

[5] M. C. Q. Farias, M. S. Moore, J. M. Foley, and S. K. Mitra, "Perceptual contributions of blocking, blurring, and fuzzy impairments to overall annoyance," in Proc. SPIE, Human Vision and Electron. Imaging IX, 2004, vol. 5292, pp. 109–120.

[6] C. C. Koh, S. K. Mitra, J. M. Foley, and I. Heynderickx, "Annoyance of individual artifacts in mpeg-2 compressed video and their relation to overall annoyance," in Proc. SPIE, Human Vision and Electronic Imaging X, 2005, vol. 5666, pp. 595–606.

[7] J. Xia, Y. Shi, K. Teunissen, and I. Heynderickx, "Perceivable artifacts in compressed video and their relation to video quality," Signal Process . Image Commun., vol. 24, no. 7, pp. 548–556, 2009.

[8] M. Shen and C. J. Kuo, "Review of postprocessing techniques for compression artifact removal," J. Vis. Commun. Image Rep., vol. 9, no. 1,
pp. 2–14, 1998.

[9] J. Luo, C. Chen, K. Parker, and T. S. Huang, "Artifact reduction in low bit rate dct-based image compression," IEEE Trans. Image Processing, vol. 5, pp. 1363–1368, 1996.

[10] K. Zon and W. Ali, "Automated video chain optimization," IEEE Trans. Consum. Electron., vol. 47, pp. 593–603, 2001.

[11] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in

images," in Proc. IEEE Int. Conf. Image Processing, 2000, vol. 3, pp. 981–984.
[12] H. Liu and I. Heynderickx, "A perceptually relevant no-reference blockiness metric based on local image characteristics," EURASIP J. Adv. Signal Process., vol. 2009, 2009.
[13] Z. Wang and A. C. Bovik, Modern Image Quality Assessment (Synthesis Lectures on Image, Video and Multimedia Processing). San Rafael, CA: Morgan and Claypool, 2006.
[14] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process., vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.
[16] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006.
[17] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," IEEE Trans. Image Process., vol. 14, no. 11, pp. 1918–1927, Dec. 2005.
[18] R. V. Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," Signal Process., vol. 87, no. 6, pp. 1493– 1503, 2007.
[19] P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," J Electron. Imaging, vol. 14, no. 3, p. 033004, 2005.
[20] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," IEEE Signal Process. Lett., vol. 4, no. 11, pp. 317– 320, Nov. 1997.
[21] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, Apr. 2010, to be published.

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue.2, pp-270-277 ISSN: 2249-6645

A Statistical Study on Optimal Usage of Intelligent Character Recognition Technology to Fetch Quality Data for Automation of University Exam System

Vikas Sharma

International Centre for Distance Education and Open Learning,
Himachal Pradesh University, Summer Hill, Shimla, India

ABSTRACT :

The automation of examination system for a typical university has been overwhelmed with various data quality problems because of involvement of manual data entry process from candidates' filled examination forms to evaluators' handwritten award lists, importing or exporting of data, limited resources, etc. Poor data quality can have a significant negative impact on organizations' success especially for a university where its entire credibility is dependent on accuracy and timely processing of results. As a result, organizations are implementing latest technologies to fetch quality data to achieve competitive advantage as well as to satisfy the varying needs of users. In this paper, a case study of Intelligent Character Recognition (ICR) system "AutoRec" is presented with regard to different data quality parameters such as accuracy, time, value added services, security, timeliness, etc. by processing a good sample of manual handwritten awards lists involving both numeric as well as alphanumeric characters. The results of the study indicate that the ICR based "AutoRec" system has the potential solution to improve data quality, minimize human intervention, reduce cost and time by balanced usage of scanning parameters, validation checks and confidence levels. Further, ICR is not the substitute of human operator but can minimize manual intervention.

Keywords - Data Quality, ICR, Scanning Parameters, Validation Checks, Confidence Levels, Human Intervention

1. INTRODUCTION

Information is increasingly becoming a critical asset for success in the modern societies throughout the world. Information is being created, processed, stored and retrieved and transmitted instantaneously from one end to another but the basic question is "how much this information is fit for use?" The data Quality (DQ) is one of the key determinants that decide the success or failure of any organisation. Errors in data cause variety of problems and raise costs in several areas. Earlier an error is detected, the cheaper it is to correct [1]. A typical university examination system consists of large data volumes, heterogeneous data types, widely distributed data sources and multiple stakeholders. The very existence of any university can be threatened by poor data quality (DQ). The data on which examination results are based and upon which the future of thousands of students depends if inaccurate, incomplete or has other types of problems can put a big question mark on the credibility. The automation of such examination system demands high data quality management system in place to avoid garbage-in-garbage out. The examination wing of Himachal Pradesh University, Summer Hill, Shimla compiles the results of the students by performing manual data entry of awards using a small software (utility) "Awards Management System" but this system doesn't provide quality data for processing. Further, the manual data entry is a slow, laborious, and expensive process compared to automatic recognition of text and its subsequent processing. Data entry specialists are becoming increasingly difficult to employ since it is tedious and boring work [5]. Computer has the ability to perform numerous tasks simultaneously and efficiently on scanned images by recognizing characters using artificial intelligence power. The most significant among these technologies is the Intelligent Character Recognition (ICR) for hand written documents and Optical Mark Recognition (OCR) for data capture from printed documents. The ICR is seemingly a good technology to fetch data from real world and convert into computer readable form. In this paper, a case study of newly introduced ICR system "AutoRec" viz.-a-viz. manual data entry system is conduced in the examination wing of Himachal Pradesh University to know which system can provide better quality of input data for processing by involving minimal human intervention, cost and time.

1. NEED OF INTELLIGENT CHARACTER RECOGNITION SYSTEM

Paper forms are still the least expensive data capture device where individuals without network connections must provide data for entry into a computer system. Even today's much-acclaimed Internet browser interfaces do not help computer applications when the data must be collected from

www.ijmer.com Vol.1, Issue.2, pp-270-277 ISSN: 2249-6645

constantly changing individuals and constantly changing locations [5]. The forms are easy to use even today due to little initial cost per form. No doubt that the documents are increasingly originated on the computer, however, in spite of this, it is unclear whether the computer has decreased or increased the amount of paper. Documents are still printed out for reading, dissemination, and markup [4]. Colleges and universities throughout the country are struggling to find some way to deal with paper documents that must be maintained to ensure institutional accountability. The improvement in hardware and increasing use of computers for storing paper documents has paved the way for document processing and recognition. The cost of optical scanners for document input have dropped to the level that these are affordable to even small businesses and individuals. In addition to above, the advancements in document analysis software and algorithms have also improved the text and image recognition rates significantly up to the level of 90 to 95% [4].

2. ISSUES RELATED TO DATA QUALITY AND COST USING INTELLIGENT CHARACTER RECOGNITION SYSTEM

Errors in data cause a variety of problems and raise the costs in several other associated areas. The cost to recognize and detect errors is not small whereas significant amount is involved to correct these data errors. The largest cost components are the hidden costs that affect the efficiency, productivity and public image of the organisation [1]. To maintain data quality has become the essential task for universities who have to compile huge volume of students data related with admission and examination processes. A small mistake in students' result status due to consideration of poor quality data can put the credibility of whole university under scanner which further also lowers the public image as well as involvement of litigation costs, etc. Since data quality is the major issue for data processing jobs so it become essential to design and tune character recognition applications to achieve high data quality. There are two types of recognition errors in ICR system: 1) rejected errors -unrecognized characters, and 2) substitution errors- erroneously recognized characters. Rejects need to be corrected by human intervention but substitutions must first be detected and then corrected [1]. Document scanners can misread an image that is dirty or too skewed. Characters read without contextual analysis may be interpreted as letters, when only numbers should exist in a field [5]. Substitution errors are the most dangerous because during these errors, an incorrect character is substituted for the correct character [3]. An image of the rejected character is presented to the data entry operator who corrects it by re-entering the actual character and program control automatically moves to the next rejected character. The rejected character cost is dependent on three factors: 1) missing characters- unreadable character, 2) extraneous characters, and 3) key entry speed of operator [1]. So, in nutshell, there are two factors which influence the cost to repair reject characters: 1) accuracy of the ICR recognition engine, and 2) reject re-entry speed. An increase in reject re-entry rate also leads to substantial error cost reduction. This is because the data entry operator can often key the entire field faster than repairing several rejects [1].

3. RESEARCH OBJECTIVES

- To study the opinions of technical staff about quality of input data essential for compilation of examination results.
- To analyse and compare quality dimensions of input data fetched through ICR based system "AutoRec" viz.-a-viz. manual data entry system "Awards Management System".
- To study the effect of image quality, confidence levels and validation checks on character recognition level of ICR based system "AutoRec".
- To analyse the effect of image quality, confidence levels and validation checks on cost, time and human intervention involved in ICR based system "AutoRec".

4. RESEARCH DESIGN AND METHODOLOGY

The research methodology of this study has been divided into three main parts, namely: 1) Scope of Study, 2) Population and Sample, and 3) Research Tools.

5.1 SCOPE OF STUDY

This study is conducted in the Examination Wing of Himachal Pradesh University and specifically on two examination classes - B.Com. and B.Sc. whose results are complied using computers.

5.2 POPULATION AND SAMPLE

To study the first two objectives, this study is based on convenient sample survey of nine technical staff members-four programmers and five data entry operators who actually use the ICR based "AutoRec" system- a product of FilFlan Technologies as well as in-house developed Awards Management System (AMS) based on Visual Basic 6.0 as front end and MS-Access 2000 as back end to fetch input data for compilation of results. To study the last two objectives, five samples of ICR compatible awards lists 2 from B. Com. part-III and 3 from B. Sc. Part -III, regular examinations, March 2010 were selected again using convenient sampling technique to know the effect of scanning parameters, validation checks and confidence levels on character recognition level of "AutoRec" system and involvement of human intervention, cost, time, etc. The

www.ijmer.com Vol.1, Issue.2, pp-270-277 ISSN: 2249-6645

above sample had 208 data fields which in turn summed to recognition of 761 characters. The opinion of ICR system "AutoRec" was classified using three point Likert Scale – matched characters, unrecognised characters and substituted characters.

5.3 RESEARCH TOOL

To study the first two objectives, the data collection tool was self designed questionnaire having two parts: 1) first part was used to rate the importance of input data quality parameters in context of University Examination System, and 2) second part was used to observe the quality dimensions of input data fetched through manual data entry system "AMS" as well as ICR based "AutoRec" system separately. A 5-point Likert Scale (5 = highly important and 1 = not important at all) was used to observe the opinion of the dealing persons corresponding to each quality dimension. To build an initial list of data qualities, the fifteen data quality dimensions defined [6] as: 1) access security, 2) accessibility, 3) accuracy, 4) appropriate amount of data, 5) believability, 6) completeness, 7) concise representation, 8) ease of understanding, 9) ease of understanding, 10) interpretability, 11) objectivity, 12) relevancy, 13) representational consistency, 14) reputation, and 15) timeliness were discussed in detail followed by brainstorming sessions to conclude a raw list relevant in context of compilation of results. These items were arranged then in logical order to give a questionnaire format. Using the literature on information/data quality and by looking carefully for overlap of data qualities in context of examination system, the items in the questionnaire were reduced to a more manageable 9 items with small description to provide readily available detail for observers while completing their questionnaires. The data collection process was carried out firstly by using the first part of the questionnaire followed by second part of questionnaire. Further to study the third objective, the following methods were used:

To analyse the effect of image quality on character recognition level of "AutoRec" system, four cases (T1C1, T1C2, T2C1 and T2C2) were designed using different scanning parameters where T1 (128 units), T2 (184 Units), C1 (128 units) and C2 (144 units) are default and best threshold 'T' and contrast- 'C' values on Fujitsu Scanner Fi4340C. The 128 units is the default value for threshold and contrast whereas for best visibility 184 units is the threshold value and for best sharpness the 144 units is the contrast value. Further, the red colour was dropped during scanning and the size of award lists used here was of legal (8.5"X14") size.

To study the effect of data validations checks on recognition accuracy of ICR system, two types of validation checks namely: 1) NVC (No Validation Checks), and 2) ONC (Only Numeric Checks) were applied on all above four cases (T1C1, T1C2, T2C1 and T2C2) separately.

- To study the recognition accuracy of ICR system at different confidence levels, four confidence levels (50, 75, 90 and 100 units) were experimented separately.
- To study the fourth objective, the response of ICR system based on three points Likert Scale (matched, unrecognised and substituted characters) was divided into two segments- characters needed human intervention (unrecognised and substituted characters) and characters needed no human intervention (matched characters). Total numbers of character needed human interventions were compared with actual number of characters which needed human intervention (manual data entry) to analyse the cost and time involved in both systems. In manual data entry system, double entry of every single award is done to have good accuracy level and to avoid any kind of discrepancy. The above observed opinions of the ICR system then converted into appropriate data tables and different statistical techniques were applied for analysis using MS-Excel 2007 spreadsheet.

5. RESULTS AND DISCUSSION

The table 1 shows the summary averages for weighted and unweighted data sets. Firstly, the importance score (IS) shows the average importance ranking for each question as rated by technical staff. Secondly, the average scores per data item for ICR based "AutoRec" system and manual data entry system (MDES) "Awards Management System (AMS)" are given. This is displayed in two modes: 1) raw score (RS) as unweighted ratings (with a theoretical Likert Scale range of 1 to 5), and 2) weighted score (WS). The weighted score (theoretically ranging from 1 to 25) is obtained by multiplying the unweighted score by the importance score for each respondent. The data quality considered most important by the technical staff, e.g. upper quartile (5.00) are all about accuracy, representational consistency and access security. The data qualities considered least important, e.g. below lower quartile (4.8) is value added features. Other quality dimensions are in between and the median for above importance score is 4.9.

The above total scores make it difficult to analyse the quality of input data fetched through above two system, so Data Quality Index (DQI) was calculated for each system using weighted score against the total possible score (Importance Score X 5). Overall, it appears that input data fetched through "AutoRec" system scored (0.9 points) followed by Manual Data Entry System (0.7 points). The Data Quality Index (DQI) for different input data quality dimensions is shown in fig. 1 for "AutoRec" system as well as Manual Data Entry System. Though there is no much

www.ijmer.com Vol.1, Issue.2, pp-270-277 ISSN: 2249-6645

more difference on some data quality dimensions such as appropriate amount of data, completeness, ease of understanding, representational consistency, value-added features and accessibility but a clear circle around by "AutoRec" system over Manual Data Entry System on other data quality dimensions such as accuracy, timeliness and access security indicates the difference.

TABLE I
Summary Averages for Weighted and Unweighted Data Sets

Sr. No.	Quality Dimensions	IS	SD	WIS	ICR "AutoRec"			MDES "AMS"		
					RS	WS	DQI	RS	WS	DQI
1	**Accuracy** Fetched Data is correct, i.e. free of errors.	5.0	0.0	25.0	4.2	21.1	0.84	2.9	14.5	0.58
2	**Appropriate Amount of Data** The quantity or volume of obtained data is appropriate for compilation of results	4.9	0.3	24.5	4.4	22.2	0.91	4.4	22.2	0.91
3	**Completeness** Fetched data is sufficient for compilation of results.	4.9	0.3	24.5	4.3	21.7	0.89	4.2	21.1	0.86
4	**Ease of Understanding** Obtained data is clear, without ambiguity and easy to comprehend.	4.8	0.4	23.9	4.6	22.8	0.95	4.4	22.2	0.93
5	**Timeliness** Time taken to convert manual data into digital form is reasonably good.	4.8	0.4	23.9	4.2	21.1	0.88	1.9	9.5	0.40
6	**Representational Consistency** Fetched data is represented in the specified format and compatible with previous data.	5.0	0.0	25.0	4.7	23.4	0.93	4.6	22.8	0.91
7	**Accessibility** Obtained data is available for usage easily and quickly.	4.9	0.3	24.5	4.1	20.6	0.84	3.6	17.8	0.73
8	**Value-Added** Obtained data is beneficial and provide advantages for value added services.	4.6	0.5	22.8	2.9	14.5	0.63	2.7	13.4	0.59
9	**Access Security** Access to fetched data is restricted and hence kept secure.	5.0	0.0	25.0	4.8	23.9	0.96	3.4	17.2	0.69
	Total	43.8	--	219.0	38.2	191.1	0.90	32.1	160.6	0.70

IS-Importance Score, SD- Standard Deviation, WIS- Weighted Importance Score, RS-Raw Score, WS-Weighted Score, ICR-Intelligent Character Recognition, MDES-Manual Data Entry System, DQI-Data Quality Index

www.ijmer.com Vol.1, Issue.2, pp-270-277 ISSN: 2249-6645

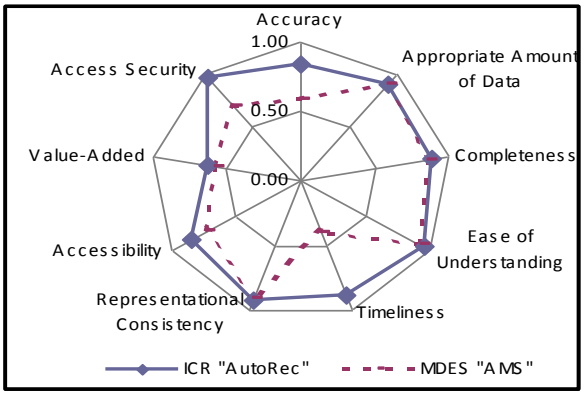


Fig 1. Input data quality dimensions or ICR and manual data entry system

5.1 EFFECT OF SCANNING PARAMETERS ON CHARACTER RECOGNITION, HUMAN INTERVENTION, COST AND TIME INVOLVED

The major difficulties of character recognition have to do with locating and correcting errors. The quality of a document's appearance is critical. A paper document marred by smudges, fingerprints, dot-matrix print, or fuzziness can be nearly as disastrous as skewed placement or a dirty scanner glass. Other errors are caused by coloured inks or papers, outsized or otherwise unrecognized fonts, etc. [2].

A good quality document with well-delineated text is the first requirement to get accuracy. Paper colour and type also affect the quality of the scanning and resulting image [1]. It was observed that different scanning parameters (threshold and contrast values) affected the recognition accuracy of ICR system. Higher character recognition accuracy rate (89.56 percent) was observed using 184 units threshold and 128 units contrast value whereas low character recognition rate (86.29 percent) was obtained using 128 units threshold and 144 units contrast value. The fig. 2 shows the effect of scanning quality on recognition accuracy level of "AutoRec" system.

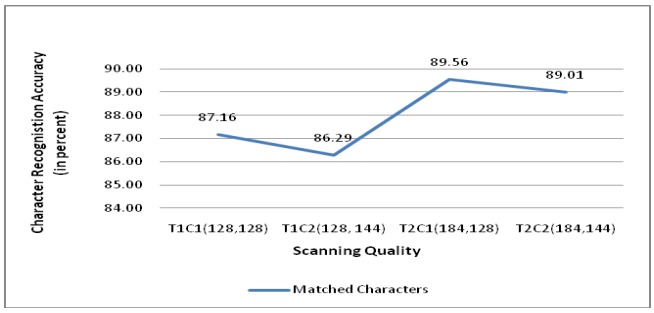


FIG.2. Effect of scanning quality on recognition accuracy level of "AutoRec" System

The Fig 3 shows the effect of scanning parameters on different errors types using "AutoRec" system. The highest errors rate (10.94 percent) for unrecognized characters is for the T1C2 images and lowest rate (7.21 percent) for T2C1 images. Similarly, highest substitution character error rate (3.55 percent) for the T1C1 images whereas lowest substitution character error rate (2.51 percent) for T2C2 images. This indicates that good combination of threshold and contrast values are required to enhance character recognition level of "AutoRec" system.

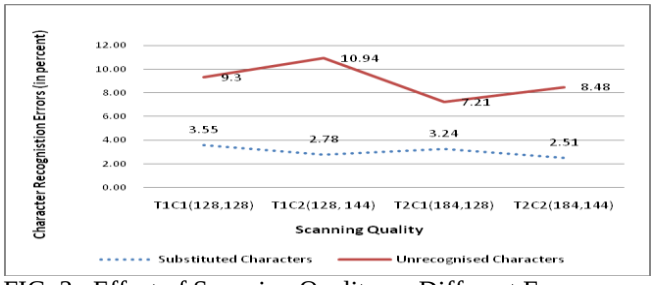


FIG. 3 Effect of Scanning Quality on Different Errors Types

The human intervention is manual efforts required at the end of the computer operator to make each individual character understandable to the computer system where "AutoRec" system is not able to recognise characters. Overall human intervention per character using "AutoRec" system was 6.0 percent whereas cost was just 0.06 units. This means for 100 characters to be entered by an operator manually, only 6 characters needs human intervention using "AutoRec" system. Similarly, if cost for manually entered 100 characters is 100 units then using "AutoRec" the cost for the same number of characters would be just 6 units. As for as promptness of "AutoRec" system is concerned, it was able to fetch 17 characters at a given time as compared to one character entered manually. It was also observed that different scanning parameters also affected the human intervention involved in "AutoRec" system. The maximum human intervention (6.86 percent) was involved for the T1C2 scanned images and minimum human intervention (5.22 percent) for T2C1 scanned images. It is concluded that the ICR based "AutoRec" system provides better performance as compared to manual data entry system in terms of time, cost and involvement of minimum human

intervention. The table 2 shows the performance of "AutoRec" system over manual data entry system using different scanning parameters.

TABLE 2
Performance of "AutoRec" System over Manual Data Entry System using Different Scanning Parameters

Images	TC	SC	UC	MC	HI (in % age)	Prm.	C
T1C1	6088	216	566	5306	782 (6.42)	15.57	0.06
T1C2	6088	169	666	5253	835 (6.86)	14.58	0.07
T2C1	6088	197	439	5452	636 (5.22)	19.14	0.05
T2C2	6088	153	516	5419	669 (5.49)	18.2	0.05
Total	24352	735	2187	21430	5.99	16.87	0.06

TC- Total Characters, **SC**-Substituted Characters, **UC**-Unrecognised Characters, **MC**-Matched Characters, **HI**- Human Intervention, **Prm**-Promptness**, C**-Cost
* Manual Data Entry involves double data entry of awards

5.2 Effect of Data Validation Checks on Character Recognition, Human Intervention, Cost and Time Involved

Data validation checks affect the recognition accuracy of ICR system. Appropriate use of various data validation checks can provide high data throughput and able to minimise human intervention. The fig. 4 shows the effect of data validation checks on recognition accuracy level of ICR system.

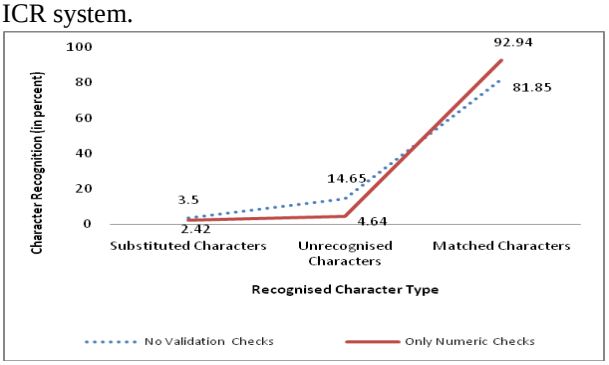


FIG. 4 Effect of data validation checks on recognition accuracy level of ICR system

The "AutoRec" system uses its own intelligence power within the domain of validation checks. It was observed that recognition accuracy level of 92.94 percent achieved using Only Numeric Checks (ONC) on award lists followed by 81.85 percent by applying No Validation Checks (NVC). Further by applying NVC, the substituted error rate and unrecognised character rate were 3.5 percent and 14.65 percent respectively whereas these were 2.42 percent and 4.64 percent by applying ONC. So, it is concluded that specific validation checks must be applied to minimise human intervention and to enhance overall accuracy of "AutoRec" system. The table 3 shows the performance of "AutoRec" system over manual data entry system using different data validation checks.

TABLE 3
Performance of "AutoRec" System over Manual Data Entry System using Different Data Validation Checks

VC	TC	SC	UC	MC	HI (in % age)	Prm.	C
NVC	12176	426	1784	9966	2210 (9.08)	11.02	0.09
ONC	12176	294	565	11317	859 (3.53)	28.35	0.04
Total	24352	720	2349	21283	3069 (6.31)	15.87	0.06

VC-Validation Checks, **TC**- Total Characters, SC-Substituted Characters, **UC**- Unrecognised Characters, **MC**-Matched Characters, **HI**- Human Intervention, **Prm**-Promptness, **C**-Cost, NVC-No Validation Checks, ONC-Only Numeric Checks
* Manual Data Entry involves double data entry of awards

Further, It was observed that the human intervention of an operator reduced by applying specific validation checks. Maximum human intervention (9.08 percent) was observed using NVC whereas minimum human intervention (3.53 percent) by applying ONC. Overall human intervention per character using "AutoRec" system was 6.31 percent whereas character recognition cost per character was just 0.06 units as compared to manually entered character. The maximum character recognition cost for ICR system was 0.09 units by applying NVC whereas minimum character recognition cost is 0.04 units on applying ONC. Further, the ICR system is 15.87 times faster to recognise characters as compared to similar number of characters punched by an operator manually in which maximum character recognition promptness (28.35 characters) using ONC whereas minimum character recognition promptness (11.02 characters) using NVC. This further indicates that "AutoRec" system has better performance as compared to manually data entry system in terms of time, cost and involvement of human intervention.

5.3 Effect of Confidence Levels on Recognition Accuracy Level, Human Intervention, Cost and Time Involved

An ICR recognition engine assigns a specific confidence value for every character to be recognized. Confidence

thresholds may be modified within the software for certain fields or characters (Phillips, 2000). The confidence level of ICR system also affects the accuracy level as well as human intervention. At confidence level 50 units, the recognition accuracy rate was 87.75 percent and error rate was 12.25 percent whereas at confidence level 75 units, the recognition accuracy rate was highest (88.14 percent) and errors rate was minimum (11.86 percent). Further, as confidence level of "AutoRec" increases above 75 units, a continuous decrease in recognition accuracy level was witnessed. The recognition accuracy rate was 88.09 percent for confidence level 90 units whereas 88.02 percent for confidence level 100 units. The fig. 5 shows the effect of confidence levels on character recognition level accuracy of "AutoRec" system.

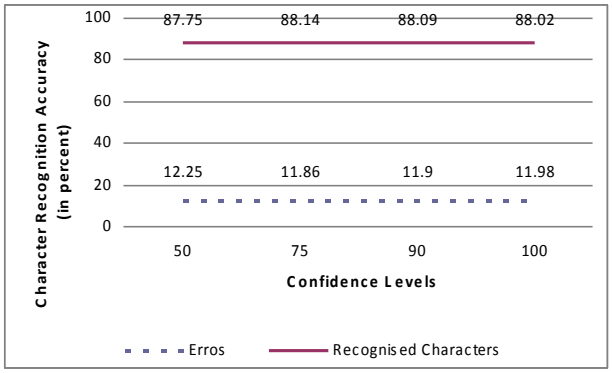


Fig. 5 Effect of confidence levels on recognition accuracy level of "AutoRec" system.

This indicates that there is a need to choose the confidence level of appropriate level to get high accuracy and minimize human intervention. It is also observed that after reaching a certain threshold limit of confidence level, there is increase in specific types of errors such as substitutional characters and these types of errors are not only very hard to detect but also very costly to correct. The fig. 6 shows the effect of confidence levels on different character recognition errors:

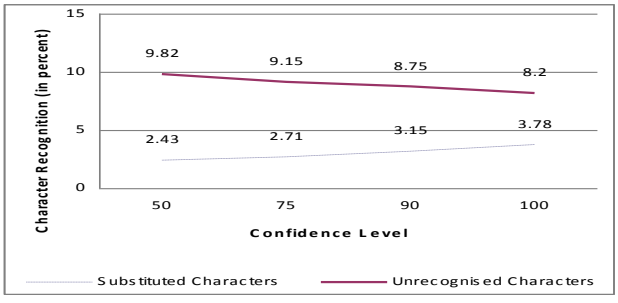


Fig. 6 Effect of confidence levels on different character recognition errors

Maximum human intervention (6.13 percent) was involved at confidence level 50 units and minimum human intervention (5.93 percent) at confidence level 75 units. The table 4 shows the performance of "AutoRec" system over manual data entry system using different confidence levels.

TABLE 4
Performance of ICR System over Manual Data Entry System using Different Confidence Levels

Confd Level	TC	SC	UC	MC	HI (in % age)	Prm.	C
50	6088	148	598	5342	746 (6.13)	16.32	0.06
75	6088	165	557	5366	722 (5.93)	16.86	0.06
90	6088	192	533	5363	725 (5.95)	16.79	0.06
100	6088	230	499	5359	729 (5.99)	16.7	0.06
Total	24352	735	2187	21430	2922 (6.00)	16.66	0.06

Confd. Level- Confidence Levels, **TC**- Total Characters, **SC**-Substituted Characters, **UC**-Unrecognised Characters, **MC**-Matched Characters, **HI**- Human Intervention, **Prm**.-Promptness, **C**-Cost
* Manual Data Entry involves double data entry of awards

The overall human intervention per character using ICR System was 6.0 percent whereas character recognition cost was just 0.06 units as compared to manually data entry system. Further, the overall character recognition promptness of ICR system was 16.66 characters as compared to one character punched by an operator manually in which maximum character recognition promptness (16.86 characters) at confidence value 75 units and minimum character recognition promptness (16.32 characters) at confidence value 50 units.

6. CONCLUSIONS

The results of the above study show that ICR technology has the potential to maintain data quality, minimise manual data entry load and increase overall productivity & efficiency of university examination system where limited human manpower, time and cost are the major constraints to process huge volume of data using papers. But to make effective utilisation of this technology, there is a need to take care of certain factors which affect the recognition accuracy level of ICR system such as quality of scanned image, use of data validation checks and confidence levels. The balanced usage of data validation checks and confidence levels do not only facilitate in minimisation of human intervention, reduction in cost and time but also increases the overall data quality. ICR system functions on individual character basis and not on entire data field so a single false recognition or substitutional error has a very

high probability to corrupt the whole record and which in turn provides poor data quality. Further, ICR system is not the substitute of human operator but an aid to minimize manual intervention for conversion of data available on papers into computer readable form. Based on the study presented in this paper, it is recommended that usage of ICR based technology "AutoRec" can be extended to other areas where huge volume of paper work is involved such as for admission, examination, settlement of result discrepancies, re-evaluation cases, etc. In addition to above, the usage of ICR technology can be used to create data centres for universities where paper is still the dominant media for exchange of information.

7. REFERENCES

[1] (2005). "The Importance of Power/Precision Data Entry to Document Imaging". [Online]. Available: http://www.vikingsoft.com/pdf/importanceofppde.pdf

[2] Arrington, Daniel V. (1992). "Departmental Document Imaging: Issues and Concerns." [Online]. Available: http://cool-palimpsest.stanford.edu/bytopic/imaging/depimgng.html.

[3] Gingrande, Arthur. "Forms Automation: From ICR to E-Forms to the Internet". *Silver Spring, MD: AIIM International*.1998.

[4] O'Gorman Lawrence and Rangachar Kasturi, "What is a Document Image and What Do We Do With It." *Document Image Analysis. IEEE Computer Society Executive Briefings*. [Online]. Available: http://www.ijcaonline.org/volume10/number5/pxc3871991.pdf

[5] Phillips, JOHN T (2000). "Does ICR Keep Paper Forms Viable?" [Online]. Available: https://www.entrepreneur.com/tradejournals/article/62194277.html

[6] Wang, Richard, Diane Strong, and Lisa Guarascio. "An empirical investigation of data quality dimensions: A data consumer's perspective." Working paper *TDQM-94-01, MIT TDQM Research Program, E53-320, 50 Memorial Drive*, Cambridge.1994.

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue.2, pp-278-287 ISSN: 2249-6645

A Wavelet-based Feature Selection Scheme for Palm-print Recognition

Hafiz Imtiaz, Shaikh Anowarul Fattah

(Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh)

ABSTRACT

In this paper, a multi-resolution feature extraction algorithm for palm-print recognition is proposed based on two-dimensional discrete wavelet transform (2D-DWT), which efficiently exploits the local spatial variations in a palm-print image. The entire image is segmented into several small spatial modules and a palm-print recognition scheme is developed, which extracts histogram-based dominant wavelet features from each of these local modules. This not only drastically reduces the feature dimension but also results in a very high within-class compactness and between-class separability of the extracted features. Moreover, the improvement of the quality of the extracted features as a result of illumination adjustment has also been analyzed. A principal component analysis is performed to further reduce the feature dimension. From our extensive experimentations on different palm-print databases, it is found that the performance of the proposed method in terms of recognition accuracy and computational complexity is superior to that of some of the recent methods.

Keywords **- Feature extraction, classification, discrete wavelet transform, entropy based information content, histogram, dominant wavelet-domain feature, palm-print recognition, modularization**

I. INTRODUCTION

Conventional ID card and password based identification methods, although very popular, are no more reliable as before because of the use of several advanced techniques of forgery and password-hacking. As an alternative, biometrics, such as palm-print, finger-print, face and iris being used for authentication and criminal identification [9]. The main advantage of biometrics is that these are not prone to theft and loss, and do not rely on the memory of their users. Moreover, they do not change significantly over time and it is difficult for a person to alter own physiological biometric or imitate that of other person's. Among different biometrics, in security applications with a scope of collecting digital identity, the palm-prints are recently getting more attention among researchers [4, 11].

Palm-print recognition is a complicated visual task even for humans. The primary difficulty arises from the fact that different palm-print images of a particular person may vary largely, while those of different persons may not necessarily vary significantly. Moreover, some aspects of palm-prints, such as variations in illumination, position, and scale, make the recognition task more complicated [8].

Palm-print recognition methods are based on extracting unique major and minor line structures that remain stable throughout the lifetime. In this regard, generally, either line-based or texture-based feature extraction algorithms are employed [16, 17]. In the line-based schemes, generally, different edge detection methods are used to extract palm lines (principal lines, wrinkles, ridges, etc.) [15, 12]. The extracted edges, either directly or being represented in other formats, are used for template matching. Canny edge detector is used for detecting palm lines in [15], whereas in [12], feature vectors are formed based on a low-resolution edge maps. In cases where more than one person possess similar principal lines, line based algorithms may result in ambiguous identification. In order to overcome this limitation, the texture-based feature extraction schemes can be used, where the variations existing in either the different blocks of images or the features extracted from those blocks are computed [1, 2, 3, 6, 14]. In

www.ijmer.com Vol.1, Issue.2, pp-278-287 ISSN: 2249-6645

this regard, generally, principal component analysis (PCA) or linear discriminant analysis (LDA) are employed directly on palm-print image data or some popular transforms, such as Fourier, wavelets and discrete cosine transforms (DCT), are used for extracting features from the image data. Because of the property of shift-invariance, it is well known that wavelet based approach is one of the most robust feature extraction schemes, even under variable illumination [7]. Given the extracted features, various classifiers, such as decision-based neural networks and Euclidean distance based classifier, are employed for palm-print recognition [15, 12]. Despite many relatively successful attempts to implement face or palm-print recognition system, a single approach, which combines accuracy, robustness, and low computational burden, is yet to be developed.

In order to extract distinguishable features among different persons, in this paper, we propose to extract precisely spatial variations from each local zone of the entire palm-print image instead of concentrating on a single global variation pattern. In the proposed palm-print recognition scheme, the entire palm-print image of a person is segmented into several small modules. A wavelet domain feature extraction algorithm using 2D-DWT is developed to extract histogram-based dominant wavelet coefficients corresponding to the spatial modules residing within the image. In comparison to the discrete Fourier transform, the DWT is used as it possesses a better space-frequency localization. Moreover, the improvement of the quality of the extracted features as a result of illumination adjustment has also been analyzed. Apart from considering only the dominant features, further reduction of the feature dimension is obtained by employing the PCA. Finally, recognition task is carried out using a distance based classifier.

II. BRIEF DESCRIPTION OF THE PROPOSED SCHEME

A typical palm-print recognition system consists of some major steps, namely, input palm-print image collection, pre-processing, feature extraction, classification and template storage or database, as illustrated in Fig. 1. The input palm-print image can be collected generally by using a palm-print scanner. In the process of capturing palm images, distortions including rotation, shift and translation may be present in the palm images, which make it difficult to locate at the correct position. Pre-processing sets up a coordinate system to align palm-print images and to segment a part of palm-print image for feature extraction. For the purpose of classification, an image database is needed to be prepared consisting template palm-images of different persons. The recognition task is based on comparing a test palm-print image with template data. It is obvious that considering images themselves would require extensive computations for the purpose of comparison. Thus, instead of utilizing the raw palm-print images, some characteristic features are extracted for preparing the template. It is to be noted that the recognition accuracy strongly depends upon the quality of the extracted features. Therefore, the main focus of this research is to develop an efficient feature extraction algorithm.

The proposed feature extraction algorithm is based on extracting spatial variations precisely from the spatial modules of the palm-print image instead of utilizing the image as a whole. In view of this, a modularization technique is employed first to segment the entire palm-print into several small segments. It should be noted that variation of illumination of different palm-print images of the same person may affect their similarity. Therefore, prior to feature extraction, an illumination adjustment step is included in the proposed algorithm. After feature extraction, a classifier compares two palm-print features and a database is used to store registered templates and also for verification purpose.

www.ijmer.com Vol.1, Issue.2, pp-278-287 ISSN: 2249-6645

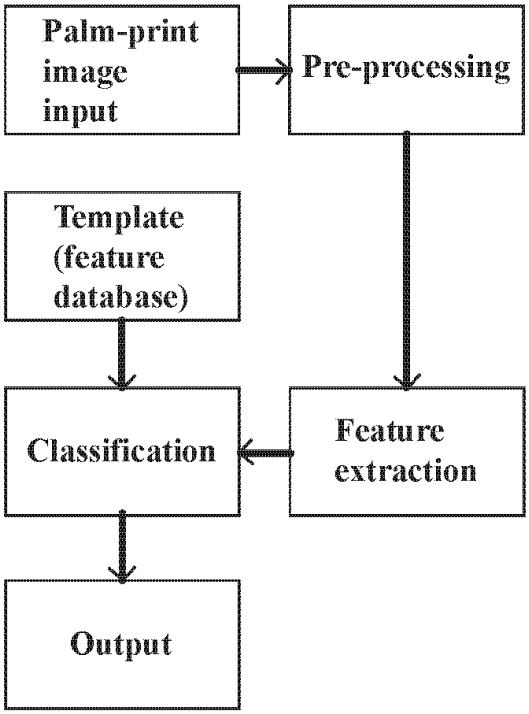


Figure 1: Block diagram of the proposed method.

III. PROPOSED METHOD

For any type of biometric recognition, the most important task is to extract distinguishing features from the template data, which directly dictates the recognition accuracy. In comparison to person recognition based on face or voice biometrics, palm-print recognition is very challenging even for a human being. For the case of palm-print recognition, obtaining a significant feature space with respect to the spatial variation in a palm-print image is very crucial. Moreover, a direct subjective correspondence between palm-print features in the spatial domain and those in the wavelet domain is not very apparent. In what follows, we are going to demonstrate the proposed feature extraction algorithm for palm-print recognition, where spatial domain local variation is extracted from wavelet domain transform.

A. Wavelet-based Feature Extraction from Spatial Modules

For biometric recognition, feature extraction can be carried out using mainly two approaches, namely, the spatial domain approach and the frequency domain approach [13]. The spatial domain approach utilizes the spatial data directly from the palm-print image or employs some statistical measure of the spatial data. On the other hand, frequency domain approaches employ some kind of transform over the palm-print image for feature extraction. In case of frequency domain feature extraction, pixel-by-pixel comparison between palm-print images in the spatial domain is not necessary. Phenomena, such as rotation, scale and illumination, are more severe in the spatial domain than in frequency domain. Recently, multi-resolution analysis, such as wavelet analysis, is also getting popularity among researchers. In what follows, we intend to develop a feature extraction algorithm based on multi-resolution transformation.

It is well-known that Fourier transform based palm-print recognition algorithms involve complex computations and choices of spatial and frequency resolution are limited. In contrast, DWT offers a much better space-frequency localization. This property of the DWT is helpful for analyzing images, where the information is localized in space. The wavelet transform is analogous to the Fourier transform with the exception that it uses scaled and shifted versions of wavelets and the decomposition of a signal involves sum of these wavelets. The DWT kernels exhibit properties of horizontal, vertical and diagonal directionality.

The continuous wavelet transform (CWT) of a signal $s(t)$ using a wavelet $\psi(t)$ is mathematically defined as

$$C(a,b) = \frac{1}{\sqrt{a}} \int s(t)\psi(\frac{t-b}{a})dt, \quad (1)$$

where a is the scale and b is the shift. The DWT coefficients are obtained by restricting the scale (a) to powers of 2 and the position (b) to integer multiples of the scales, and are given by

$$c_{j,k} = 2^{j/2} \int_{-\infty}^{\infty} s(t)\psi(2^j t - k)dt, \quad (2)$$

where j and k are integers and $\psi_{j,k}$ are orthogonal baby wavelets defined as

www.ijmer.com Vol.1, Issue.2, pp-278-287 ISSN: 2249-6645

$$\psi_{j,k} = 2^{j/2}\psi(2^j t - k) \quad (3)$$

The approximate wavelet coefficients are the high-scale low-frequency components of the signal, whereas the detail wavelet coefficients are the low-scale high-frequency components. The 2D-DWT of a two-dimensional data is obtained by computing the one-dimensional DWT, first along the rows and then along the columns of the data. Thus, for a 2D data, the detail wavelet coefficients can be classified as vertical, horizontal and diagonal detail.

Palm-prints of a person possess some major and minor line structures along with some ridges and wrinkles. A person can be distinguished from another person based on the differences of these major and minor line structures. Fig. 2 shows sample palm-print images of two different persons. The three major lines of the two persons are quite similar. They differ only in minor line structure. In this case, if we considered the line structures of the two images locally, we may distinguish the two images.

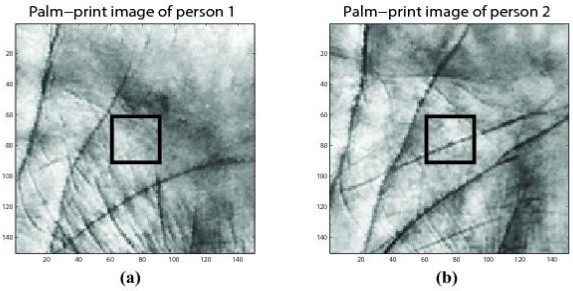
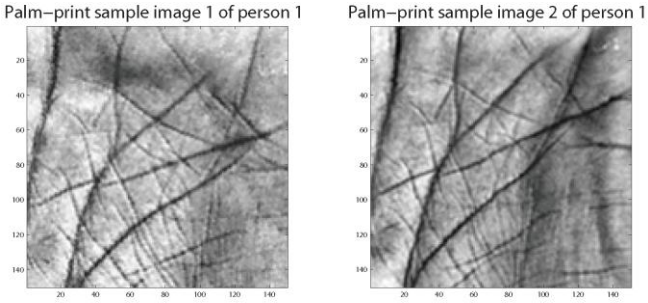


Figure 2: Sample palm-print images of two persons. Square block contains portion of images (a) without any minor line (b) with a minor line

B. Effect of Illumination

It is intuitive that palm-images of a particular person captured under different lighting conditions may vary significantly, which can affect the palm-print recognition accuracy. In order to overcome the effect of lighting variation in the proposed method, illumination adjustment is performed prior to feature extraction. Given two palm-print images of a single person having different intensity distributions due to variation in illumination conditions, our objective is to provide with similar feature vectors for these two images irrespective of the difference in illumination conditions. Since in the proposed method, feature extraction is performed in the DWT domain, it is of our interest to analyze the effect of variation in illumination on the DWT-based feature extraction.

In Fig. 3(a), two palm-print images of the same person are shown, where the second image has a slightly lower average illumination level. 2D-DWT operation is performed upon each image, first without any illumination adjustment and then after performing illumination adjustment. Considering all the 2D-DWT approximate coefficients to form the feature vectors for these two images, a measure of similarity can be obtained by using correlation. In Figs. 3(b) and (c), the cross-correlation values of the 2D-DWT approximate coefficients obtained by using the two images without and with illumination adjustment are shown, respectively. It is evident from these two figures that the latter case exhibits more similarity between the DWT approximate coefficients indicating that the features belong to the same person.



www.ijmer.com Vol.1, Issue.2, pp-278-287 ISSN: 2249-6645

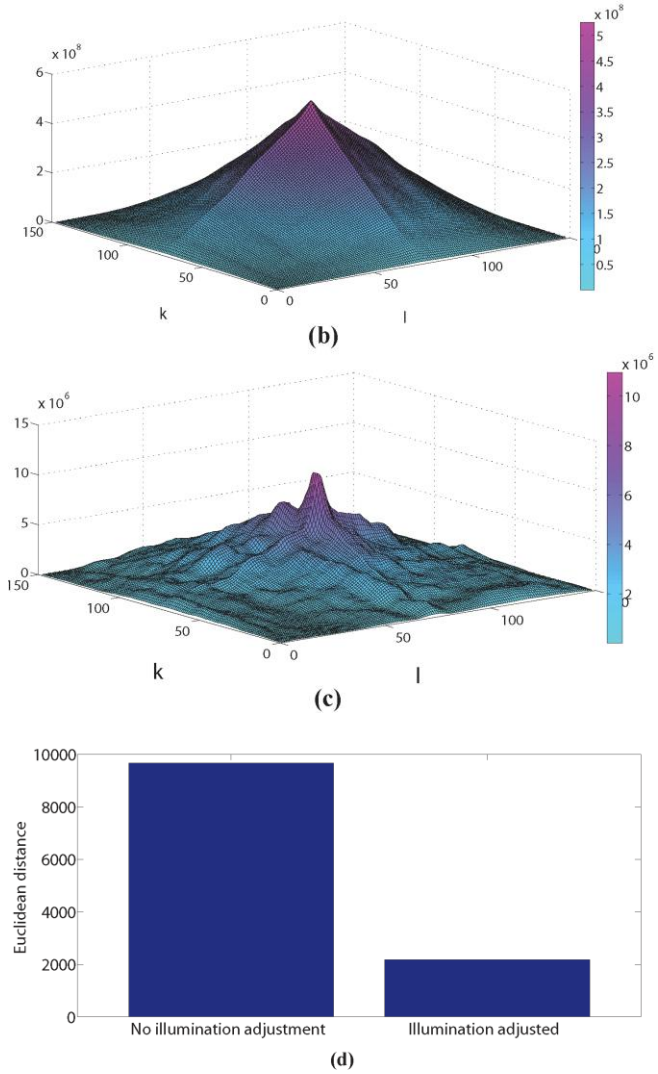


Figure 3: (a) Two sample palm-print images of the same person under different illumination; correlation of the 2D-DWT approximate coefficients of the sample palm-print images shown in Fig. 3(a): (b) no illumination adjustment (c) illumination adjusted; and (d) Euclidian distance between 2D-DWT approximate coefficients of sample palm-print images

The similarity measure in terms of Euclidean distances between the 2D-DWT approximate coefficients of the two images for the aforementioned two cases are also calculated and shown in Fig. 3(d). It is observed that there exists a huge separation in terms of Euclidean distance when no illumination adjustment is performed, whereas the distance is very small when illumination adjustment is performed, as expected, which clearly indicates that a better similarity between extracted feature vectors.

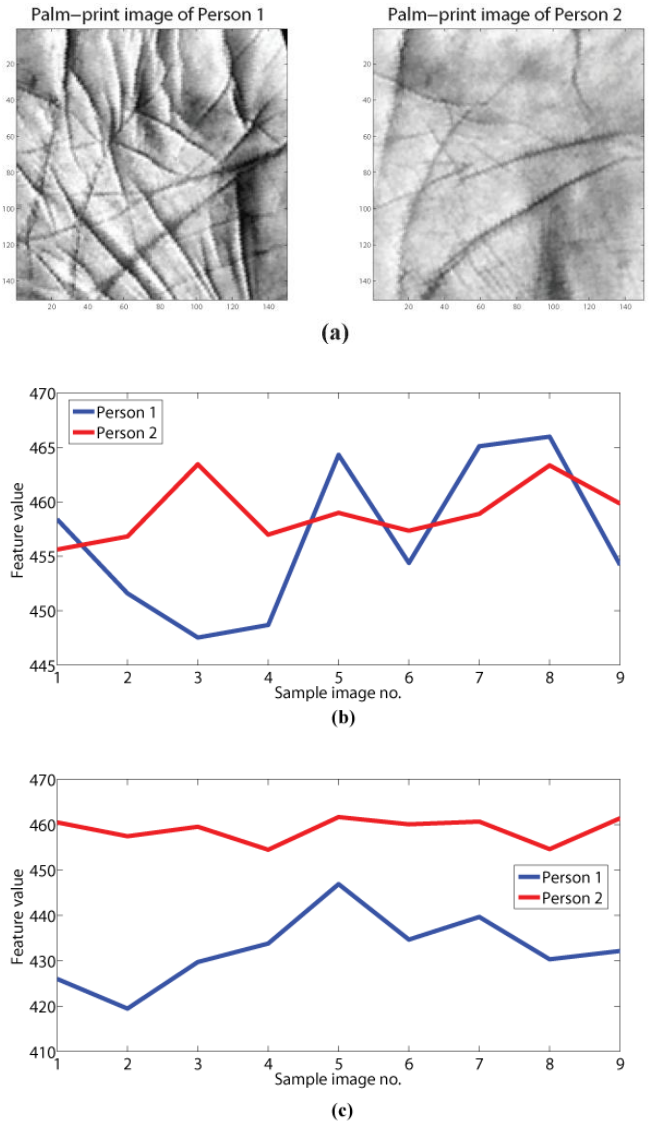
C. Proposed Wavelet Domain Dominant Feature

Instead of considering the DWT coefficients of the entire image, the coefficients obtained form each modules of the palm-print image are considered to form the feature vector of that image. However, if all of these coefficients were used, it would definitely result in a feature vector with a very large dimension. In view of reducing the feature dimension, we propose to utilize wavelet coefficients, which are playing the dominant role in the representation of the image. In order to select the dominant wavelet coefficients, we propose to consider the frequency of occurrence of the wavelet coefficients as the determining characteristic. It is expected that coefficients with higher frequency of occurrence would definitely dominate over all the coefficients for image reconstruction and it would be sufficient to consider only those coefficients as desired features. One way to visualize the frequency of occurrence of wavelet coefficients is to compute the histogram of the coefficients of a segment of a palm image. In order to select the dominant features from a given histogram, the coefficients having frequency of occurrence greater than a certain threshold value are considered.

It is intuitive that within a palm-print image, the image intensity distribution may drastically change at different localities. In order to select the dominant wavelet coefficients, if the thresholding operation were to be performed over the wavelet coefficients of the entire image, it would be difficult to obtain a global threshold value that is suitable for every local zone. Use of a global threshold in a palm-print image may offer features with very low between-class separation. In order to obtain high within-class compactness as well as high between-class separability, we have considered wavelet coefficients corresponding to the smaller spatial modules residing within a palm-print image, which are capable of extracting variation in image geometry locally. In this

www.ijmer.com Vol.1, Issue.2, pp-278-287 ISSN: 2249-6645

case, for each module, a different threshold value may have to be chosen depending on the wavelet coefficient values of that segment.



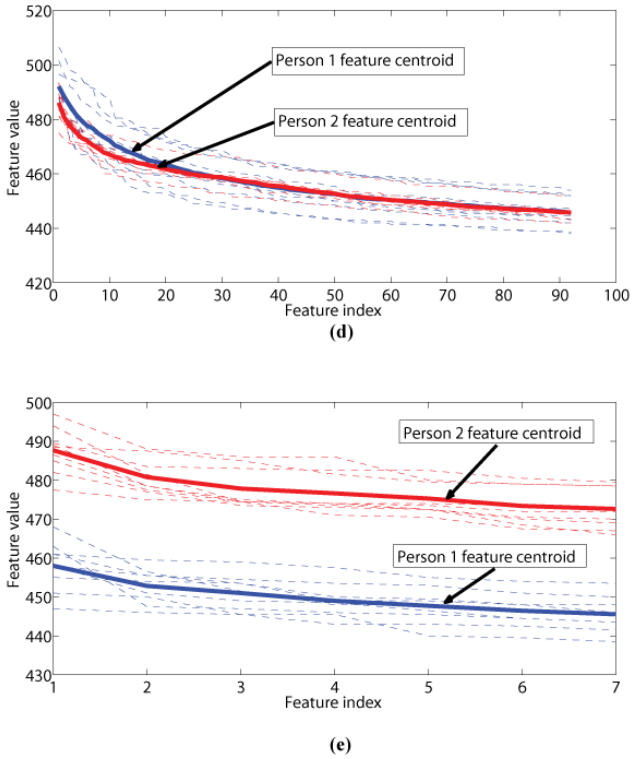


Figure 4: (a) Sample palm-print images of two persons; Feature centroids of different images for: (b) un-modularized palm-print image (c) modularized palm-print image; Feature values for: (d) un-modularized palm-print image (e) modularized palm-print image

We propose to utilize the coefficients (approximate and horizontal detail) with frequency of occurrence greater than $\theta\%$ of the maximum frequency of occurrence for the particular module of the palm-print image and are considered as dominant wavelet coefficients and selected as features for the particular segment of the image. This operation is repeated for all the modules of a palm-print image.

Next, in order to demonstrate the advantage of extracting dominant wavelet coefficients corresponding to some smaller modules residing in a palm-print image, we conduct an experiment considering two different cases: (i) when the entire palm-print image is used as a whole and (ii) when all the modules of that image are used separately for feature extraction. For these two cases, centroids of

the dominant approximate wavelet coefficients obtained from several poses of two different persons (appeared in Fig. 4(a)) are computed and shown in Figs. 4(b) and (c), respectively. It is observed from Fig. 4(b) that the feature-centroids of the two persons for different sample palm-print images are not well-separated and even for some images they overlap with each other, which clearly indicates poor between-class separability. In Fig. 4(c), it is observed that, irrespective of the sample images, the feature-centroids of the two persons maintain a significant separation indicating a high between-class separability, which strongly supports the proposed local feature selection algorithm.

We have also considered dominant feature values obtained for various sample images of those two persons in order to demonstrate the within class compactness of the features. The feature values, along with their centroids, obtained for the two different cases, i.e., extracting the features from the palm-print image without and with modularization, are shown in Figs. 4(d) and (e), respectively. It is observed from Fig. 4(d) that the feature values of several sample palm-print images of the two different persons are significantly scattered around the respective centroids resulting in a poor within-class compactness. On the other hand, it is evident from Fig. 4(e) that the centroids of the dominant features of the two different persons are well-separated with a low degree of scattering among the features around their corresponding centroids. Thus, the proposed dominant features extracted locally within a palm-print image offer not only a high degree of between-class separability but also a satisfactory within-class compactness.

D. Feature Dimensionality Reduction

For the cases where the acquired palm-print are of very high resolution, even after selection of dominant features from the small segments of the palm-print image, the feature vector length may still be very high. Further dimensionality reduction may be employed for reduction in computational burden.

Principal component analysis (PCA) is a very well-known and efficient orthogonal linear transformation [10]. It reduces the dimension of the feature space and the correlation among the feature vectors by projecting the original feature space into a smaller subspace through a transformation. The PCA transforms the original p-dimensional feature vector into the L-dimensional linear subspace that is spanned by the leading eigenvectors of the covariance matrix of feature vector in each cluster $(L < p)$. PCA is theoretically the optimum transform for given data in the least square sense. For a data matrix, X^T, with zero empirical mean, where each row represents a different repetition of the experiment, and each column gives the results from a particular probe, the PCA transformation is given by:

$$Y^T = X^T W = V\Sigma^T \quad (4)$$

where the matrix Σ is an $m \times n$ diagonal matrix with nonnegative real numbers on the diagonal and $W\Sigma V^T$ is the singular value decomposition of X. If q sample palm-print images of each person are considered and a total of M dominant DWT coefficients (approximate and horizontal detail) are selected per image, the feature space per person would have a dimension of $q \times M$. For the proposed dominant features, implementation of PCA on the derived feature space could efficiently reduce the feature dimension without loosing much information. Hence, PCA is employed to reduce the dimension of the proposed feature space.

E. Palm-print Recognition

In the proposed method, for the purpose of recognition using the extracted dominant features, a distance-based similarity measure is utilized. The recognition task is carried out based on the distances of the feature vectors of the training palm-images from the feature vector of the test palm-image. Given the m-dimensional feature vector for the k-th sample image of the j-th person be $\{\gamma_{jk}(1), \gamma_{jk}(2),..., \gamma_{jk}(m)\}$ and a test sample image f with a feature vector $\{v_f(1), v_f(2),..., v_f(m)\}$, a similarity measure

between the test image f of the unknown person and the sample images of the j-th person, namely *average sum-squares distance,* Δ, is defined as

$$\Delta_j^f = \frac{1}{q}\sum_{k=1}^{q}\sum_{i=1}^{m} |\gamma_{jk}(i) - v_f(i)|^2, \quad (5)$$

where a particular class represents a person with q number of sample palm-print images. Therefore, according to (5), given the test sample image f, the unknown person is classified as the person j among the p number of classes when

$$\Delta_j^f \le \Delta_g^f, \ \forall j \ne g \ and \ \forall g \varepsilon \{1,2,...,p\} \quad (6)$$

IV. EXPERIMENTAL RESULTS

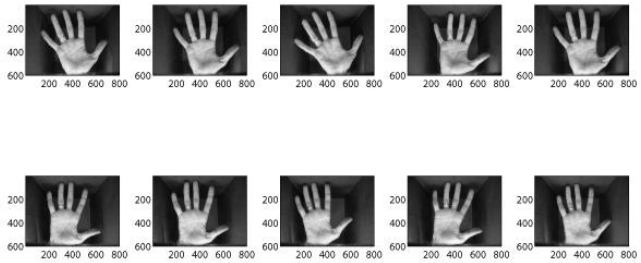
Extensive simulations are carried out in order to demonstrate the effectiveness of the proposed method of palm-print recognition using the palm-print images of several well-known databases. Different analyses showing the effectiveness of the proposed feature extraction algorithm have been shown. The performance of the proposed method in terms of recognition accuracy is obtained and compared with those of some recent methods [2, 5].

A. Palm-print Recognition

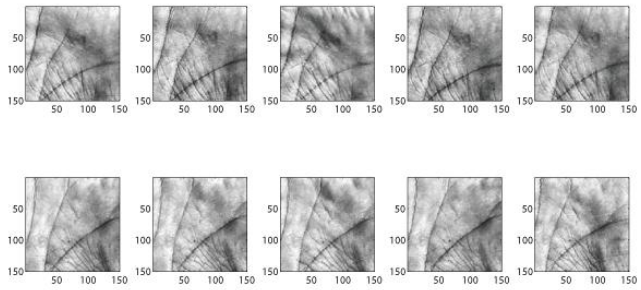
In this section, palm-print recognition performance obtained by different methods has been presented using two standard databases, namely, the PolyU palm-print database (version 2) (available at http://www4.comp.polyu.edu.hk/~biometrics/) and the IITD palm-print database (available at http://web.iitd.ac.in/~ajaykr/Databasen_Palm.htm). In Figs. 5(a) and (b), sample palm-print images from the PolyU database and the IITD database are shown, respectively. The PolyU database (version 2) contains a total of 7752 palm-print images of 386 persons. Each person has 18 to 20 different sample palm-print images taken in two different instances. The IITD database, on the other hand, consists a total of 2791 images of 235 persons, each person having 5 to 6 different sample palm-print images for both left hand and right hand. It can be observed from Figs. 5(a) and (b) that not all the portions of the palm-print images are required to be considered for feature extraction [4]. The portions of the images containing fingers and the black regions are discarded from the original images to form the regions of interest (ROI) as shown in Figs. 5(c) and (d).

B. Performance Comparison

In the proposed method, dominant features (approximate and horizontal detail 2D-DWT coefficients) obtained from



(a)



(c)

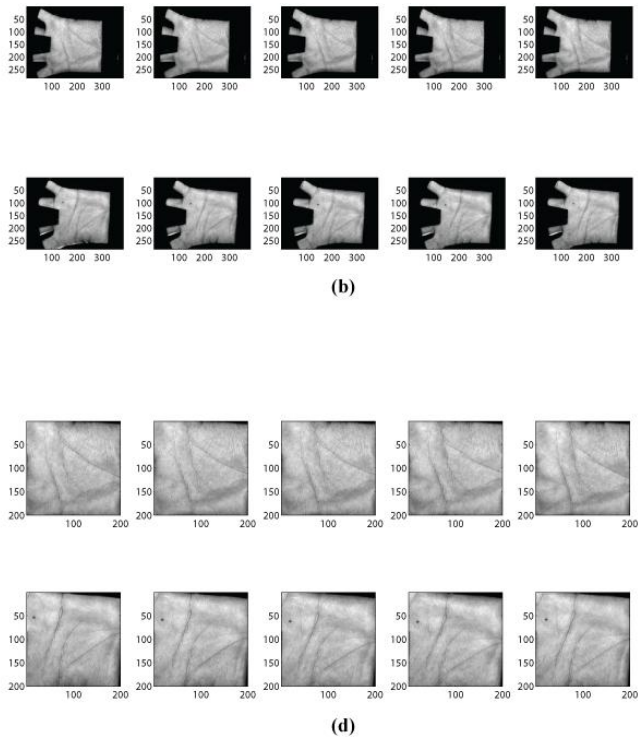


Figure 5: Sample palm-print images from: (a) the IITD database and (b) the PolyU database; Sample palm-print images after cropping: (c) from the IITD database and (d) from the PolyU database

all the modules of palm-print image are used to form the feature vector of that image and feature dimension reduction is performed using PCA. The recognition task is carried out using a simple Euclidean distance based classifier as described in Section 3.E. The experiments were performed following the leave-one-out cross validation rule.

For simulation purposes, the module size for the PolyU database and the IITD database has been chosen as 16×16 pixels and 8×8 pixels, respectively. The dominant wavelet coefficients corresponding to all the local segments residing in the palm-print images are then obtained using $\theta = 20$. For the purpose of comparison, recognition accuracy obtained using the proposed method along with those reported in [2] and [5] are listed in Table 1. It is evident from the table that the recognition accuracy of the proposed method is comparatively higher than those obtained by the other methods. The performance of the proposed method is also very satisfactory for the IITD database (for both left hand and right hand palm-print images). An overall recognition accuracy of 99.82% is achieved.

Table 1: Comparison of recognition accuracies

Method	**Recognition accuracies**
Proposed method	99.79%
Method [2]	97.50%
Method [5]	98.00%

V. CONCLUSION

In the proposed DWT-based palm-print recognition scheme, instead of operating on the entire palm-print image at a time, dominant features are extracted separately from each of the modules obtained by image-segmentation. It has been shown that because of modularization of the palm-print image, the proposed dominant features, that are extracted from the sub-images, attain better discriminating capabilities. The proposed feature extraction scheme is shown to offer two-fold advantages. First, it can precisely capture local variations that exist in the major and minor lines of palm-print images, which plays an important role in discriminating different persons. Second, it utilizes a very low dimensional feature space for the recognition task, which ensures lower computational burden. For the task of classification, an Euclidean distance based classifier has been employed and it is found that, because of the quality of the extracted features, such a simple classifier can provide a very satisfactory recognition performance and there is no need to employ any complicated classifier. From our extensive simulations on different standard palm-print databases, it has been observed that the proposed method, in comparison to some of the recent methods, provides excellent recognition performance.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude towards the authorities of the Department of

Electrical and Electronic Engineering and Bangladesh University of Engineering and Technology (BUET) for providing constant support throughout this research work.

REFERENCES

[1] T. Connie and A.T.B. Jin and M.G.K. Ong and D.N.C. Ling. An automated palmprint recognition system. *Image and Vision Computing*, 23:501--515, 2005.

[2] Manisha P. Dale and Madhuri A. Joshi and Neena Gilda. Texture Based Palmprint Identification Using DCT Features. *Proc. Int. Conf. Advances in Pattern Recognition*, pages 221--224, 2009.

[3] Xiao Yuan Jing and David Zhang. A Face and Palmprint Recognition Approach Based on Discriminant DCT Feature Extraction. *IEEE Trans. Systems, Man and Cybernetics*, 34:167--188, 2004.

[4] Adams Kong and David Zhang and Mohamed Kamel. A survey of palmprint recognition. *Pattern Recognition*, 42:1408-1418, 2009.

[5] J. Lu and Y. Zhao and J. Hu. Enhanced Gabor-based region covariance matrices for palmprint recognition. *Electron. Lett.*, 45:880-881, 2009.

[6] Imitaz, H. and Fattah, S.A., "A spectral domain feature extraction scheme for palm-print recognition," Int. Conf. Wireless Communications and Signal Processing (WCSP), 2010, pp.1-4.

[7] Ekinci, Murat and Aykut, Murat. Palmprint Recognition by Applying Wavelet-Based Kernel PCA. *Journal of Computer Science and Technology*, 23:851-861, 2008.

[8] C.C. Han and H.L. Cheng and C.L. Lin and K.C. Fan. Personal authentication using palm-print features. *Pattern Recognition*, 36:371--381, 2003.

[9] Jain, A.K. and Ross, A. and Prabhakar, S. An introduction to biometric recognition. *IEEE Trans. Circuits and Systems for Video Technology*, 14(1):4 - 20, 2004.

[10] I.T. Jolloffe. Principal Component Analysis. *Springer-Verlag, Berlin*, 1986.

[11] A. Kong and D. Zhang and G. Lu. A study of identical twins palmprint for personal verification. *Pattern Recognition*, 39:2149-2156, 2006.

[12] S.Y. Kung and S.H. Lin and M. Fang. A neural network approach to face / palm recognition. *Proc. IEEE Workshop Neural Networks for Signal Processing*, pages 323--332, 1995.

[13] Li, W. and Zhang, D. and Zhang, L. and Lu, G. and Yan, J. 3-D Palmprint Recognition With Joint Line and Orientation Features. *IEEE Trans. Systems, Man, and Cybernetics, Part C*, 41(2):274 -279, 2011.

[14] Jiwen Lu and Erhu Zhang and Xiaobin Kang and Yanxue Xue and Yajun Chen. Palmprint Recognition Using Wavelet Decomposition and 2D Principal Component Analysis. *Proc. Int. Conf. Communications, Circuits and Systems Proceedings*, pages 2133--2136, 2006.

[15] X. Wu and K. Wang and D. Zhang. Fuzzy direction element energy feature (FDEEF) based palmprint identification. *Proc. Int. Conf. Pattern Recognition*, pages 95--98, 2002.

[16] Xiangqian Wu and Zhang, D. and Kuanquan Wang. Palm line extraction and matching for personal authentication. *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans*, 36(5):978 -987, 2006.

[17] Imtiaz, H. and Fattah, S.A., "A DCT-based feature extraction algorithm for palm-print recognition," IEEE Int. Conf. Communication Control and Computing Technologies (ICCCCT), 2010, pp.657-660.

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

Comparison Among Ambiguous Virtual Keyboards For People With Severe Motor Disabilities

A.J. Molina[1], O. Rivera[1], I. Gómez[1], M. Merino[1], J. Ropero[1]

*(Department of Electronic Technology, University of Seville, Spain)

ABSTRACT

This paper presents an exhaustive study on the different topologies of ambiguous soft keyboards, analyzing the text entry average time per character and the average number of user inputs necessary for its creation. Various topologies and design criteria are investigated. In addition, an analytical model is also proposed. This model allows one to compare among different topologies and estimate the sensitivity that different keyboards offer when compared with dictionary hit rates. It has been found that ambiguous keyboards, with six keys, are better to use.

Keywords **- Descriptive User Models, Human-Computer Interaction, Indirect Text Input, System Model, Virtual Keyboard.**

I. INTRODUCTION

1.1 The Need of Adapted Communication Systems

Communication is a fundamental element to obtain social integration. People with severe motor disabilities often have reduced communication skills. Many of them have speech impairments that make them difficult to be understood, and external aid systems are required to carry out daily communication tasks. These are augmentative and alternative communication systems (AAC).

1.2. Scanning-based Virtual Keyboards

Some of the AAC systems are text entry systems that are equipped with a text-to-speech component, letting an oral communication from an entered text, augmenting communication possibilites. However, conventional text entry devices may not be used by motor disables, especially when they have severe disabilities because of their lack of precision movements. Therefore, devices that are adapted to their skills have to be used.

An alternative that is widely accepted is the use of a virtual keyboard (VK) used as a substitute for a conventional keyboard. A VK is a software application whose graphical user interface represents a keyboard. A VK that has one character in each key is named unambiguous keyboard. On the other hand, an ambiguous VK contains more than one character in some keys. These require disambiguation of the character contained in the key. Ambiguous VK demonstrates some advantages with respect unambiguous ones [1].

1. The efficiency of an ambiguous keyboard is near to one keystroke per letter. 2. Apart from literacy, no memorization of special encodings is required. 3. Attention to the display is required only after the word has been typed. 4. A keyboard with fewer keys can have larger keys for direct selection. 5. The average time to select a key by scanning is reduced considerably. 6. Simple linear scanning can be used efficiently to select a key. 7. Fewer keys may allow direct selection with various input devices.

The keys of a VK may be selected using an interaction method adapted to user's skills, such as a stylus over a touchable screen, an adapted mouse, eye-gaze trackers [2,3], head movement detectors [4,5], etc. However, a simple interaction method is required for severe motor handicapped people. This method is based on detecting a residual voluntary movement or some brain activity, such as the one used in Brain Computer Interaction (BCI) Systems. The most simple device is capable of detecting only one kind of user input, used to select the current item, such as a single switch, blink detectors, wink detectors, saccadic movement detectors, etc. [6-8]. To Use this kind of devices, an indirect text entry method to select the desired option from the currently highlighted options has to be implemented. In this sense, a scanning method is a possible alternative. An option or a group of options is highlighted in each scanning step. If the desired option is in the currently highlighted ones, a selection has to be done.

The scanning method can be automatic or manual. In automatic scanning, an internal timer establishes the dwell time or the time that any key or row is

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

highlighted. A switch keystroke makes the selection of the current highlighted key or row. In manual scanning, a keystroke makes the current highlighted group to advance to the next. A second switch or the timeout of a timer makes a selection.

On the other hand, the scanning method can be implemented in a linear or row-column mode. In linear scanning, each key is highlighted, one after the other. In the latter, a cluster of keys are highlighted. In matrix VK, a cluster can be made by a row of keys. A selection when a row is highlighted performs a linear scanning of its keys. A new linear scanning on character contained in a key can be performed if a preselected key has more than one character on it. Finally, once a character is chosen, it is convenient to restart scanning from the first row [9].

1.3. The importance of proper configuration

The main drawback of the use of a scanning-based text entry system is its low text entry rate. In [10], it is estimated that the maximum communication rate using this kind of systems is 10 words per minute (WPM). In a normal conversation, able-bodied people may pronounce between 180 and 200 WPM. Some situations in which handicapped people may not participate in a normal conversation could occur because of this threshold, and therefore, this may drive to a social exclusion. Instead of this rate, these systems are the unique alternative in case of severe motor disabilities.

A proper selection and configuration of the system is important to obtain a good communication rate. There are many VKs and scanning methods. Once a given keyboard is chosen, tuning it to the user's skills and preferences may yield significant performance and comfort benefits. A study of optimal configurations of the input devices for people with physical impairments is shown in [11].

VKs are usually set in a rectangular matrix of keys, and each one of them may contain different amounts of characters, although optimal ones can have a button arrangement, different from the rectangular matrix. A study that includes different keyboards is depicted in [12], where a nonmatricial unambiguous keyboard, whose arrangement is determined by character frequency, is compared with other matricial ambiguous ones, such as Huffman, alphabetic, or mobile distribution keyboards. More examples of matricial VK can be found in [13], where a QWERTY-type one is used for brain-injured people, and in [14], where several ambiguous VK with nine keys called Levine, TOC, alphabetic, and frequency are compared.

An important issue that affects the text entry rate is how the characters are distributed on the VK keys. Keyboards, whose character layout is based on their frequency, have better performance than QWERTY, whose layout is a reproduction of the traditional keyboard. This is because the most likelihood characters are placed in or nearby the row or key, where the automatic scan starts, so that the mean character access time is reduced, and therefore, the text entry rate could be increased without any negative effect on the number of user inputs (UI). The fixed character layout can be established based on the character frequency in the language and character access time in VK. On the other hand, a dynamical character layout can also be established, where the characters are automatically rearranged depending on the probability of the previous character sequence. It is obvious that analytically, the second option is the best. However, studies with fluctuating keyboards [15] have shown that there is a toll on time taken using this keyboard owing to the high mental load needed to locate the position of the characters. Thus, the performance of fluctuating keyboard matches or worsens that of fixed ones.

On the other hand, the methods used to improve text entry rates in everyday devices can be applied to improve the communication capabilities for the disables. Thus, for example, the T9[1] method could be implemented in an ambiguous VK to enhance the text entry rate. Character or word prediction can be used to improve performance. Prediction can be accomplished in VK showing the most likely character that follows a preselected character sequence (or prefix) [16], [17] or the most likely word that matches with that prefix [18]. Predictor requires the existence of a dictionary and/or a prefixes table with word/prefixes frequency information included.

1.4. How to Compare

VKs testing can be accomplished by experiments or simulation. The part of the system that is tested has to be made as a prototype or a final product in experiment option, implying a cost in time and budget. In addition, end users have to participate in this option, and a trial programming has to be made

[1] T9 is a registered trademark of Tegic Communication and it means text in 9 keys

carefully to obtain correct measurements. The sample of participants should be representative of the end user group, making it difficult to carry out trials and make prototypes in certain situations. In addition, hardware components are required to test software solutions and the testing results may be influenced by them.

On the other hand, simulation consists of an application software that tries to emulate the system behavior, user behavior and interaction among them. The system behavior is more or less stable and may be translated to a program language. However, other components depend on the user and they may change in each trial or during a trial. Simulators have a group of input parameters that let set the simulation context to obtain correct results. In this way, the participation of users is limited and the need of making a prototype or a final product is removed.

The results obtained by these methodologies are representative of the conditions in which they were elaborated. If we want to know how the keyboard performance is affected by external conditions, the experiment or the simulation should be repeated.

There are many prediction systems, each having different characteristics, making it difficult to compare their performances because of the diversity of heterogenous parameters used to measure them. Some authors, such as Gillette and Hoffman [19] or Heinisch and Hecht [20] have carried out studies on commercial predictive products. A study on non-commercial prediction system is presented in [21].

It is necessary to set some metrics to compare different text entry systems or different configurations of one of these systems. VK for motor disabled people gaims to measure two items: the text entry rate and the number of movements (interactions) that have to be done to enter a text. First, a parameter measured in WPM is often used. On the other side, the most used parameter to measure the second item is keystrokes per character (KSPC). It is also clear that the term number of user inputs per character (UI_c) is more general for that diversity of devices instead of KSPC, and hence, we prefer to use this term, although the meaning is completely equivalent.

Comparison UI_c of different VKs can be carried out by simulation, employing extensive texts from a corpus built using several sources, such as digital journals, magazines, dictionaries, etc. This is due to the UI_c parameter that only depends on the operational mode of a specific KSPC, and it is independent of the user. However, obtaining WPM strongly needs a user's model, and hence, the most frequent method to test a VK is in experiments where the users have to use the VK to type a preselected text fragment with a high degree of correlation with the user's language.

1.5. Goals and document structure

A comparison among different ambiguous VKs is presented in this paper. Different disambiguation modes are considered. A simple user model is used to obtain a proper value of the reaction time, and some system models are presented. The latter are probabilistic models obtained from a dictionary that have been complied. One of these models is a mathematical model associated with various VKs that work in Tn (Text in n-keys[2]) operated by single-switch users. The model estimates both the average time per character ($\overline{t_c}$) and average number of user inputs per character($\overline{UI_c}$). By assuming that the average length of a word, $\bar{l}$, including space character, is 6 for English or 5.5 for Spanish, we have WPM = $\frac{60}{\bar{l} \cdot \overline{t_c}}$. The model lets us to test how a VK layout or a dictionary may influence $\overline{t_c}$ and $\overline{UI_c}$.

In section II, a review of VK software is presented with special emphasis on ambiguous ones. Section III presents a common structure of the proposed models. Section IV describes several topologies and operational modes for ambiguous VKs. Two methods are shown: disambiguation by scanning or word approach. In section IV a simple letter scanning-based model is depicted. In addition, the results are also reported in this section. The Tn mode is presented in section IV, and the required NEXT and SPACE functions are discussed. In addition, a probabilistic model and its validation are shown in this section. Finally, in section V, the model is used to establish a comparison among different ambiguous VKs and to state which VK could be better for an user under different user preferences and external conditions. An appendix, in which VKs considered in the analysis are represented, completes the paper.

[2] A generalization of T9 method

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

II. BACKGROUND

Nowadays, abundant scientific studies about text entry using a numerous varieties of methods and devices, such as mobile phones, PDA, etc., could be found. For instance, a model to predict WPM in mobile based on Fitts' law [22] is depicted in [23]. The study included different methods of text input such as multi-tap, T9, or two-keys. Other studies based on this law are shown in [24], [25], and [26]. These studies try to predict the performance of an expert user. A model for predicting the text entry speed of novice users based on Fitts' law is described in [27]. In [28], a combination of the power law of learning and theorical upper limit predictions is used to describe the development of text entry rates from users first contact to asymptotic expert usage. In addition, interesing studies based on GOMS models, such as those depicted in [29], [30], [31], and [32] have also been carried out. In [33], a study about indirect text entry methods and a model based on the notion of a containment hierarchy are presented. An evaluation of unambiguous VKs with character prediction is depicted in [34]. In this paper, a probabilistic model to predict WPM and $\overline{UI_c}$ using an indirect text entry method is shown.

An experimental study of WPM and KSPC for a mobile using keys disambiguation method based on prefixes (instead of a dictionary, as in T9) is presented in [16]. In that paper, character prediction establishes the likeliest character on the selected key, so that it will be shown in the first place according to the previous prefix. A KSPC next to 1.15 is obtained using this method. Other letter reassignments of a mobile keyboard are shown in [35], where an improved text entry is verified with different users. In other devices, such as PDAs, in which the number of keys are strongly reduced in favor of wider screens, software or VKs are developed for entring text toward the focused application. These VKs are representations of unambiguous keyboards (such as QWERTY) or ambiguous ones (such as mobile keyboards) that are controlled by a stylus. Studies of unambiguous VKs are presented in [24] and [36], where predictive models and user tests are included. In [37], a VK with 4 keys is described and tested using several languages. In [17], a study of the application of character prediction on ambiguous VK is depicted.

II. METHODOLOGY

A comparative study of user perfomance (measured in WPM and $\overline{UI_c}$) using scanning-based ambiguous VKs has been carried out in this paper. VKs with different number of keys using different scanning methods and implementing three disambiguation methods have been studied. Two considerations have been set to obtain the values of performance parameters:

1) A free error context. It is not necessary to implement a method to fix errors.

2) Expert users. Mental times, as search times, that are related to cognitive task, are optimal. The text entry is carried out in the optimal time.

Two methodologies can be applied, as mentioned earlier; one based on simulations and the other based on experiments. The users considered in this study have been suffering from severe motor disability. Thus, the use of a simple input device[3] is required. The chosen device depends on the user's skills. Each device has different characteristics that can influence the selection time, and thus, user performance. This study has tried to compare the VKs independently of the chosen input device. Using a methodology based on experiments, a trial by each configuration of VK is necessary to compare them. Furthermore, some previous sessions have to be carried out to obtain the desired level of experience. Much time and effort may be required by the user. In this study, a methodology based on simulations have been followed because of the difficulty of contact with severe motor handicapped people and aforementioned drawbacks.

To use a methodology based on simulations, some models that lead to predict the value of parameters in each case are required. In this sense, the general structure of these models is shown in Fig. 1.

[3] This is able to detect only one kind of user input

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

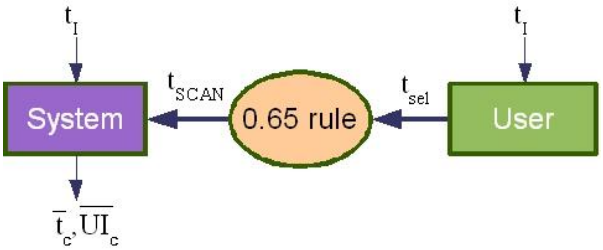


Fig. 1: The structure of models.

First, the reaction time, t_r, is predicted by a model of user behavior. This model is based on Keystroke-Level Model (KLM) [38] and [39]. The t_r is the elapsed time between the presentation of a sensory stimulus and the subsequent behavioral response as a button press. There are three kinds of reaction time experiments [40]: simple, recognition, and choice. In simple reaction time experiments, there is only one stimulus and one response. In recognition reaction time experiments, there are some stimuli that should have a response and others that should get no response. In choice reaction time, the user must give a response that corresponds to the stimulus. The value of t_r depends on the used input device. Therefore, this model has an input parameter that represents the interaction time, t_I[4]. Choice reaction times depend on the number of different stimulus according to Hick-Hymman's law [41]. In our context, with expert users who have a "mental map" of VK layout, the reaction time is reduced to accept or not the highlighted row, key or letter. Thus, the recognition reaction time seems to be more adequate. Henceforth, we will follow KLM notation [42], where the reaction time would include the mental preparation time, M, and the time to make a keystroke, K, or generate a user input. Therefore, the time M can be considered constant for all the processes related to the use of VK.

Once the selection time is predicted, a proper value of dwell time, T_{scan}, has to be calculated. [43], [44], and [45] have shown that the optimal scanning time is related to the reaction time by a constant equal to 0.65. Therefore, the reaction time establishes a lower bound for T_{scan} (Equation 1).

$$\min\{T_{scan}\} = \frac{t_r}{0.65} \qquad (1)$$

Then, the 0.65 rule may be used to obtain proper value of T_{scan}. Subsequently, T_{scan} is used as an input parameter in the model of system behavior. This one is a probabilistic model that emulates the given VK behavior.

III. AMBIGUOUS VKs WITH AUTOMATIC SCANNING

As mentioned earlier, ambiguous VKs require implementation of a disambiguation method. This method allows selection of a character among those on a chosen key. Disambiguation may be carried out in several ways: by using a new scanning of the letters on the chosen key, by character-level prediction, or by word-level approach.

4.1. Disambiguation by scanning: letter scanning

In this case, accessing the characters of a VK is performed by the scanning method. Once a key is selected, the characters on this key are scanned. This scan is usually linear. The value of the dwell time among the characters may be different from that set for scanning a row, column, or key. In this sense, this disambiguation method may be seen as an automatic version of multi-tap that is used on mobiles.

Another alternative is described in [46]. Fig. 2 represents a 12-key VK, where this method has been implemented. Two keys have special codes, called 2nd and 3rd, whose function is to enable next scan through the second or third option of each key, respectively. Instead of other VKs, here, row-column scans only include the first character of each key when the scan starts. To access the second or third character of a key, the user, first of all, has to select the key that contains the code 2nd or 3rd. Subsequently, the scan restarts, including only the characters in the positions indicated by the preselected code. The more likely characters, located in the first position of each key, need only two user inputs, and the rest need four user inputs. It must also be noted that the letter placement in VK with this operation mode differs from the row-column ones.

[4] Time required to interact with the given input device

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

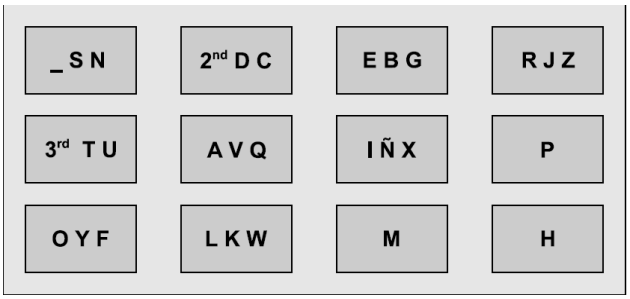


Fig. 2: 12-key VK using scanning depicted in [46]: RED12rcv

4.1.1. System Model

When using this disambiguation method, the value of $\overline{t_c}$ only depends on the times for accessing the characters in the considered context[5]. However, it is necessary to weigh the frequency of usage of each character in the language to obtain the average value of entry time per character. On the other hand, the character access time depends on the position of this character in the layout of VK, the set value of the dwell time, and the interaction time, t_I. The last one is an input parameter of the model, whose value is constant, because it only depends on used input device. As mentioned earlier, the value of the dwell time is obtained by applying the 0.65 rule to the value of the reaction time estimated by the user's model. Therefore, the value of T_{scan} is constant. In this way, once a layout of VK is set, the value of the access time of each character may be considered constant. Hence, the value of $\overline{t_c}$ is given by relationship shown in equation 2, where p represents a matrix of probability of the use of characters in an alphabet and t_{acc} is a matrix containing the access times of characters in a given layout.

$$\overline{t_c} = p^T \cdot t_{acc} \quad (2)$$

Similarly, the value of average UI is obtained in the same way. Let ui_{acc} be the matrix that represents the numbers of necessary user inputs to access a specific character in VK. Then, $\overline{UI_c}$ is given by Equation 3.

$$\overline{UI_c} = p^T \cdot ui_{acc} \quad (3)$$

[5] Free error context and an expert user.

4.1.2. Results

In this paper, ambiguous VKs with 4, 6, 9, 12, and 16 keys have been studied. The chosen layout of VKs is an optimal frequency-letter arrangement of characters in its keys. Two scanning methods have been implemented, a linear and a row-column. The character scanning method is linear. The dwell time is fixed to 1 s, and therefore, the scaling of $\overline{t_c}$ shown in Tables I and II by the appropriated value may give us the results for other dwell times. Obviously, $\overline{UI_c}$ is unaffected by the dwell time.

Although this study is based on ambiguous VKs, the values of $\overline{t_c}$ and $\overline{UI_c}$ using unambiguous VKs may be obtained with the proposed models. In this sense, the values of these parameters have been estimated considering an optimal frequency-letter arrangement of unambiguous VKs with a linear scanning method (CONVl) or a row-column scanning method (CONV).

Table I shows the $\overline{t_c}$ and $\overline{UI_c}$ for the VKs that use letter-scanning. The nomenclature used to identify a VK is RED_NK_KSM, in which NK is the number of VK keys and KSM may be l for linear or rc for row-column.

Table I: Letter-scanning operation mode results for different layouts using row-column or linear key scanning, and linear scanning for characters into a key and T_{scan}=1.}

	$\overline{t_c}$ (s)	$\overline{UI_c}$
RED 4l	5.42	2.00
RED6l	5.35	2.00
RED9l	5.42	2.00
RED12l	5.69	2.00
RED16l	6.65	2.00
CONVl	9.59	1.00
RED4rc	5.94	3.00
RED6rc	5.64	3.00
RED9rc	5.51	3.00

RED12rc	5.51	3.00
RED16rc	5.61	3.00
CONV	5.35	2.00

It can be seen for row-column scanning that $\bar{t}_c$ progressively decreases in relation to the increase in ambiguity, reaching a minimum of 6 keys in VK. Likewise, for linear key scanning, something similar occurs. On comparing the two scanning methods, linear ones are observed to achieve better times than the row-column ones, excluding RED16, and have better number of user inputs. A keyboard with 4, 6, or 9 keys, when a linear scan is used, is able to obtain a $\bar{t}_c$ close to the unambiguous one and with an identical $\overline{UI_c}$.

Table II shows the values that different VKs achieve using the disambiguation method described in [46]. It can be observed that as the ambiguity of the keyboard increases, the benefits worsen. Furthermore, when compared with their counterparts presented in Table I, only keyboards with 12 or 16 keys are found to show improvement. VKs with 9 or less keys show worse results, because they require a larger number of special codes to control the following scan.

Table II: Letter-scanning operation mode results for different layouts using scanning depicted in [46] with T_{scan}=1 s.

	$\bar{t}_c$ (s)	$\overline{UI_c}$
RED4rcv	7.13	4.40
RED6rcv	6.71	3.65
RED9rcv	5.75	2.93
RED12rcv	5.55	2.66
RED16rcv	5.41	2.35

To summarize, better results for ambiguous VKs are obtained by 4-, 6-, 9-, and 12-key VKs with linear scanning of its keys, and RED16rc with the row-column scanning depicted in [46]. Only 4-, 5-, 9-key VKs are found to be very close to unambiguous keyboard.

4.2. Disambiguation by word approach: Tn mode

This disambiguation method is based on T9 method for mobiles. To use this method, two functions are required, NEXT and SPACE. As the text entry is continued, the most likely word associated with the sequence of the selected keys is shown. In most cases (95\% according to [23] for mobile phones), the sought word is predicted. Thus, the SPACE key has to be selected to accept this word and then the text entry is continued by the next word. Only in 5\% of the cases, the suggested word is not accepted. Under these circumstances, the NEXT function has to be selected repeatedly to show other suggestions. In addition, the sought word may not be found in the dictionary and obviously it will not be shown as a suggestion. This is the worst case scenario, and an alternative text entry method has to be used in this case, for example, multi-tap, to type the wished word completely.

In short, the T9 system for mobile phones is very efficient. In [23], it has been shown that WPM is greater using T9 than using multi-tap, and in [16], a comparative table for mobile phones has been presented, where KSPC, using a T9 method, is close to the unit. Nonetheless, in both cases, if the word is not in the dictionary or is not shown in the first position, KSPC will be slightly greater than 1.

4.2.1 NEXT and SPACE functions

To implement this method in an ambiguous VK with n keys (Tn method), a NEXT key is required. Once the NEXT key is selected, the next suggested word is shown. This procedure should be repeated until the sought word is shown or the last suggestion is reached. However, this may produce fatigue to the user and increase the number of user inputs. Hence, it would be preferable to use an automatic variant that can scan the items of the suggestion list in a linear way. This variant requires only an additional user input to choose the word of the list.

On the other hand, the SPACE character has an additional function: to accept the first suggested word. For this reason, the SPACE character may not be integrated with other characters in the same key, because the system would not be able to distinguish if the user is selecting the suggested word or he/she is typing a new character. In this sense, two alternatives of layout may be used to implement the Tn mode: to

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

integrate both the functions, NEXT and SPACE, in the same key, or to use two separate keys.

Some difficulties have been encountered when two new keys are integrated in an ambiguous VK. First, in VKs that are highly ambiguous, for example, from 4 up to 6 keys, WPM may be affected because it would be necessary to carry out a new arrangement of characters in the keys. Second, finding an optimal layout of VK would be a complex task, because the letter arrangement depends on the NEXT key position, and this depends on the used dictionary and, in turn, the letter arrangement in the VK. Third, an optimal VK using Tn mode locates the most likely characters inside the keys that are closest to the scanning start point. In highly ambiguous VKs, this fact could drive to a time penalty, because a sequence of the selected keys is associated with a great number of words. Hence, the probability of the desired word is that the first suggestion is reduced and the use of NEXT key is increased. Such a situation could increase the $\overline{t_c}$ as the user has to search the desired word on the suggestion list. Thus, it is possible that the scrambled arrangement would have a great probability of finding the desired word in the first position, enhancing the text entry rate. Both these situations will be studied in this paper.

The general structure of a VK, where the NEXT and the SPACE functions are separated in two keys, is shown in Fig. 3. The first suggested word is shown on the viewer. If the SPACE key is selected, the suggested word is accepted and the SPACE character between the words is automatically introduced. If the NEXT key is selected, other suggestions are shown in a new window. In this case, a word is highlighted in each scanning step. In such a situation, a user input has to be made to introduce the current highlighted word in the text. Nevertheless, when the end of the list is reached by the scanning method, the alternative text entry method is used to type the desired word completely.

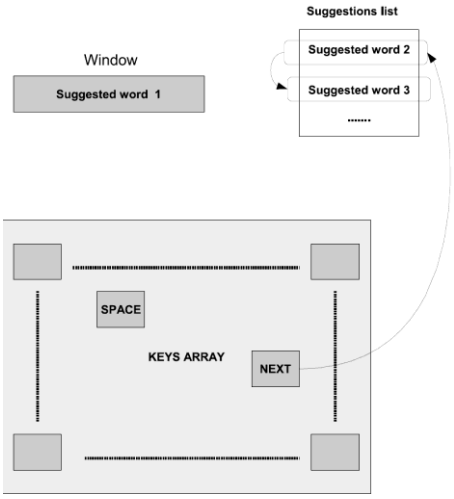


Fig. 3: Ambiguous VK with SPACE and NEXT in independent keys. See text for description.

The general structure of a VK using Tn mode, where the NEXT and the SPACE function are integrated in the same key, is shown in Fig. 4. Once all keys containing all letters of the desired word are selected, the SPACE-NEXT key has to be selected. Subsequently, the list of suggestions appears, and the linear scanning starts. An additional user input is required to select the highlighted word. As before, when the end of the list is reached by the scanning method, the alternative text entry method is used to type the desired word completely. In this solution, an extra user input has to be made if the sought word is shown in the first position. In addition, there is also an increment in the entry time per character, because an additional scanning cycle is always necessary to the focus on the list obtained after the SPACE-NEXT key is selected.

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

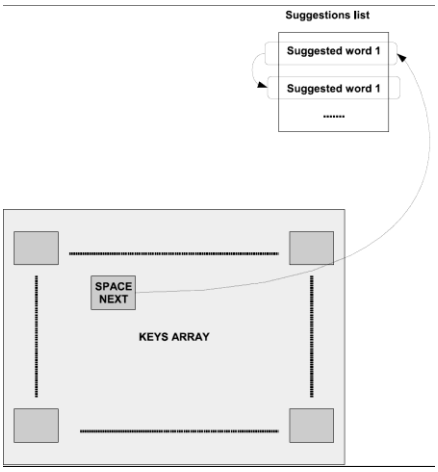


Fig. 4: An ambiguous keyboard with the SPACE-NEXT installed in the same key. See the description of the operation in the text.

4.2.2 System Model

Text entry using Tn mode starts by selecting all keys that contain the characters of the desired word. Accordingly, three cases may be presented:

1) The desired word is the first option in the suggestion list. Therefore, it is shown on the viewer. In this case, the SPACE key has to be selected to enter the word. The average time required to do this operation is represented by $\bar{t}_{first}$. On the other hand, the average number of user inputs required to carry out this operation is represented by $\overline{UI}_{first}$.

2) The desired word is in the suggestion list, but it is not the first option. Hence, the NEXT key has to be selected to scan the next suggestions. Once the word is highlighted, a user input has to be made to select it. $\bar{t}_{nofirst}$ represents the average time to carry out this operation, and $\overline{UI}_{nofirst}$ represents the average number of user input.

3) The desire word is not in the suggestion list. As in the last case, the NEXT key has to be selected to scan the next suggestions. Once the last suggestion is highlighted, the letter-scanning mode is activated. Then, the word has to be entered using this text entry method. The average time required to do this operation is represented by $\bar{t}_{fail}$ and the average number of user input is represented by $\overline{UI}_{fail}$.

The time and the number of user inputs to enter a word depends on the presented case. Therefore, the value of $\bar{t}_c$ and $\overline{UI}_c$ also depends on this. Hence, these cases have to be considered by the system model. A descriptive model of a user interacting with a VK using Tn mode is shown in Fig. 5. Four tasks are represented by a rectangle in this descriptive model: selecting the keys that contain the character of the word (it also includes the selection of NEXT-key or SPACE-key), scanning the suggestion list and typing the word using the letter-scanning mode. In this figure, the three above-mentioned cases are represented. Each case is related to a path that is represented by a dashed line. The first case is related to path 1 (Red), the second one to path 2 (blue) and the third one to path 3 (green).

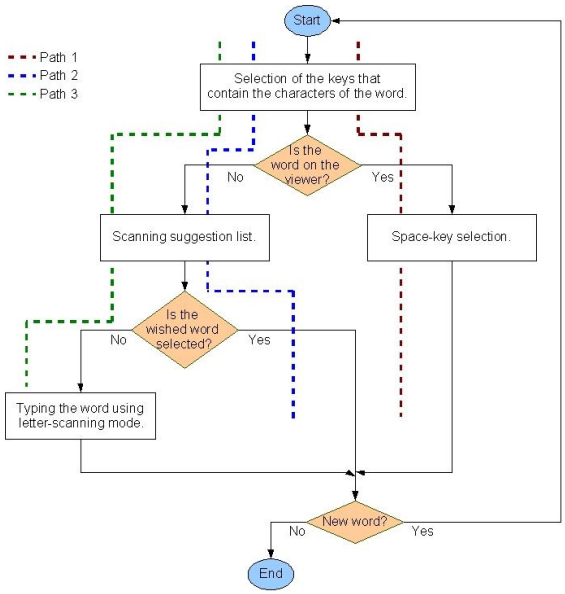


Fig. 5: Model structure.

The average value of time and number of user inputs to enter a word using each path has to be obtained separately. Subsequently, the value of $\bar{t}_c$ and $\overline{UI}_c$ may be estimated by a weighted addition. For that, the average value of time and number of user inputs using each path are calculated by the length of the word and some probabilistic parameters are required. The value of some of these parameters is obtained

using the information contained in a dictionary. Others are input parameters of the model. In this sense, Equation 4 and 5 represent the value of $\bar{t}_c$ and $\overline{UI}_c$ respectively. The probability of entering a word that has n characters (n-word) is represented by $p(w_n)$. The unit in the denominator is due to a space character that follows each word.

$$\bar{t}_c = \frac{\sum_n p(w_n) \cdot T_w(n)}{1 + \sum_n n \cdot p(w_n)} \quad (4)$$

$$\overline{UI}_c = \frac{\sum_n p(w_n) \cdot UI_w(n)}{1 + \sum_n n \cdot p(w_n)} \quad (5)$$

To estimate the value of $T_w(n)$ and $UI_w(n)$, the three aforementioned cases are considered, as it can be checked in Equations 6 and 7. In these, the probability of a n-word in the prediction list is represented by $p(w_n \in L_\Omega)$, and the probability of a n-word in the first position in this list is represented by $p_{in}(w_n \in L_\Omega)$.

$$T_w(n) = p(w_n \in L_\Omega) \cdot \left[p_{in}(w_n \in L_\Omega) \cdot \bar{t}_{first}(n) + (1 - p_{in}(w_n \in L_\Omega)) \cdot \bar{t}_{nofirst}(n) \right] + (1 - p(w_n \in L_\Omega)) \cdot \bar{t}_{fail}(n) \quad (6)$$

$$UI_w(n) = p(w_n \in L_\Omega) \cdot \left[p_{in}(w_n \in L_\Omega) \cdot \overline{UI}_{first}(n) + (1 - p_{in}(w_n \in L_\Omega)) \cdot \overline{UI}_{nofirst}(n) \right] + (1 - p(w_n \in L_\Omega)) \cdot \overline{UI}_{fail}(n) \quad (7)$$

4.2.3 Validation

To check the model effectiveness, a program has been made in C, which interacts with a database MySQL that houses the dictionary. This dictionary consists of 10,000 words with their absolute and relative frequencies from a Corpus [47]. The words included in the dictionary represent 17,672,326 words from the Corpus (if the Corpus were used as a sample text, it would generate a hit rate in the dictionary of 88%). The program emulates the VK behavior having some input parameters, as mentioned in section III. Furthermore, the context set in that section is considered by this emulation. Besides, this program uses a sample text selected from a collection of some kinds of documents: sports, religion, etc., with a total of 960,180 words and an average of 4.91 characters per word. The sample text is divided into 30 segments of an approximate equal size. For each segment, a set of parameters is gathered: frequency of letters, frequency of n-character words, probabilities vectors described in the above-mentioned sections, etc. Thus, each segment of the sample text represents different simulation conditions. In addition, the value of $\bar{t}_c$ and $\overline{UI}_c$ for each segment is also calculated. The mean hit rate is equal to 85%.

A Matlab program is also made to obtain the model results. This Matlab program reads the dictionary, builds statistical information from it, and calculates the value of $\bar{t}_c$ and $\overline{UI}_c$ according to the model. By using Matlab and C programs, the goodness of the model may be tested.

The simulation (C program) results (blue line) and model results (circles) are shown in Fig. 6. Furthermore, a simplification of the model has been carried out, and its results are shown by a red line with x-marks in this figure. Each segment is represented in x-axis, while the value of $\bar{t}_c$ in seconds is represented in y-axis. As it can be seen, the results from both the models are very close to simulation ones. The maximum relative error for $\bar{t}_c$ is 0.61% using the model or 3.26% using the simplification of the model.

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

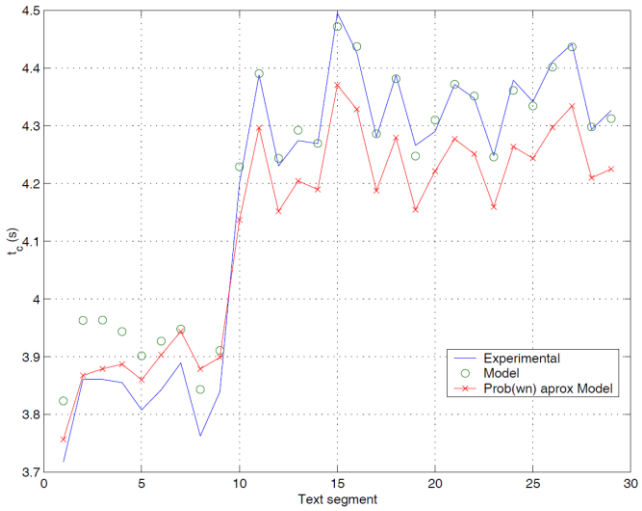


Fig. 6: Average text entry time per character for a RED4 keyboard obtained in each segment: simulation and model results.

The results for $\overline{UI_c}$ are shown in Fig. 7. As in the above-mentioned case, the results of the models are very close to those of the simulation ones: the maximum relative error for $\overline{UI_c}$ is 0.18% using the model or 4.14% using the simplification of the model.

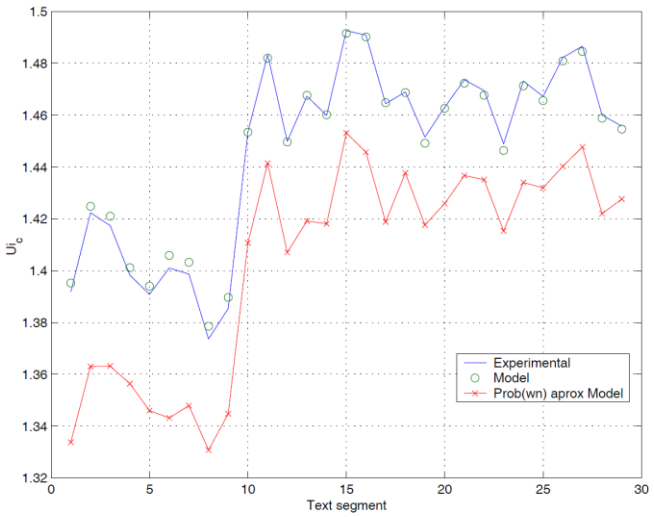


Fig. 7: Average number of user inputs per character for RED4 keyboard obtained in each segment: simulation and model results.

This trial has been repeated using some different VKs obtaining similar results. Thus, it can be concluded that these low relative errors are a proof of the goodness of the model and simplification.

4.2.4 Results

The performance of different ambiguous VKs has been compared in this section. The layouts of these VKs are shown in Appendix A. VKs are sized from 4, 6, 9, and 12 keys, and two different character arrangements are considered: PT and nonPT. In the first case, the time to access a key in Tn mode is used to place the most frequent characters. In the other case, the time to access a position in the letter-scanning method is used to set the character arrangement. In this sense, RED4PT represents an ambiguous VK with 4 keys and PT arrangement, and RED16 is related to a 16-key ambiguous VK with non-PT arrangement. Only a linear-scanning method is considered because of its low value of $\overline{UI_c}$.

The time required to enter a word with n characters (n-word) that are in the dictionary is shown in Fig. 8. Word length is represented in x-axis and the value of this time in seconds is represented in y-axis. In general, as VK ambiguity diminishes, time increases. This is owing to the fact that the scanning method needs longer time to reach different keys. As expected, the PT arrangement results are better in 9-key and 12-key VKs. However, it is completely opposite for 4-key and 6-key VKs. Taking into account the fact that the most frequent words have lengths between 2 and 9, RED4 and RED6 are found to present better results. It must be noted that RED4, RED4PT, and RED6PT show great values of this time, which makes the plot to move away from the constant slope in other cases. For RED4PT VK, this deviation is particularly important. These deviations are due to the fact that as ambiguity is increased, the number of words that are matched with a sequence of selected keys is increased, and hence, the length of the suggestion list is longer, and this involves a greater time to select the desired one. In the case of RED4PT, almost every most-frequent character is placed on the same key, and thus, the number of words that are matched with a sequence of selected keys is huge. Highly ambiguous VKs using a PT arrangement show worse results when compared with the others. Obviously, the probability of a word in the dictionary ($p(w \in \Omega)$) should be close to 1, and hence, the best results of $\overline{t_c}$ may be obtained by RED6, RED4, and RED6PT.

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

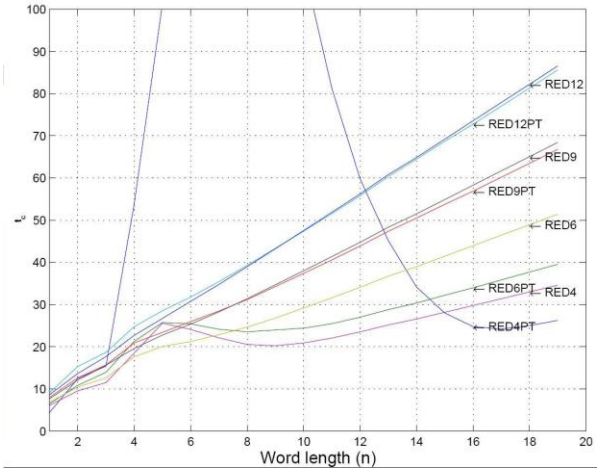


Fig. 8: Time required to enter a word with n characters that is in the dictionary using the studied VKs.

If $p(w \in \Omega)$ is moving away from 1, the letter-scanning mode is used. In such a situation, two cases may occur: 1. the sequence of the selected keys is matched with a word or some words in the dictionary, generating a suggestion list; or 2. this sequence is not matched with any words in the dictionary. Therefore, the time to type a n-word may be broken down into two terms: time to type in the first case, $\bar{t}_{fail\_List}(n)$, and time to type in the second case, $\bar{t}_{fail\_NoList}(n)$. Obviously, $\bar{t}_{fail\_List}(n)$ is greater than $\bar{t}_{fail\_NoList}(n)$. Both the terms are weighted using the probabilistic parameters. Thus, the best scenario is represented by the situation in which if the desired word is not in the dictionary, an empty suggestion list is always predicted. Fig. 9 shows $\bar{t}_{fail\_NoList}(n)$ for the VKs under study. In this situation, the best results are shown by RED6, RED4, and RED6PT.

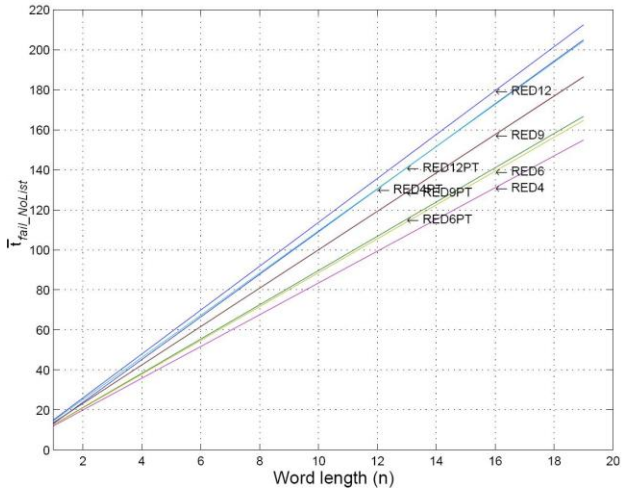


Fig. 9: Time required to enter a word with n characters that is not in the dictionary, using the studied VKs when a suggestion list is never generated.

On the other hand, the worse scenario, where a suggestion list is always predicted when the desired word is not in the dictionary, is shown in Fig. 10. In this case, the user always has to wait for the scanning of the suggestion list before starting to type in letter-scanning mode. The behavior of VKs is similar to that shown in Fig. 8. It must be noted that RED6 continues to be a good choice in the range between 1 and 0 n-words. RED9 and RED9PT also show a good behavior for $\bar{t}_c$ indicator.

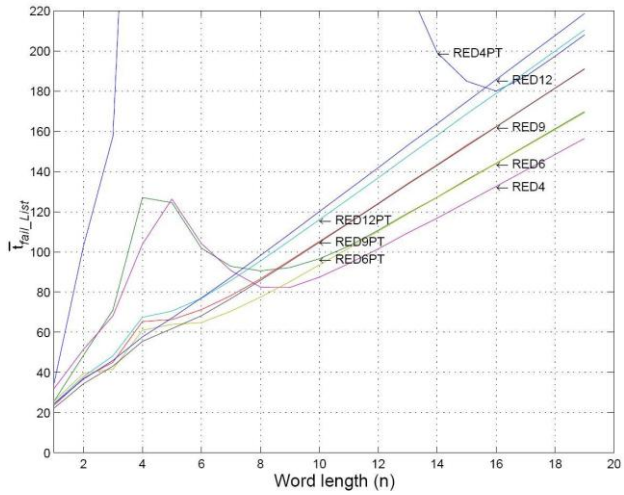


Fig. 10: Time required to enter a word with n characters that is not in the dictionary, using the studied VKs when a suggestion list is always generated.

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

Fig. 11 shows the number of user inputs when the desired word is in the dictionary. Besides, the best VK considering $\overline{UI_c}$ may be set from this figure, which is independent of the value of $p(w \in \Omega)$, because the number of user inputs is the same in any case. The best results are obtained for RED12, RED9, RED6, and RED12PT VKs.

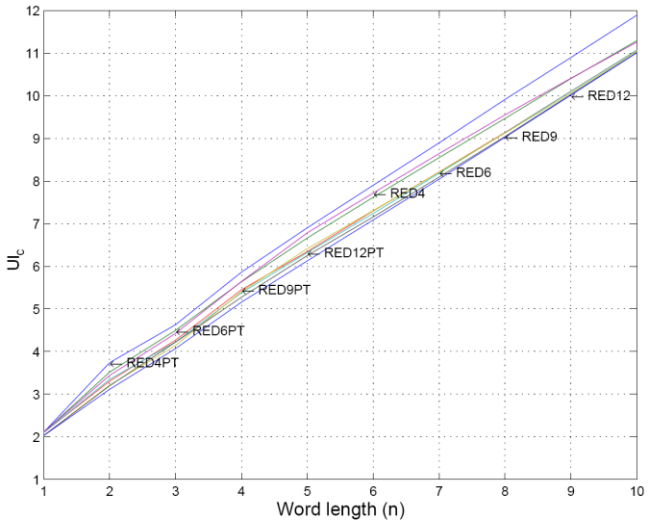


Fig. 11: Number of user inputs when the desired word is in the dictionary.

V. SELECTING A VK

In the previous sections, a study on the behavior of different VKs according to $\overline{UI_c}$ and $\overline{t_c}$ have been presented. The model can be used to estimate $\overline{t_c}$ and $\overline{UI_c}$ for any VK and $p(w \in \Omega)$. A relationship between these parameters has to be established to select the best VK under certain conditions. In this sense, the best option may be defined as the VK that reduces both the parameters. A function F that measures the VK performance may be built. This function must consider user preferences in relation to the text entry rate and the number of user inputs. Accordingly, two variables have been included, a_1 and a_2. Their values could change from 0 up to 1, and a_1 is related to the text entry rate. In this way, a value that is close to 1 shows that the user gives much relevance to the text entry rate. On the other hand, a_2 is related to the number of user inputs. In a similar way, a value that is close to 1 shows that the user gives much relevance to the number of user inputs. These estimations can be employed in Equation 8, so that a numerical value to F_k for the keyboard k can be obtained depending on a_1 and a_2.

$$F_k = a_1 \cdot \frac{\bar{t}_{c_k}}{\max_{\forall j}\left(\bar{t}_{c_j}\right)} + a_2 \cdot \frac{\overline{UI}_{c_k}}{\max_{\forall j}\left(\overline{UI}_{c_j}\right)} \qquad (8)$$

In this equation, the text entry time per character using a given VK, k, is represented by $\bar{t}_{c_k}$. The number of user inputs per character using a given VK is represented by $\overline{UI}_{c_k}$. Both the parameters are scaled according to the maximum values associated with the set of VKs in this study. Subsequently, given a_1 and a_2, the best VK would be the having the lowest F_k.

The function F_k can be made into a bi-dimensional matrix according to values assigned to a_1 and a_2 parameters. For example, if a_1 and a_2 change from 0 up to 1 by 0.1 steps, the matrix will be 11 x 11. The following figures are representations of the comparison results among the F matrices obtained for all the VKs, according to a_1 and a_2 value. The figures show the areas where a VK is the best or have the lowest F_k value. For example, we can see from Fig. 17 that there exists four areas, and therefore, four VKs achieve better performance when $p(w \in \Omega)$ =1. If a user prefers to obtain better type speed (a_1 close to 1) and does not care the number of inputs, an RED4 VK must be chosen. If another user does not care about the type speed and wants to minimize the number of inputs (a_2 close to 1), a CONVL must be chosen. If a third user wants to minimize both the parameters, an RED6 VK must be selected.

As mentioned earlier, in Tn mode, $\overline{t_c}$ and $\overline{UI_c}$ are estimated by the length of the word and some probabilistic parameters are required. Therefore, to predict the value of $\overline{t_c}$ and $\overline{UI_c}$ using a given VK, it is necessary to set the value of these parameters: the probability of typing a n-word, $p(w_n)$, the probability of a sequence of the selected keys generating a suggestion list when the desired word is not in the dictionary, x, and the probability of the desired word is in the dictionary, $p(w \in \Omega)$. First, $p(w_n)$ may be set using the information obtained in the validation. To set an appropriate value of x, it is necessary to take into account the VK layout. For example, in a 4-key VK, the probability of a sequence of the selected keys generating a suggestion list is almost 1, because

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

of the great amount of characters associated with a key. However, for 9-key and 12-key VKs, this probability is close to 0. This is owing to the fact that all words included in the dictionary are matched with only a sequence, and hence, the probability of a sequence that is not matched with a word in the dictionary generates a suggestion list is practically 0. For this reason, the following approximation may be carried out: for 4-key VKs, x = 1, for 9-key and 12-key VKs, x = 0, and for 6-key VKs, x = 0.5. Finally, values from 0 to 1 may be assigned to $p(w \in \Omega)$ to contemplate different conditions.

In this study, the following VKs were included: CONVL, CONV, RED4, RED6, RED6PT, RED9, RED9PT, RED12, and RED12PT, which comprise, all VKs in Appendix A, excluding 4PT. For each VK, the values of F considering different values of a_1 and a_2 have been estimated. In each case, a value of $p(w \in \Omega)$ has been set. In this way, the areas in function of the values of a_1 and a_2 may be defined. Each area represents a VK having the lowest value of F in relation to others. A special area is represented by coordinates a_1, a_2 in [0.9; 1], in which both the parameters, $\overline{t_c}$ and $\overline{UI_c}$, should be minimized. This area is called Maximum Interest Area (MIA).

By assuming a $p(w \in \Omega)$ =0.5 (Fig. 12) only CONVL, CONV, and RED6 VKs may be offered to a user depending on their preferences. By focusing on the piece of area delimited by coordinates a_1, a_2 in [0.9; 1], CONVl can be selected.

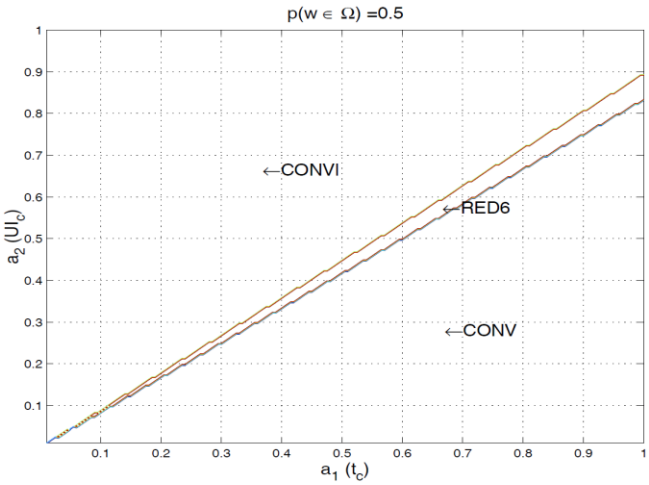


Fig. 12: Areas with $p(w \in \Omega)$ =0.5.

Areas assuming a $p(w \in \Omega)$ =0.6 are shown in Fig. 13. An increase in RED6 area is presented, reducing both CONV and CONVl areas. In this case, using values of a_1, a_2 in [0.8; 1], RED6 could be selected, which includes MIA.

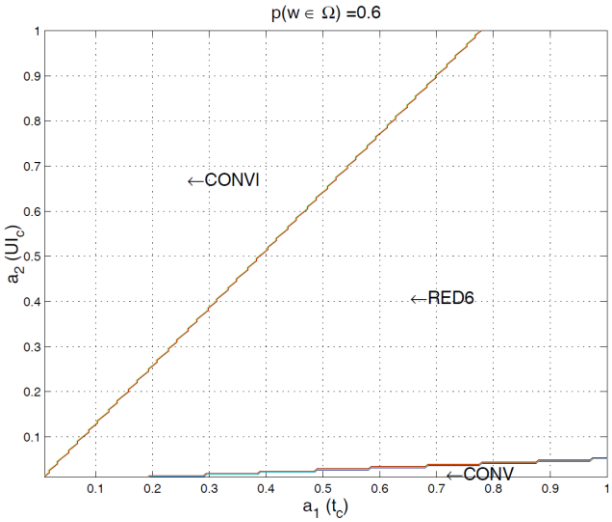


Fig. 13: Areas with $p(w \in \Omega)$ =0.6.

The CONV area disappears when $p(w \in \Omega)$ >0.6 as it may be seen in Fig. 14 - 17. Furthermore, as $p(w \in \Omega)$ increases, the CONVl area decreases. RED6 could be selected for values of a_1 > 0.5 assuming $p(w \in \Omega)$ =0.7, for values of a_1 > 0.3 assuming $p(w \in \Omega)$ =0.8 and for values of a_1 > 0.15 assuming $p(w \in \Omega)$ =0.9.

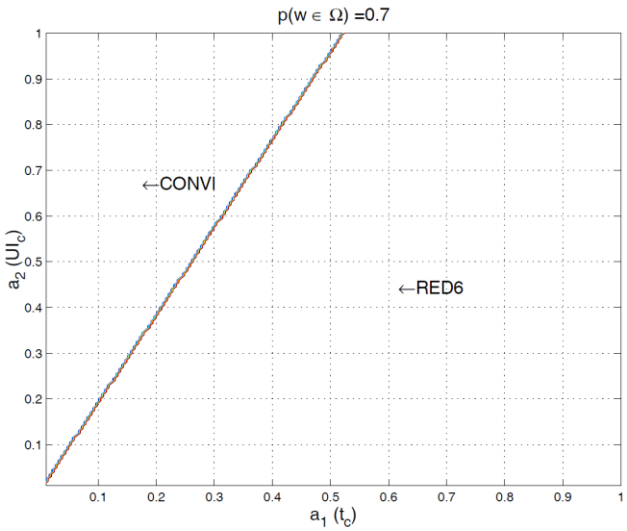


Fig. 14: Areas with $p(w \in \Omega)$ =0.7.

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

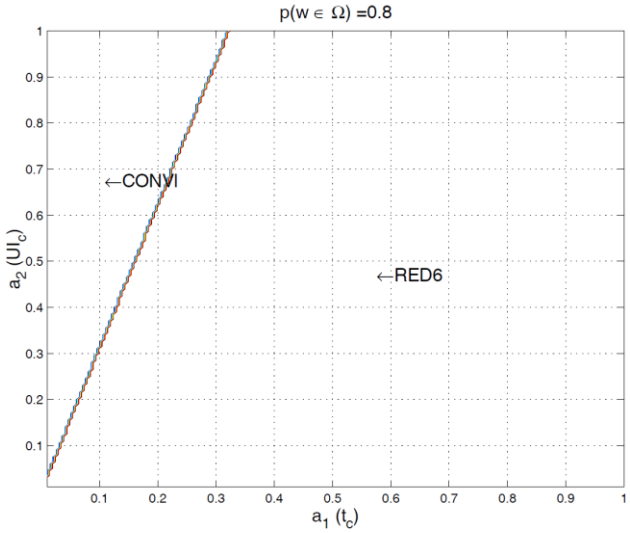


Fig. 15: Areas with $p(w \in \Omega)$ =0.8.

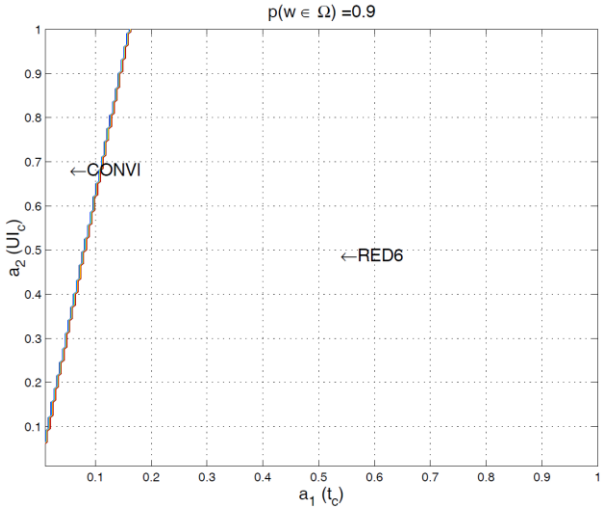


Fig. 16: Areas with $p(w \in \Omega)$ =0.9.

By assuming $p(w \in \Omega)$ =1, where the desired word is always in the dictionary, CONVl, RED12, RED6, and RED4 may be selected. The RED6 area is the greatest one and is placed in the central position. However, for low values of a_2, RED4 is the most likely option.

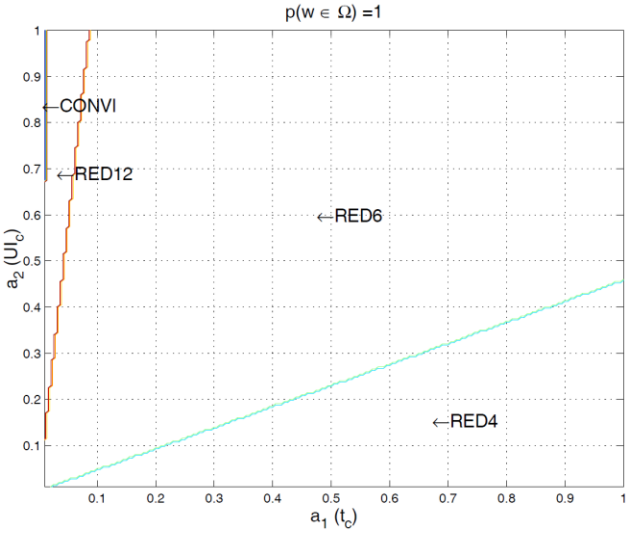


Fig. 17: Areas with $p(w \in \Omega)$ =1.

Furthermore, the effect due to assigning different values of x has also been checked. By assuming that $p(w \in \Omega)$ >=0.8, no changes have been found on the previously obtained values. For lower $p(w \in \Omega)$ <0.8, RED9 or RED12 might replace RED6 on AMI.

VI. CONCLUSIONS

In this paper, some ambiguous VK layouts have been studied. The text entry rate using these VKs in Tn mode is greater than the one using an unambiguous VK. In this mode, a dictionary, a NEXT function, and a SPACE function are required. Furthermore, the convenience of having the SPACE and NEXT key separately on the VK without additional functions or characters in them has also been discussed. Besides, it has been tested that PT arrangement neither improves $\overline{t_c}$ nor $\overline{UI_c}$.

In addition, a model based on the dictionary has been developed. The proposed model allows estimating the values of both the $\overline{t_c}$ and $\overline{UI_c}$. A simple user model based on KLM has been used to estimate the value of t_r. This fact allows making comparison independently of the used input device. The 0.65 rule has been used to estimate a proper value of the dwell time. No simulations are required to compare different VKs. The values of some needed parameters have been set using the information in the dictionary to compare them, and model validity has been demonstrated even in its simplified version. The model allows us to

www.ijmer.com Vol.1, Issue.2, pp-288-305 ISSN: 2249-6645

understand how VKs behave according to the changes in $p(w \in \Omega)$. When using this model, participation of real users in trials is not necessary. Hence, a prototype of the system is not required, and the behavior of any input device can be emulated. A function that assesses the user preferences on $\overline{t_c}$ and $\overline{UI_c}$ indicators and allows comparison of VKs has also been proposed. According to that function, the RED6 VK has been found to show better performances even in a wide range of $p(w \in \Omega)$.

APPENDIX A: CONSIDERED VKs

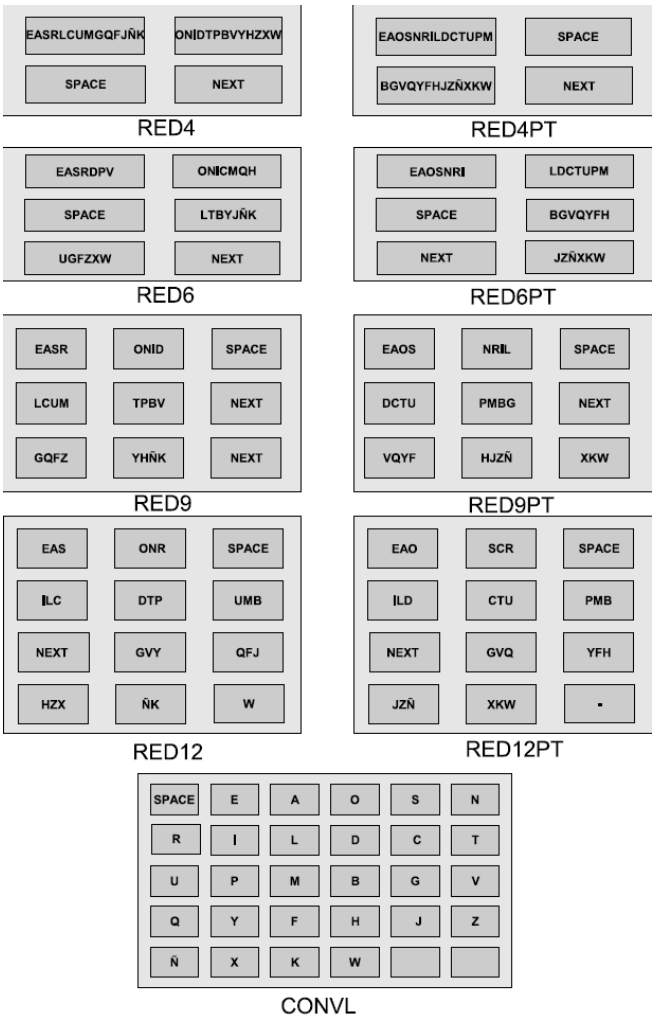


Fig. 18: Considered VKs.

REFERENCES

[1] Kushler, "AAC using a reduced keyboard," *Proceedings of CSUN* 98, 1998.

[2] K.-N. Kim and R. S. Ramakrishna, "Vision-based eye-gaze tracking for human computer interface," vol. 2, 1999, pp. 324–329.

[3] B. Noureddin, P. D. Lawrence, and C. F. Man, "A non-contact device for tracking gaze in a human computer interface," Computer Vision and Image Understanding", vol. 98, n0. 1, pp. 52-82, 2005, special issue on Eye Detection and Tracking. Available: http://www.sciencedirect.com/science/article/B6WCX-4D99BSP-1/2/24b89233ae82e7d1385951212a3151a7

[4] Y.-L. Chen, F.-T. Tang, W. Chang, M.-K. Wong, Y.-Y. Shih, and T.-S. Kuo, "The new design of an infrared-controlled human-computer interface for the disabled," Rehabilitation Engineering, IEEE Transactions on, vol. 7, no. 4, pp. 474–481, Dec 1999.

[5] D.G. Evans, R. Drew, and P. Blenkhorn, "Controlling mouse pointer using an infrared head-operated joystick", IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 1, pp. 107-117, Mar 2000.

[6] O. Rivera, A. J. Molina, and I. M. Gómez, "Versatile system to computer control based on infra-red light," 24th Human Factors and Ergonomics Society Europe Chapter. Num. 24. Shaker Publishing, pp. 4–5, 2007.

[7] I.M. Gómez, A. J. Molina, O. Rivera, "Multipurpose interface for handling general computer applications", AAATE 2007, 9th European Conference for the Advancement of Assistive Technology in Europe, pp. 7-8, 2007.

[8] F. L. Castro, "Class I infrared eye blinking detector", Sensors and Actuators A: Physical, vol. 148, no. 2, pp. 338-394, 2008. Available: http://www.sciencedirect.com/science/article/B6THG-4TDK6SC-4/2/6826bf33afa9efa561f71c97b3992b20

[9] P. O'Neil, C. Roast, and M. Hawley, "Evaluations of scanning user interfaces using real-time-data usage logs", Proceedings of the fourth international ACM conference on Assistive technologies (ACM SIGACCESS Conference on Computers and Accessibility), pp. 137-141, 2000.

[10] N. Alm, J. L. Arnott, and A. Newell, "Prediction and conversational momentum in an

augmentative communication system." Comm ACM, pp. 46–57, 1992.

[11] H. Horstmann, E. Lopresti, and R. C. Simpson, "Toward automatic adjustment of keyboard setting for people with physical impairments." Diability and Rehabilitation: Assistive Technology., pp. 261–274, 2007.

[12]] K. Harsbush and M. Khn, "An evaluation study of two-button scanning with ambiguous keyboards," Proceedings AAATE, pp. 954–958, 2003.

[13] P. Gnanayutham, C. Bloor, and G. Cockton, "Soft keyboard for the disabled," Computers Helping People with Special Needs, Lecture Notes in Computer Science, 2004.

[14] G. W. Lesher, B. J. Moulton, and D. J. Higginbotham, "Techniques for augmenting scanning communication," International Society on Augmentative and Alternative Communication(ISAAC), no. 2, pp. 81-101, 1998.

[15] T. Bellman and I. MacKenzie, "A probabilistic character layout strategy for mobile text entry," In Proc. Graphics Interface, pp. 168–176, 1998. [Online]. Available: citeseer. comp.nus.edu.sg/bellman98probabilistic.html

[16] I. S. MacKenzie, H. Kober, D. Smith, T. Jones, and E. Skepner, "Letterwise: prefix-based disambiguation for mobile text input," in UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology. New York, NY, USA: ACM, 2001, pp. 111– 120.

[17] J. Arnott, Probabilistic character disambiguation for reduced keyboards using small text samples. Taylor & Francis, September 1992, vol. Augmentative and Alternative Communication. Available: http://www.ingentaconnect.com/content/tandf/taac/1992/00000008/00000003/art00004

[18] H. Horstmann and S. Levine, "Modeling the speed of text entry with a word prediction interface," IEEE transactions on rehabilitation engineering, vol. 2, pp. 177–187, 1994.

[19] Y. Gillette and J. L. Hoffman, "Getting to word prediction: developmental literacy and aac." Proceedings of the 10th annual international conference, technology and persons with disabilities, 1995.

[20] B. Heinisch and J. Hecht, "Predictive word processors: a comparison of six programs." Tam News., pp. 4–9, 1993.

[21] N. Garay-Vitoria and J. Abascal, "Text prediction: a survey," Univ Access Inf Soc, pp. 188–203, 2006.

[22] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement." J Exp Psychol, vol. 47, no. 6, pp. 381–391, June 1954. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/13174710

[23] M. Silfverberg, I. S. MacKenzie, and P. Korhonen, "Predicting text entry speed on mobile phones," in CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems. New York, NY, USA: ACM, 2000, pp. 9–16.

[24] I. S. Mackenzie, S. X. Zhang, and R. W. Soukoreff, "Text entry using soft keyboards," Behaviour & Information Technology, vol. 18, pp. 235–244, 1999.

[25] R. W. Soukoreff and I. S. MacKenzie, "Theorical upper and lower bounds on typing speeds using a stylus and soft keyboard." Behavior & Information Technology, pp. 370–379, 1995.

[26] I. S. MacKenzie and R. W. Soukoreff, "A model of two-thumb text entry." Proceedings of Graphics Interface, pp. 117–124, 2002.

[27] A. Pavlovych and W. Stuerzlinger, "Model for non-expert text entry speed on 12-button phone keypads." Proceedings of the SIGCHI Conference of Human Factors in Computing Systems., pp. 351–358, 2004.

[28] P. Isokoski and I. S. MacKenzie, "Combined model for text entry rate development." Proceedings of the ACM CHI 2003 Conference on Human Factors in Computing Systems, pp. 752–753, 2003.

[29] H. Horstmann and S. P. Levine, "Validation of a keystroke-level model for a text-entry system used by people with disabilities." Proceedings of the First ACM Conference on Assistive Technologies., pp. 115–122, 1994.

[30] H. Horstmann and S. P. Levine, "Effect of a word prediction feature on user performance." AAC Augmentative and Alternative Communication, pp. 155–168, 1996.

[31] H. Horstmann, "Keystroke-level models for user performance with word prediction," ISAAC, pp. 239–257, 1997.

[32] H. Horstmann and S. P. Levine, "Model simulations of performance with word prediction." Augmentative Alternative Communication., pp. 25–35, 1998.

[33] M. Baljko and A. Tam, "Indirect text entry using one or two keys." Proceedings of the 8th international ACM SIGACCESS Conference on Computers and Accessibility, pp. 18–25, 2006.

[34] A. J. Molina, O. Rivera, I. M. Gómez, and G. Sánchez, "Evaluation of unambiguous virtual keyboards with character prediction," AAATE 2009. Assistive Technology from Adapted Equipment to Inclusive Environments, pp. 144–149, 2009.

[35] H. Ryu and K. Cruz, "Letterease: Improving text entry on a handheld device via letter reassignment," in OZCHI '05: Proceedings of the 17th Australia conference on Computer-Human Interaction. Narrabundah, Australia, Australia: Computer-Human Interaction Special Interest Group (CHISIG) of Australia, 2005, pp. 1–10.

[36] I. S. MacKenzie and S. X. Zhang, "The design and evaluation of a highperformance soft keyboard," in CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems. New York, NY, USA: ACM, 1999, pp. 25–31.

[37] K. Tanaka-Ishii, Y. Inutsuka, and M. Takeichi, "Entering text with a four-button device," in Proceedings of the 19th international conference on Computational linguistics. Morristown, NJ, USA: Association for Computational Linguistics, 2002, pp. 1–7.

[38] S. K. Card, T. P. Moran, and A. Newell, "The keystroke-level model for user performance time with interactive systems." Communications of the ACM, pp. 396–410, 1980.

[39] S. K. Card, T. P. Moran, and A. Newell, The psychology of human-computer interaction., 1983.

[40] A. T. Welford, "Choice reaction times: Basic concepts," Reaction Times. London Academic Press, pp. 73–128, 1980.

[41] W. Hick, "On the rate of gain of information," Journal of Experimental Psychology, vol. 4, pp. 11–36, 1952.

[42] J. Carroll, HCI models theories and frameworks: towards a multidisciplinary science. Kaufmann Publishers, 2003.

[43] R. Simpson and H. Horstmann, "Adaptive one-switch row-column scanning," IEEE Transactions on rehabilitacion engineering vol7. N 4, no. 4, pp. 464–473, 1999.

[44] R. Simpson, H. Horstmann, and E. F. Lopresti, "Selecting an appropriate scan rate: the .65rule," Proceedings of RESNA 2006 Annual Conference, Atlanta, GA. Arlington, VA: RESNA Press, 2006.

[45] G. W. Lesher, D. J. Higginbotham, and B. J. Moulton, "Techniques for automatically updating scanning delays," In Proceedings of the RESNA 2000 Annual Conference, pp. 85–87, 2000.

[46] H. S. Venkatagiri, "Efficient keyboard layouts for sequential access in augmentative and alternative communication," Augmentative and Alternative Communication, 15(2), pp. 126–134, June 1998.

[47] R. Almela, P. Cantos, A. Sánchez, R. Sarmiento, and M. Almela, Diccionario y estudios léxicos y morfológicos. Universitas SA, 2005. [Online]. Available: http://www.um.es/lacell/proyectos/dfe/

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-306-311 ISSN: 2249-6645

Geographic Information System: A Conceptual Enterprise Model for Bangalore Metropolitan City

M. Raghunath[1], B. Shankar[2]
[1](Project Officer SPFU-TEQIP Directorate of Technical Education, Government of Karnataka, Bangalore)
[2](Associate Professor in Urban and Regional Planning, Institute of Development Studies, University of Mysore, Mysore)

ABSTRACT
Geographical Information System (GIS) is an effective tool for planning and management of a metropolitan city. In developing countries like India, metropolitan governments, planning authorities and parastatals (viz. Water Supply Board, Transport Corporation etc.) have executed GIS projects independently. In recent years, many municipalities in the World have switched from stand-alone GIS systems to integrated approaches that share resources and applications. Enterprise GIS is an organization-wide approach, integrates spatial data and technology across the different departments of an organization coupling centralized management with decentralized use. Geographic Information System projects have been implemented in Bangalore City by the different stakeholders independently, without much common resources and integration. This calls for integrating both spatial and non-spatial data of all the stakeholders for effective planning, governance and management. This paper presents experiences of GIS implementation in Bangalore City and suggests a conceptual Enterprise GIS Model for Bangalore.

***Key Words*: Enterprise GIS, Integration, Metropolitan City Planning, Spatial**

I. INTRODUCTION
Geographic Information System (GIS) has been in use in local government since many years for isolated application such as tax collection, town planning etc. Enterprise GIS is considered to be the highest level of GIS development, which involves a large scale data, provides an information and operational framework for major portions of the activities and applications within an organization or consortium of organizations. The term 'enterprise' refers to looking at the entire organization as a single entity. An Enterprise GIS approach will provide a framework for integration of the requirements of all the departments of municipal governments including collection of data, sharing of information, collaborating and conducting cross-departmental analysis. The Enterprise GIS (EGIS) also, extends both vertically to state and central governments and horizontally to other organizations, stakeholders and parastatals (ESRI, 2003).

II. BACKGROUND OF BANGALORE
Bangalore is the third largest metropolitan city with a population of over 9.5 million (2011 provisional figures of Census of India) rose from 5.7 million in 2001. The City is an Information Technology (IT) hub and is also called as Silicon City of India. The City Government was called Bangalore Mahanagara Palike (BMP), with 100 wards, had an area 226 sq.km. It was renamed as Bruhat Bangalore Mahanagara Palike (BBMP) in 2007, by increasing its area to 800sq.kms and wards to 147 wards. The Bangalore Development Authority (BDA) is a statutory Planning Authority, has a Local Planning Area (LPA) namely Bangalore Metropolitan Area (BMA), with an extent of 1279 sq.kms. The Bangalore Metropolitan Regional Development Authority (BMRDA) was established in 1985 having its jurisdiction of Bangalore Urban, Rural and Ramanagaram districts. Bangalore Metropolitan region Development Authority (BMRDA) was set up for the purpose of planning, coordinating and supervising the proper and orderly development of the area within the Bangalore Metropolitan Region (BMR) with Local Planning Areas to the extent of 8005 sq.km.

III. INITIATION OF GIS IN BANGALORE
The BMRDA engaged Indian Resource Information and Management Technologies (INRIMT), Hyderabad for preparation of structure plan for metropolitan region in 1994. The satellite photography was procured from the Indian Remote Sensing (IRS) Programme, available through National Remote Sensing Agency (NRSA), which has influenced the urban planning in Bangalore for the first time. The INRIMT constructed the base map 1:250,000 scale from Survey of India topographic sheets including the boundaries of BMR. The overlays for different variables including topography, drainage, water bodies, ground cover, annual rainfall, geology and mineral, ground water table, and transportation network, water supply, sewerage and electricity lines etc are some of the prominent features. The BMRDA set up a GIS centre by hiring a consultant to support the System and installed hardware and software along with the technical personnel. Similarly, Bangalore Development Authority also initiated Integrated Urban Environmental Improvement Project (IUEIP), which covered neighbourhoods in four layouts in 1998 for over a

period of 3 years. The GIS was initiated for water distribution system in Bangalore Water Supply and Sewerage Board (BWSSB) for by engaging the French consultants, Strategie-Conceptions-Erudes (SCE) in 1999 and implemented in 2001. Bangalore Mahanagara Palike (BMP) also initiated GIS project in a selected wards on a pilot basis in 2002. The details of the Implementation of GIS of BMP, BWSSB, BDA and BMRDA are as follows:

3.1 Bangalore Mahanagara Palike (BMP)

A pilot GIS project was initiated in two wards of BMP in 2002 covering two layers viz. property and road as shown in figure 1 and figure-2, by the Indian Space Research Organization (ISRO).

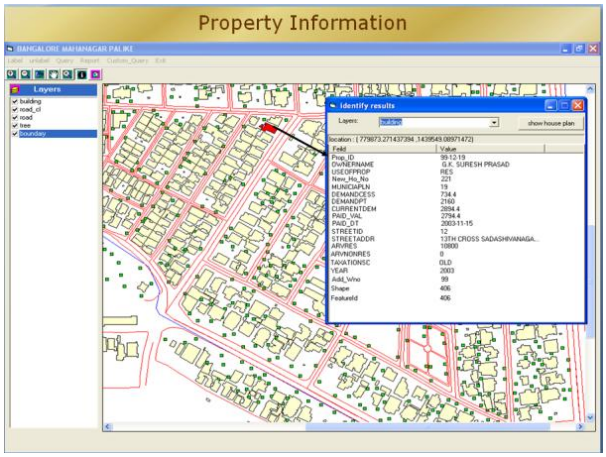


Fig.1: Property Information of BMP Ward No-99

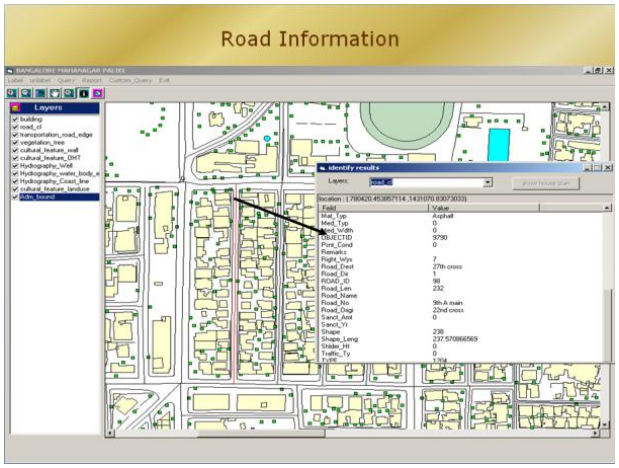


Fig.2: Road Information of BBMP Ward No: 60
(Source: BBMP, Bangalore)

The objective was to develop customized GIS software and the Management Information System (MIS). In 2004, the Karnataka State Remote Sensing Applications Centre (KSRSAC) conducted another pilot project for BMP in Ward No.64 by capturing 15 features namely roads, properties, storm water drain, cross drainage works, streetlights, trees, public taps, location of water tanks, public taps and bore-wells etc., covering four important departments namely Revenue, Health, Horticulture and Engineering Departments. Quick Bird satellite image of 0.6m spatial resolution captured in 2004 and the image was used for creating the base map. In the pilot project, tasks of physical surveying, collection of attributes of various layers, digital photographing, data validation at the field and linking attribute and spatial data were carried-out.

Later, BMP extended the geo-database model, spatial data for 29 wards of BMP involving KSRSAC as coordinating agency with the support from the National Remote Sensing Agency (NRSA). The data model is represented in the fig.5. However, only 29 wards were covered out of 100 wards of BMP in the first phase of the project. As of now, the data validation is completed in few wards. Further, BMP GIS model excluded Solid Waste Management, Emergency Services, Crime, Education, Health Services, Election, Property Management, Poverty Management, Housing etc. which are of utmost importance for urban governance.

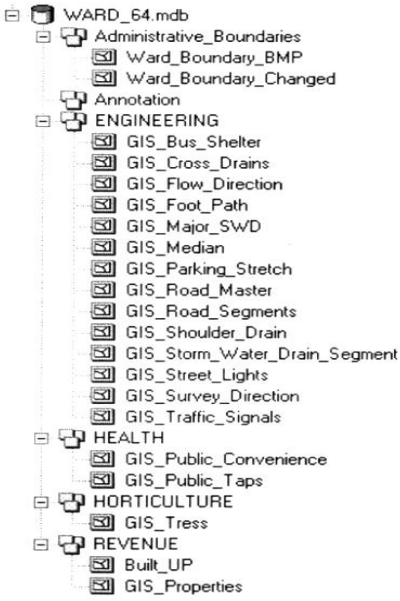


Fig. 3: Geo-database Model Adopted by BMP

3.2 Bangalore Water Supply and Sewerage Board

The Bangalore Water Supply and Sewerage Board (BWSSB) has implemented GIS for its water supply and sanitary services in the year 2001.The base map was prepared using aerial photographs supplied by NRSA. The scale of the base map was 1:2000. The pilot project covered an area of 290 sq.kms, 4000 kms of water pipelines, 3000 kms of sewerage lines and 320,000 consumer connections in the BWSSB jurisdiction. It was linked to the attribute data viz. water pipeline, valve, fire hydrant, reservoir, pumping stations, BWSSB offices, storm water drain, manholes, sewer pipe, service station boundary, consumer point etc. The employees' database along with the billing system was integrated in GIS to improve the bill collection and transparency among the organisations. Data integration was achieved through data entry of alphanumerical data using Excel sheets; report of graphical data collected on

the field on reference base maps (A0 size – 1: 1200 scale); conversion of NRSA base maps into the ARCINFO /ARCFM GIS; digitization of graphical data; ARCINFO conversion after Quality Control; corrections and validation by BWSSB and final integration of data after Topology Building. 98% of attribute data was collected directly from the field.

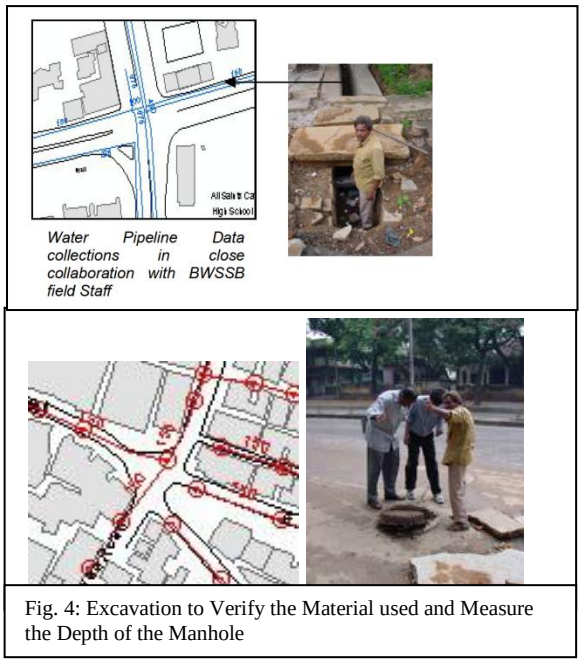


Fig. 4: Excavation to Verify the Material used and Measure the Depth of the Manhole

3.3 Bangalore Metropolitan Transport Corporation

The Bangalore Metropolitan Transport Corporation (BMTC) implemented the 'Passenger Information System (PIS)'. The PIS helps the passengers, who are travelling in BMTC buses, to get real-time information about arrival of buses at stops enroute. The corporation has introduced GPS on 1,500 buses and eventually brings all its 4,000-odd buses under GPS. The display unit, was placed at a noticeable place in the bus shelter, would digitally display arrival timing of the buses enroute. The unit was connected to the control room through GPRS. The corporation was to able provide the route map, timetable, number of buses on a particular route and other required details to the service provider.

3.4 Bangalore Metropolitan Regional Development Authority

In 2001, BMRDA established the GIS centre, which houses a spatial base map at 1:50000 including land use/land cover, drainage, water bodies, irrigation systems, contours and slopes, land geomorphology and soils, roads, rail, electricity networks and administrative boundaries. These data were essentially used for the preparation of structural and development plans. The base map of very small scale was used for urban applications.

3.5 Bangalore Development Authority

The Bangalore Development Authority (BDA) initiated Metropolitan Spatial Data Infrastructure project (MSDI) in partnership with Groupe SCE. This project was funded under an Indo-French Protocol signed between the French and the Indian Government.

The MSDI project is thus a unique "spatial data" vehicle developed for Greater Bangalore to address various issues, like the CDP revision, and to help create physical infrastructure through the deployment of a multidisciplinary approach: IT tools and GIS applications. The governing principle of the MSDI project as well as its main challenge is to build, in parallel, a Sustainable Urban Geographical Information System along with a renewed approach to city planning through the CDP revision. That is to say to provide, collect, organize and standardize all kinds of data in a mega urban database ranging from satellite images to building footprints that interest all public and private stakeholders. This database must be both spatially meaningful and usable to all kinds of experts as well as to the building of the sustainable common asset and spatial repository acting as the spatial memory of the territory. This intertwining of scales (metropolitan, city, ward, village, parcel, building), dates, issues (urban planning versus urban management) and the variety of actors make the MSDI project a showcase of GIS technology applied to a complex mega city like Bangalore.

One hundred and ten experts were mobilised comprising of town planners, architects, economists, demographers, sociologists, GIS & IT specialists, geographers, cartographers, and infrastructure and transport specialists

The main challenge and principle of the MSDI project was to build a sustainable urban geographical information system through the creation, collection, organization and standardization of huge amounts of data from over 30 public and private sources into mega urban spatial database and make use of this GIS to modify and streamline the planning process thereby making it highly efficient. This spatial repository would become a common asset to the stakeholders acting as the 'spatial memory' of the territory.

Table-1: List of Geodatabases Created for MSDI of BDA

Geo-database	Contents
Boundaries	The local administrative limits of various stakeholders
Topography	Geology, hydrography and relief information
Transportation	Network of existing railway lines, roads and proposed metro
Landmarks	Extensive list of well known landmarks
Land use	Existing land use situation as well as proposed land use for 2015
Housing_land	Buildings, development layouts, slums, urban fabrics etc
Socio_Economic	Tables containing information on demographic, health, education etc.
Environment	State forest limits, borewell distributions and depths

Utility_services	Bangalore Water and Sewerage Board network, power lines, oil pipelines
Raster_Photos	Historical maps, satellite images, DEM etc.
Cdp_support	Various supporting elements for the mapping of the Existing and proposed Land use
Technical	Support database containing grids and other templates

The database model was created keeping in mind the simplicity for use and flexibility for evolution and growth. Since the platform chosen for the GIS was ESRI's ArcGIS(R), the spatial repository was designed as multiple personal geodatabases as given in the table-1. The MSDI project was the first GIS project of this scale that the BDA was undertaking. Therefore, the choice of using simpler personal geodatabases instead of larger geodatabase along platforms like Oracle(R) was made in order to avoid complexity of managing databases for the clients.

The DUSR comprises of a digital geo-referenced large scale map (scale 1:2000) covering an area of Bangalore is 1500 km^2, 553 villages, 55,000 parcels, 6.5 lakh buildings, 15.5 km of roads, 230 km of railway, 2546 places of worship, 100691 manholes, 330,903 consumer connections, 4008 km of water pipe lines, 3245 km of sewage lines, 8115 km of drains, 450 km of HT, 90,000 parcels of existing land use, 15,000 various landmarks and 400 layouts totalling 2 GB of records in 12 geo-databases. Attribute data gathered from 1991 & 2001 Census, BWSSB, BBMP, the Slum Clearance Board and others –totally 27 stakeholders. The database model of DUSR is yet to be commissioned due to ongoing development of IT Applications.

As it is evident stakeholders such as BMRDA, BBMP, BDA, BWSSB, BMTC etc. have executed GIS projects independently without having any common spatial reference except the MSDI initiative. Also, the GIS project initiated by BMP is yet to be commissioned and may not be able to meet the present day challenges of temporal, administrative, functional, technical, jurisdictional and institutional transformations. However, some of the huge amount of field data already collected and validated can be used appropriately.

IV. NEED FOR ENTERPRISE GIS

The GIS projects were implemented by few stakeholders in Bangalore Metropolitan Area by limiting to few requirements. Even though, Bangalore Mahanagar Palike initiated the first pilot GIS project with the help of ISRO, but it was restricted to roads and properties. The second pilot project was initiated by involving KSRSAC and NRSA covering 15 features of few departments in BMP. Few of the utility and service organisations namely BWSSB, BMTC, BESCOM etc., have developed their MIS system integrating the GIS limiting to their few departmental needs. The GIS developed by BDA and BMRDA are not synchonised and oprationised. The City Government is yet form Metropolitan Planning Committee, which can act as a coordinating agency as per the 74th CAA of all the stakeholders coming within the jurisdiction of the Bangalore Metropolitan Area. The BBMP's responsibilities are also increases with the new functions that are going to be assigned to the Mahanagara Palike in the context of 74th CAA. In the absence of robust integrated GIS, the planning, co-ordinating, resource mobilization and management tasks in metropolitan area is a difficult task which are not synchronizing for meeting the demands of the citizen. It is inevitable to have a coordinated and integrated GIS System with sound technologies not only to reduce the cost but also to increase the effectiveness in service delivery and good urban governance. The experiences of other municipalities in the world have proved that Enterprise GIS would be appropriate model for Metropolitan Bangalore.

V. SWOT ANALYSIS OF GIS PROJECTS

Strengths	Weaknesses
1.Stakeholders are willing to co-ordinate the GIS in Bangalore 2.GIS projects are already operationilised in Utility and Service Departments 3.Efforts have been made to collect large scale field data in GIS projects of BMP, BMRDA and BDA	1.There is no synchronization of data of different stakeholders 2. Many of the GIS Projects of BMP and BDA are not operationsed. 3. There is no mechanism for updating both spatial and attribute data. 4.Data collection is limited to few functional aspects of metropolitan governance
Opportunities	Threats
1.E-governance initiatives under JNNURM and State Government Reforms Projects for initiating Integrated and collaborative GIS	1.Non-constitution of Metropolitan committee

VI. A CONCEPUTAL EGIS MODEL FOR BANGALORE

The methodology for developing the conceptual model of Enterprise Geographic Information (EGIS) is as follows. The methodology for implementation of EGIS in any organization involves

[1]. Developing an organization-wide GIS approach using standards and consistent methodologies that address the needs of all units of the organization
[2]. Migrating existing GIS applications and data to current GIS technology capable of supporting all potential users in a cohesive manner
[3]. Integrating GIS data and services with other information systems within the organization as part of an overall enterprise information systems solution
[4]. Adapting the GIS staffing structure to support the enterprise approach
[5]. Training the IT and GIS staff to design, develop and maintain the enterprise GIS resources
[6]. Training staff in the departments new to GIS in the effective use of GIS specific to their business needs

VII. PROPOSED EGIS CONCEPTUAL MODEL

BDA, as a planning authority has felt the need for having a common digital base map which could be used by various stakeholders for their own independent GIS implementations. The base map was prepared by digitizing over a mosaic of large scale aerial photographs (source: NRSA) being geo-referenced with Quickbird satellite images consisting of 514 layers of information. Over 300 Survey of India (SOI) ground control points were used for geo-referencing of image. Data has been collected from 27 stakeholders and is organized into 80 GIS feature classes. Ground surveys were conducted to prepare existing land-use maps, water network, urban amenities, Central Business District (CBD), land marks, road widths etc. All the data is available at the BDA central repository for internal use as well as sharing with other stakeholders.

The urban governance in a Metropolitan city like Bangalore having an administrative area of 800 sq.km, and its demands an innovative new technology for faster and timely decision making. Embracing enterprise wide GIS at the earliest is the only possible option for good urban governance. BBMP and other stakeholders have the advantage of bringing their own data into the existing common spatial database. BBMP had already initiated the GIS programs but without much success. BBMP is headed by the Commissioner and assisted by two Special-Commissioners: one for Engineering Projects and the other for Administrative functions at the central office. The administrative boundary of BBMP, after formation of Greater Bangalore has been divided into 8 zones and new 147 wards altering the boundaries of earlier 100 wards. Each zone is governed by a Joint Commissioner and is responsible for the overall functions of various departments such as Revenue, Engineering, Town Planning, Health etc.

We propose a distributed network data model as shown in figure-5. The spatial data and the available attribute data necessary for the functioning of BBMP shall be replicated in an intermediate server from the DUSR of BDA. The central server, located at the Head office will have to be replicated and synchronized database from intermediate server. The EGIS-Coordinator controls the central server and co-ordinates the zonal GIS units. He also determines data to be updated back in the DUSR repository and publishes the relevant GIS information to public who are accessing through Internet. The central data repository is replicated and synchronized to zonal servers. The replication may be in full or partial with respect to changes in the database. The administrators, at the zonal offices maintain the zonal servers and coordinate various departments for data viewing, editing and updating the central server. Each zonal server is connected to departmental servers. Each department will have an expert who will view, edit and update by synchronization with the zonal server in their respective domain through field or office staff directly. In order to reduce the load on the communication network, a separate server may be maintained by each department for viewing and editing and another replica server for updating and synchronizing with zonal server. Thus zonal offices are responsible for the maintenance of geodatabase and the central office will have the compiled and updated version of the geodatabase which may be published to Public and perform municipal functions.

CONCLUSIONS

The GIS projects were executed independently by BBMP, BDA, BWSSB, BMTC and other stakeholders without having any common spatial data reference. Also, the GIS projects, which were initiated by BMP are unable to implement effectively due to the present day administrative, functional, technical, jurisdictional and institutional challenges. BBMP has grown both population-wise and area-wise in recent time. The responsibility and quantum of services to be delivered have drastically increased due administrative, functional, institutional changes. Therefore, BBMP is compelled to adopt geospatial technology to cope up with service delivery and timely decision making. The proposed Conceptual Model of Enterprise GIS with decentralized functions at zonal level would BBMP for achieving better planning, good urban governance and management.

REFERENCES

1. Aranya, Rolee (2003), Globalization and Urban Restructuring of Bangalore, India, 39th ISOCARP Congress, 2003
2. Bangalore Mahanagara Palike (2006), BMP GIS Project – Write Up, BMP, Bangalore, India
3. BWSSB (2002), Computerized Mapping, GIS Development and Utilities Computerization for Bangalore Water Supply & Sewerage Board (BWSSB), Bangalore, India
4. Commissioner, BDA (2003), Metropolitan Spatial Data Infrastructure (MSDI), Map India Conference 2003
5. ESRI White Paper (2003), Enterprise GIS for Municipal Government, ESRI, 380, New York, St. Redlands, CA.
6. ESRI White Paper (2007), An Overview of Distributing Data with Geodatabases, ESRI, 380, New York.
7. ESRI White Paper (2007), Enterprise GIS for Local Government, ESRI, 380, New York, St. Redlands, CA.
8. Heitzman, James (2003), Geographic Information Systems in India's 'Silicon Valley': The Impact of Technology on Planning Bangalore, *Contemporary South Asia, 12(1), 57-83*
9. Radwan ,M. Mostafa et.al, Designing an Integrated Enterprise Model to Support Partnerships in the Geo-Information Industry, *International Institute for Geo-Information Science and Earth Observation (ITC), Netherlands*
10. Safe Software Inc., Data Replication and Data Sharing – Integrating Heterogeneous Spatial Databases, (Safe Software Inc, 7445-132nd St., Surrey, B.C., Canada)
11. Subash S. & Arun Padaki (2003), Enterprise GIS for Municipalities – An Integrated Approach, *Map India Conference 2003*

12. Tomaselli , Linda (2004), The Enterprise Model of GIS and the Implications for people and organizations, TSU GIS 2004, Troy State University, Troy, Alabama, USA

ACKNOWLEDGEMENTS

The authors are greatly acknowledging the Institute of Development Studies, University of Mysore and Director of Technical Education, Government of Karnataka for encouragement and support.

BIOGRAPHIES

Raghunath M. received the Bachelors degree in Civil Engineering from the B.M.S. College of Engineering, Bangalore and Masters degree in Remote Sensing and Geographical Information Systems from the Indian Institute of Remote Sensing, Dehra Dun. He is working as Project Officer in TEQIP – World Bank funded Project at State Project Facilitation Unit (SPFU), Directorate of Technical Education, Bangalore and his research areas are Geographic Information System.

B. Shankar received the B.E. degree in Civil Engineering in 1984, M.U.R.P degree in Urban and Regional Planning in 1989 and Ph.D degree in 1997 from the University of Mysore, Mysore. He is working as Associate Professor in Urban and Regional Planning at the Institute of Development Studies, University of Mysore, Mysore. His research interests include heritage conservation, planning legislation, city planning, community participation and geographic information system.

Annexure I

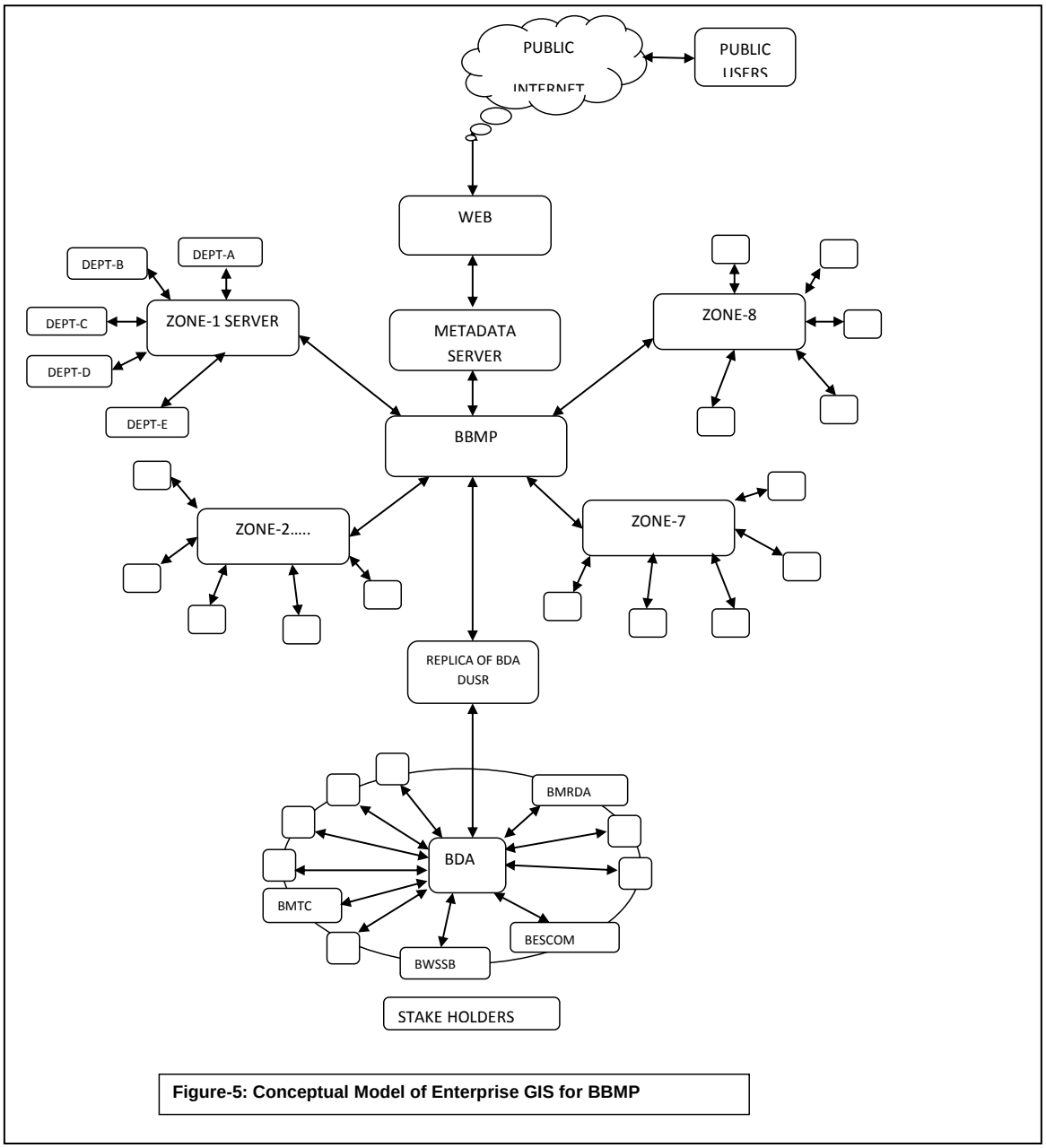


Figure-5: Conceptual Model of Enterprise GIS for BBMP

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-312-318 ISSN: 2249-6645

Power Transformer Winding Insulation
: A Review of Material characteristics

Aruna.M[1], V.V.Pattanshetti[2], Ravi.K.N[3], N.Vasudev[4],
[1](Aruna.M, EEE, ACIT, Bangalore-90)
[2](V.V.Pattanshetti, Joint Director, Dielectric materials division, CPRI, Bangalore-80)
[3](Dr.Ravi.K.N, Prof. and HOD EEE, SCE. Bangalore-73)
[4](Dr.N.Vasudev, Joint Director, HV Division, CPRI, Bangalore-80)

ABSTRACT
Transformers are one of the most important and cost-intensive components of electrical energy supply networks, thus it is of special interest to prolong their life duration while reducing their maintenance expenditures. The oil-paper insulation in power transformers is subjected to various stresses due to environmental conditions, voltage and fault stresses. Such stresses can cause deterioration of the oil-paper insulation in transformers. The condition of oil can be reversed back to some extent with the help of present technology such as on-line oil filtration by removing water and volatiles but not the acidic products of degradation. The degradation of paper insulation, however, is irreversible. Thus, the life of a transformer can be effectively determined by the life of its paper insulation. When paper degrades, it produces several by-products such as CO, CO2, and Furans and they migrate to the oil. There has been a growing trend throughout the world to study and estimate the deterioration of insulation strength of paper using such by-products as indicators. There are more direct approaches of degradation such as Tensile Strength (TS) and Degree of Polymerization (DP) measurements of paper. But these approaches require shutdown on the transformer and are considered intrusive.
It is felt that there is a need to review the insulation used in power transformer in terms of physical structure and degradation and various structural evaluation techniques available. This is being done to look for alternative materials which can substitute this natural material in terms of porosity, heat transfer, insulation and stability.

KEYWORDS
Transformer ageing, Oil Paper Insulation, moisture and ageing, paper insulation, aging, dielectric liquids, dielectric materials, insulation, insulation life.

INTRODUCTION
The most commonly used insulating materials in transformers are paper and mineral oil. Basically, apart from providing overall insulation to the transformer, the Mineral oil acts as coolant to the transformers, assisting in extinguishing arcs, and dissolves gases and moisture produced arising out of various phenomena within the transformer [1]. Whereas paper, it provides insulation to the conductor in the transformer windings. Presence of H_2O (water or moisture) in paper insulation has been linked to the decomposition of the paper fibers that is irreversible and eventually causes the paper to lose its mechanical and dielectric strength [2,3]. As for O_2 (oxygen), its presence causes oxidation on the mineral oil that leads to the deterioration on the oil insulation quality and the formation of acids. With acids present in the mineral oil, paper insulation is again exposed to deterioration and eventually ageing [3]. Ageing of paper insulation has been directly linked to its mechanical strength [2,4]. Studies have been done focusing on how long the paper can retain its mechanical strength as it ages before it loses its dielectric strength. Studies have also been done to estimate the life of transformers by studying the life of the paper insulation [4,5].

Cellulose Introduction: Cellulose is the most abundant biopolymer on Earth. About 33% of all plant matter is cellulose. Beta glucose is the monomer unit in cellulose. As a result of the bond angles in the beta acetal linkage, cellulose is a linear chain. Hydrogen bonding between the chains makes cellulose stiff and strong. There are different forms of cellulose. Porous cellulose fibers, Non-porous nano-crystalline cellulose particles, Regenerated cellulose films, Bacterial cellulose. Cellulose fibers are usually porous. This is refined cellulose with the amorphous region and impurities removed. The crystalline regions are several μm long and a few nm wide. Because of their high aspect ratio they can be regarded as nano-whiskers. These nano-crystals are further isolated in the form of independent particles. Cellulose nano-particles can be used as reinforcing phase

in thermoplastics and as novel paints because they can form liquid crystalline phases .

Regenerated cellulose film: Cellulose fibers are dissolved in alkali and carbon disulfide to make a solution called viscose, is then extruded through a slit into an acid bath to reconvert the viscose into regenerated cellulose film called cellophane. A similar process, using a hole (a spinneret) instead of a slit, is used to make a fiber called rayon.

Bacterial cellulose: It is also called microbial cellulose, a form of cellulose that is produced by bacteria. Bacteria from the genera Aerobacter, Acetobacter, Achromobacter, Agrobacterium, etc. synthesize cellulose. Only the Acetobacter xylinum produce enough cellulose to justify commercial interest.Acetobacter xylinum is reclassified as Gluconacetobacter xylinus (Yamada et al., 1997).

Bacterial cellulose has advantages over plant cellulose: Finer structure, Longer fiber length and much stronger, No hemicellulose or lignin need to be removed. But bacterial cellulose is about 100 times more expensive than plant cellulose. It is difficult to achieve large scale production capacity.

Cellulose Topology: Topology of polymers refers to the surface texture of polymers (J. Gooch, Encyclopedic Dictionary of Polymers, Springer,2007).The term surface texture designates the entirety of departures from the ideally smooth surface, inclusive of occasional flaws or other types of locally limited irregularities(M. Curtis, Dimensional Measurement, Indus. Press, 2007). Most surfaces are not smooth at atomic scale. A useful descriptor is the specific surface area (m2/g).Powders have a high specific surface area (10 to 500 m2/g). Surface texture means digression from the ideal smooth surface. It includes Topographical deviations generally associated with more of less regular waveforms, or waviness. On waveforms are the closely spaced random irregularities, called roughness, Waviness and roughness appear superimposed.
Maximum height of the profile (Rt) = Maximum peak height(Rp) -Maximum valley depth (Rv)

Techniques for examining cellulose surface topology: Surface profilometers are used to measures surface profiles, roughness, waviness and other finish parameters. Two basic surface profilometer technologies are used.Non-contact,Measure the surface texture by optically scanning a surface with a light or laser. Non-contact optical interferometer, is the technique of diagnosing the properties of two or more waves by studying the pattern of interference created by their superposition.

Linnik interferometer: A Linnik interferometer is a two-beam interferometer used in microscopy and surface contour measurements or topography.

Techniques for examining cellulose surface topology: Contact, Atomic Force Microscopy (AFM), also known as Scanning Force Microscopy (SFM), measures the height of surface features by touching the surface with an extremely sharp probe. Contact or stylus based surface profilometers use the technique to measure the surface topology.

AFM: The AFM consists of a micro-scale cantilever with a sharp tip (probe) at its end that is used to scan the specimen surface. The cantilever is typically silicon or silicon nitride with a tip with a radius of curvature on the order of nanometers. The probe is placed at the end of a cantilever with known mechanical properties. The instrument is also capable of imaging the surface while tapping the surface to minimize tip wear and sample damage. Another imaging technique involves hovering the tip above the surface. Hovering the probe tip very close to the sample causes atoms at the tip to interact with the atoms on the surface, and these interactions can deflect the cantilever. Measuring these interaction resulting tip deflection allows image analysis without ever touching the surface.

Advantages of AFM over the scanning electron microscope (SEM): AFM provides a true three-dimensional surface profile, Samples viewed by AFM do not require any special treatments (such as metal/carbon coatings) that would irreversibly change or damage the sample, SEM needs an expensive vacuum environment for proper operation, while most AFM modes can work perfectly well in ambient air or even a liquid environment, AFM can provide higher resolution. It has been shown to give true atomic resolution. Lateral resolution: 15-50 nm . Vertical resolution: 0.1 nm.

AFM Disadvantages: The SEM can image an area on the order of millimetres by millimetres with a depth of field on the order of millimeters, can only image a maximum height on the order of micrometres and a maximum scanning area of around 150 by 150 micrometres, An incorrect choice of tip for the required resolution can lead to image artifacts, Signals may require software enhancement and filtering. Such filtering could "flatten" out real topographical features.

Cellulose Porosity: Porosity refers to the ratio of the volume of voids contains within a sample of material to

the total volume, solid matter plus voids, expressed as a fraction, void fraction or percentage of voids according to Encyclopedic Dictionary of Polymers (J. Gooch, Springer,2007) Porosity(%)= (1-TrueVolume/BulkVolume)*100%

• Note that the pores collapse irreversibly when cellulose sample dries.

Techniques for examining cellulose porosity:

Mercury Porosimeter Principle:

In the porosimeter, mercury is forced into solid material. The pressure required to fill the pores completely is inversely proportional to the size of the pores

.D= -(1/P)4γ cosθ

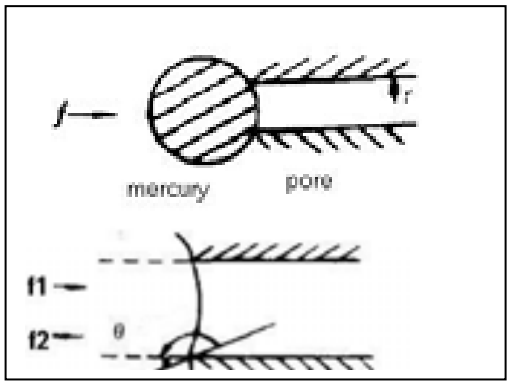
D- pore diameter P- applied pressure

γ -the surface tension θ -the contact angle

– The previous equation comes from capillary rise equation: P= -2γ cosθ/r

P-applied pressure r -radius of smallest capillary

γ- surface tension= 484 mN/m;θ-contact angle= 140deg



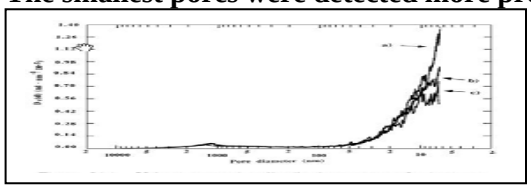
force in: f1=πr2P force out: f'= 2πr γ, f2=-2πrγ cos(180-θ) = -2πrγ cos θ

f1=f2

πr2P= - 2πr γ cos θ

P= -2γ cosθ/r

The smallest pores were detected more precisely:



Effect of scanning speed on porosimetry accuracy: The phenomenon is in agreement with Moscou and Lub's assumption* that there is no time for mercury to intrude into the sample and cover the inner surface when fast scanning speeds are used in the measurement.

Latest development in mercury intrusion porosimetry technique:

1. Poremaster series prosimeter from Quantachrome
2. Traditionally, N2 absorption method is used for pore size < 10 nm.
3. Now, Pore size ranges from about 900 μm to less than 3.5 nm in the 60,000 psi units can be detected.

Limitations: The material must not react with mercury, The porosimeter measures only those pores which open to the outside surface, The values of large openings within the sample, connected to the surface by narrow pores will be indicated as diameters of narrow pores, The limit of pore measurement is 0.003 to 360 micrometers. A maximum pressure of 60,000 is available.

Size exclusion chromatography:

SEC principle: A molecule is more or less able to enter the pores of a porous material depending on the molecule's size. When a column is packed with a porous material and an eluent is forced through, injected molecules that are too large to enter the pores will be eluted in the 'void volume' of the column. Tiny tracer molecules penetrate into almost all the pores of the material. Ve= Vo+ KVp

Ve- elution volume of the tracer, Vo- void volume of the column,K- distribution coefficient of the tracer, decided by the size and shape of the tracer and the size distribution and shape of the pores, Vp- total volume of the pores in the column

SEC conditions:

In order to characterize a porous structure reliably, these conditions should satisfied no adsorption interaction btw tracer molecules and porous material, distribution function.

$$f_p(R_p) = (1/V_p)\{dV/dR_p\}$$

where V is the volume of pores with a radius between R_p and $R_p + d R_p$

and V_p is the total pore volume of the membrane

Nuclear magnetic resonance (NMR):

NMR porosity:

The amplitude of a proton NMR measurement is directly proportional to the amount of hydrogen in the material investigated. The transverse relaxation time T2 is short in solids, of the order of 10 μs. The signal from those protons can be eliminated from the measurement by ignoring very fast component of the signal. The relaxation times of protons in pore fluids are greater than 1 ms. These protons are visible in the signal. The

relaxation times of water trapped in very small pores have intermediate values. By collecting these signal, the porosity and pore distribution can be determined:

- The H spin-spin(T2) relaxation profiles can be translated to pore size by the relationship btw T2 and surface-to volume (S/V) ratio of the pore.
  1/T2= ρ(S/V)
  ρ- surface reflexivity

NMR advantages:. Mercury Intrusion and Size Exclusion Chromatography technique may damage the delicate pore structure, means of non-destructive testing, without affecting the material structure. Mercury Intrusion and Size Exclusion Chromatography technique can test only the pores that are open to the outside surface. It can penetrate into the inner structure of the material, including the pores that are not exposed to the outside surface.

Cellulose: Electrical Properties:

Structure of Cellulose is β(1→4) linked D- glucose units. It is Crystalline and straight-chained .Cellulose is made of interwoven fibres. Mainly obtained from wood pulp and cotton. Cellulose being a natural material will have polar contaminants like lignin and other phenolics within the cellulose matrix. Further it has tendency to absorb moisture and is naturally degradable because of the weak glycosidic linkages. In order to minimize the contaminants the paper has been specially selected and impregnated to get the electrical grade papers for use in power transformer insulation applications.

Conductivity: It does not conduct as dry entity, used as an insulator, used as scaffolding for conductible materials.

Potentiometric Titration: It can be used to find the number of acid groups in a solution, used to find the dissociation constants of acidic groups. NaOH is added to acid solution, with a pH graph being created from the potential measurements. Concentrations are calculated as follows:

- **$[H^+]$ = (Vekv –Vt)COH / (Vo + Vt)**
- **$[OH^-]$ = - $[H^+]$**

Influence of Porosity in the Surface of cellulose has believed to have pores, allows for ions to flow in and out of cellulose substructure and also allows for more surface area that can be accessed for reactions. Overall effect of pores is minimal compared to surface potential. Because of Influence of Water/Drying, Conductivity increases with water content. Drying of cellulose fibers can lead to closure of pores.

Charge Density: It is nothing but Amount of electric charge on the surface area of cellulose, is very important regarding stability of colloid. It gives insight into interactions with other colloids in solution. Charge Density of Cellulose which decreases with acid groups on the cellulose surface. Raw cotton has a surface charge of 18.5mmol/kg. Regenerate Fibers have 4.7 mmol/kg,

Electrophoretic Mobility: It is the Proportionality between particle speed and electric field strength. Ions in solution can be moved with the application of an electric field. It can be calculated with various methods, namely electrophoresis. Mobility = Particle Velocity / Electric Field Strength.

Zeta Potential: Zeta potential is the potential difference between the dispersed medium and the stationary layer of fluid attached to the particle. There is a Correlation between zeta potential and mobility. It is used to describe the stability of a colloid.1.from 0 to ±5,Rapid coagulation or flocculation,2.from ±10 to ±30 Incipient instability,3. from ±30 to ±40Moderate stability,4.from ±40 to ±60Good stability,5.more than ±61Excellent stability

Zeta Potential Measurement Methods: Electrophoresis is Used for particulates (colloids). Laser Doppler Velocimetry is Used for particulates (colloids). It Uses laser refraction to measure mobility. Streaming Current is Used for flat surfaces and porous objects (films, membranes, fibers).Steaming Current Measurements measures the movement of charge between nodes. It is the measurement of mobility.

$$\zeta = \frac{\eta\{\lambda_0 + 2\lambda s\}\Delta E}{\varepsilon_r \varepsilon_0 \quad \Delta P}$$

Steaming potential high impedance

Smoluchowski Equation is the relation between mobility and zeta potential.

λ 0 -specific conductivity of solution

λ s=-specific conductivity of solution

r-capillary radius

l= capillary length

$$\mu = \varepsilon \zeta / \eta$$

remember $\varepsilon = \varepsilon_r \varepsilon_0$

Henry Equation with Ohshima is To account for double layer as well as inner electrokinetics. It is Used for materials that are not solid.

$$\mu_e = \frac{2 \varepsilon_v \varepsilon_0 \zeta f(ka)}{3 \eta}$$

$$f(ka) = 1 + \frac{1}{2\{1+\frac{2.5(1+\exp(-ka))}{Ka}\}^3}$$

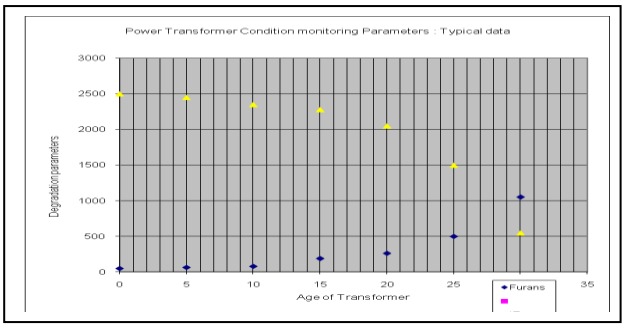


Figure.2 Graph of Degradation v/s Age of transformer

Zeta Potential of Cellulose is of Low negative values and -10 to -20 mV in water is as shown in figure 1.

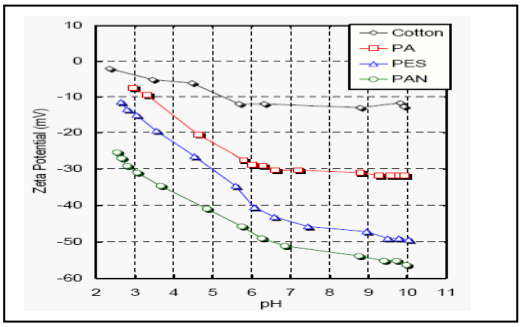


Figure 1. Zeta Potential of Cellulose

Power Transformer Condition Monitoring Parameters: Typical data

Transformer Age (Yrs)	Total Furan Content	Moisture content	Insulation Resistance Ohms	Moisture in solid insulation	Moisture in winding (Estimated)	Ambient temperature summer average	Average Load %
0	50	8	2500	--	0.05	39	60
2	65	12	1800	--	0.09	39	74
4	80	17	1745	--	1.02	40	76
6	190	19	1609	--	1.17	40	75
8	263	21	1589	--	1.61	39	75
10	500	26	1400	--	1.9	39	75
12	1053	26	1350	2.2	2.2	39	75

Figure 3. typical data of Power Transformer Condition monitoring Parameters

Transformer Age in no. of years	Furan Content	Insulation Resistance Ohms		
Age	Furans	IR	Moisture	Age
0	50	2500	0.05	0
5	65	2450	0.09	5
10	80	2350	1.02	10
15	190	2280	1.17	15
20	263	2050	1.61	20
25	500	1500	1.9	25
30	1053	550	2.2	30

Figure 4. typical data of Power Transformer Condition monitoring Parameters Age in no. of years, Furan Content, Insulation Resistance, Moisture, Age

Role of Double Layer is Zeta potential decreases with salt concentration and increases with double layer.

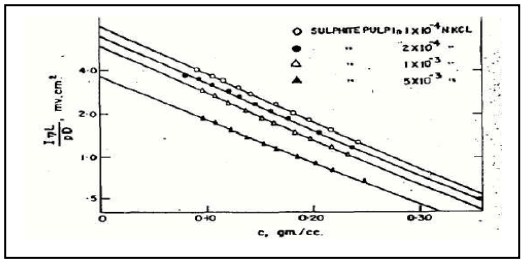
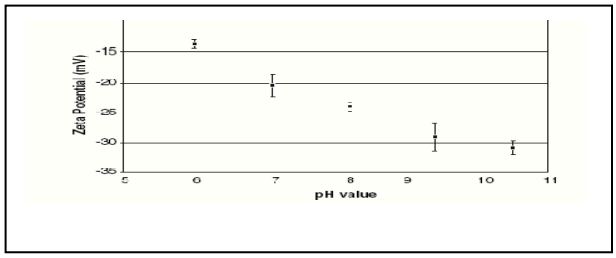


Figure 5. Zeta potential

Creating Stable Cellulose, Modification of cellulose is needed to create a larger zeta potential. Such mixtures include: Polyvinyl amine + CMC, Polyacrylonitrile + cellulose acetate, Acid treated cellulose. Cellulose Research Requiring Zeta Potential: Drug Delivery, Paper Processing, Textile Finishing, Nanocomposites



Three most common degradation factors of cellulose have been identified as thermal, oxidative, hydrolytic[3].Thermal Degradation[3],Oxidative Degradation[9],Hydrolytic Degradation[3], Degradation By-Products[3] ,Using the by-products as indicators to paper insulation condition [3,7],CO and CO2 [1],Furans [5] ,Non-Linear Relationship between Age and Degradation By-Products Concentration.

Weightages of the Degradation by-products : Typical degradation pattern on a 20 MVA trans former :It has been widely accepted that degradation is depending up the following aspects.

1. Loading of transformer : higher the load higher will be the electrical stresses and accompanied thermal stresses
2. Operating temperatures: It has been widely accepted that every 10^{o}C rise in temperature will almost double the rate of degradation. Hence, effective use of cooling accessories and maintaining the operating temperature will always of help in controlling the degradation rates.
3. Water levels in the oil and in the paper insulation has been found to be a critical factor in deciding the degradation rates.

Present demands on insulation: Our country is planning to go for 400 kV and 800 HV AC and DC transmissions. The power equipments needed will demand for good insulation will be far higher. This coupled with high ambient in the country it is felt that there is a need for looking alternative materials which can substitute cellulosic insulation.

Porous polymer preparation / Monoliths :

Porous polymers have been used in the area of chromatography wherein nano porous polymers can be prepared and used for separation of similar chemicals.
Monolithic separation media, made in one piece, contain only flow-through pores, which significantly augment the mass transfer based on convection. This enables use of high mobile phase velocities along with low back pressures and therefore fast separations without decrease of resolution. Glycidyl methacrylate stearylmethacryate-ethylenedimethacrylate and styrene-divinylbenzene monoliths have been parepared and have been characterized. Such materials can be investigated in place of cellulose insulation.

CONCLUSION

Degradation of cellulosic insulation depends up on loading, operating temperatures, water level in oil and in insulation.
Skilful maintenance of insulation under dry conditions will help preserving the insulation in dry conditions.
It has been widely accepted factors like operating temperature, moisture in winding and loading patters will have important weightages in degradation of insulation.

REFERENCES

[1] June 2003, Transformer Maintenance in Facilities Instructions, Standards, and Techniques, Hydroelectric Research and Technical Services Group, United States Department of the Interior (Bureau of Reclamation), Vol. 3-30

[2] Sparling, B.D., April 2000, "Assessing the Life of the Transformer - When Is It Time to 'Pull the Plug?'". GE Syprotec Inc

[3] Unsworth, J., Mitchell, F., August 1990,"Degradation of Electrical Insulating Paper Monitored with High Performance Liquid Chromatography". IEEE Trans. On Electrical Insulation,Vol.25,No.4, 737-746

[4] Emsley, A.M., Heywood, R.J., Ali, M., Xiao, X., November 2000, "Degradation of cellulosic in power transformers. Part 4: Effects of Ageing on the tensile strength of paper". IEE Proc. Sci. Meas. Technol., Vol 147, No. 6, 285-290

[5] Thomas, P., Shukla, A.K., Raghuveer A.K., June 25-29, 2001, "Ageing Studies On Paper – Oil to Assess the Condition of Solid Insulation Used in Power Transformer", IEEE Int. Conf. on Solid Dielectrics, 69-72

[6] Blue, R., Uttamchandani, D., Farish, O., April 1998, "Infrared Detection of Transformer Insulation Degradation Due to Accelerated ThermalAgeing". IEEE Trans. On Dielectrics and Electrical Insulation, Vol. 5, No. 2, 165-168

[7] Sans, J.R., Bilgin, K.M., Kelly, J.J., June 1998, "Large-Scale Survey of Furanic Compound in Operating Transformers and Implications for Estimating Service Life", IEEE Trans. On Electrical Insulation, 553-543

[8] Allan, D., May 25-30, 1997, "Recent Advances in the Analysis and Interpretation of Aged Insulation from Operating Power Transformers". Proc. 5th Int. Conf. on Properties and Applications of Dielectric Materials, 202-205

[9] Merhar, M. Podgomik. A., Barut M. Strancar A. & Zigon M. : BIA Separations d.o.o., Teslova 30, SI-100 Ljubijana, Slovenia & Faculty of Chemistry and Chemical Technology, University of Ljubljana, Askerceva 5, SI-100 Ljubljana :
High Performance Liquid RP Chromatography using novel monolithic supports: An internet publication.

A PROPOSED RDIT ALGORITHM FOR ANALYSIS OF FACE VERIFICATION IN VIDEOS IN IMAGE DATABASE FOR CRIME INVESTIGATION

V.S. Manjula,
Asst. Professor,
Dept. of Computer Application,
St. Joseph' College ,
Chennai- 600 122, INDIA.

Lt. Dr. S. Santhosh Baboo,
Reader, P.G & Research,
Dept. of Computer Application,
D.G. Vaishnav College,
Chennai-106. INDIA.

Abstract
In the face recognition most researches has considered about face detection and identification or a face detection and tracking, but best of our knowledge very few researches has focused on the face detection, identification and tracking. The combination of these three mechanism are been used in many real time applications .Here we consider one of the application as theft detection in real time which proposes a new mechanism named as FACE RDIT (Face Rectangular Detection, Identification and tracking) algorithms. Also here we include some preprocessing algorithms like gray scale conversion and histogram equalization for image enhancement process. Apart from this we also derive a new algorithm which is based on rectangular features of the face and multi linear training. Also the proposed technique has four main steps like training, detection, identification and tracking by using this proposed method we train the unauthorized users photo and then make the detection of the faces in the group of frames and then identify the particular personality from the group of images. Finally it tracks the identified thief in next following video frames by our proposed algorithm. Thus the effectiveness of the proposed methods is well supported by both detailed analysis and extensive experimentation on a face verification problem.

***Keywords*:** RDIT, face Rectangular Detection, histogram equalization, rectangular features.

1. Introduction

The Face detection is one of major research area in computer vision. The task of face detection is so trivial for the human being, yet it still remains a challenging and difficult problem to enable a computer to do face detection. Many difficult problem are caused by a diversity of variations, such as human races, illumination, facial expression, contrast between face and background, face regions overlapped one another and orientation of the face. The efficient detection of human faces in images, however, is fundamental in a variety of applications requiring intelligent human computer interaction[1].

We proposed used mechanism to several real time applications. we can consider the thief detection in the video frame .base on that we can propose the new mechanism named as FACE RDIT (Face Rectangular Detection, Identification and tracking).here we utilize some preprocessing algorithms like gray scale conversion and histogram equalization for image enhancement. Than we derive the new algorithm based on rectangular features of the face and multi linear training .in the proposed technique have four main steps they are training, detection, identification and tracking .that is we can train the thief photo image, than detect the faces in the video frame and than identify the thief face [2, 3]. Finally track that identified thief in next following video frames by our proposed algorithm. Face detection and tracking find applications in areas like video structuring, indexing, and visual surveillance and form active areas of research. If the application is to identify an actor in a video clip or to find a particular shot in the video sequence in which the actor is playing, then faces are the most important "basic units"[4] . This requires detection and tracking of a face through a sequence [5, 6].

The two approaches to handle these issues could be frame based detection, that is, to detect faces in each frame without taking into account the temporal information or integrated detection and tracking in which the face is detected in the first frame and tracked through the sequence. The frame-based approach completely overlooks the fact that the frames are contiguous in the sequence. In the second approach, tracking and detection are independent and information from only one source is used at a time, causing a loss of information[7]. This motivated us to develop a novel approach that integrates detection and tracking into a unified framework—the temporal approach. It uses the temporal relationships between the frames to detect

multiple human faces in a video sequence, instead of detecting them in each frame independently. Alternatively, face tracking and detection can be combined by detecting facial features like lips, mouth, nostrils, and eyes and by tracking them through the sequence, but this imposes the constraint that these features would need to be visible and, therefore, only frontal views of the face can be handled [8, 9].

2. Our Work in Perspective

The features of our approach which distinguish it from existing approaches are simultaneously detection and tracking information at each time step and is therefore able to handle changes in imaging conditions (face scale, lighting, and orientation) and changes in image content (the complexity of the background and the number of faces), as has been shown in the experiments. It is able to improve over detection results of the existing detectors. Handling pose change is a challenging problem for any detector. Most of the tracking approaches suffer from the problem of manual initialization. We avoid this by using detection for initializing the tracker. The detection information is integrated at each time step as the parameters are being propagated, that is, the probabilities are accumulated over time. This causes the algorithm to continuously detect faces even in frames where the frame-based detector fails. The detection information provides knowledge of the appearance of new faces to the trainning, which can be readily incorporated whenever they appear by a process of updating.

2.1. Training

In the training process the user will start entering the username along with a frame number as a reference for the training data sets. Secondly we check the trained data after the processing starts.Apart from this here we introduce a rectangle based approach which is used for face detection. In this proposed approach the face is been spited into different rectangle portions of features and are stored in the database for further verifications. In the Initial stage the system is been processed with basic preprocessing steps like Grey Scale Conversion and Histogram Equalization for the Color Images. Figure 1.describes about the Proposed Architeture for the face detection and tracking approach.

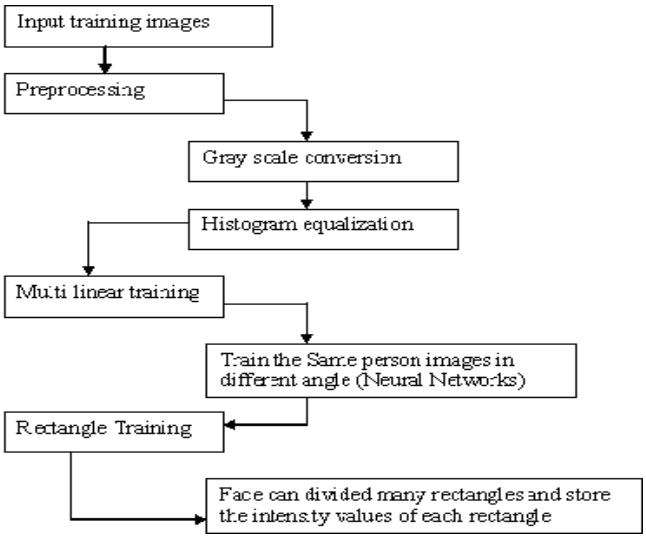


Figure 1.Proposed Architeture

The tainning approach is done by neural network which is of a back-propagation is adopted as the second part of our face detection system. In our neural system, the input vector of the network is been projected with weights of the face space and we choose hundreds of face images, which include different faces with different posses. The input images for processing are taken from different databases like ORL dataset, the MIT CMU face database, and the Wide World Web to produce different face blocks. The training face images consist of all combination of input datasets from different datasets like ORL face images, MIT_CMU images, and randomly selected WWW images. In addition, fifty eigenvectors with the largest associated eigenvalues are adopted as the face space. The projection weights of the face and non-face blocks are computed and used as the positive and negative training vectors of the neural network, respectively. In our system, the neural network consists of a hidden layer with nine hidden nodes.

Backpropagation Neural Network

- Set all weight to random value range from -1.0 to 1.0.
- Set an input pattern (binary values) to the neurons of the net's input layer.
- Active each neuron of the following layer:
  - Multiply the weight values of the connections leading to this neuron with the output values of the preceding neurons.
  - Add up these values.
  - Pass the result to an activation function, which computes the output value of this neuron.
- Repeat this until the output layer is reached.
- Compare the calculated output pattern to the desired target pattern and compute a square error value.

- Change all weights values of each weight using the formula:
  Weight (old) + Learning Rate * Output Error * Output (Neuron i) * Output (Neuron i + 1) * (1 – Output (Neuron i + 1))
- Go to the first step.
- The algorithm end, if all output pattern match their target pattern.

2.2 Image Preprocessing

Face image is what is at the input of the system. The expected result of preprocessing stage is an image which contains only the significant features of the face. To avoid variations which present in the input image we will attempt to build an algorithm of averaging the details so that only main features remain. Thus the face regions of different images which were taken in dissimilar conditions, e.g. light, background, and face to camera distance, may cause varied appearances of faces. In order to have the property of the scale invariance, the input image is resized into the image pyramid. We partition each sub-sampled image into 21 x 21 pixel blocks at every position of the image, and normalize the image blocks to have zero mean and unit variance.

2.2.1 Gray scale conversion

Grayscale images are distinct from one-bit bi-tonal black-and-white images, which in the context of computer imaging are images with only the two colors, black, and white (also called *bilevel* or *binary images*). Grayscale images have many shades of gray in between. Grayscale images are also called monochromatic, denoting the absence of any chromatic variation (i.e., one color). Grayscale images are often the result of measuring the intensity of light at each pixel in a single band of the electromagnetic spectrum (e.g. infrared, visible light, ultraviolet, etc.), and in such cases they are monochromatic proper when only a given frequency is captured. The original image contains about 1crore and seventy lakhs colors (since the input image has 2 power 24 bit colors). This original 24 bit image will be converted into gray scale image. The gray scale image contains 2 power 16bit (65536) colors. This color conversion is performed by using the pixel values of the original image.

We take an RGB color image as input

$$(R_i, G_i, B_i) \in [0.1]^3$$

And produce a grayscale image as output

$$T \in [0.1]$$

To avoid gamma correction issues.

2.2.2. Histogram equalization

However it can also be used on color images by applying the same method separately to the Red, Green and Blue components of the RGB color values of the image. Still, it should be noted that applying the same method on the Red, Green, and Blue components of an RGB image may yield dramatic changes in the image's color balance since the relative distributions of the color channels change as a result of applying the algorithm. However, if the image is first converted to another color space, Lab color space, or HSL/HSV color space in particular, then the algorithm can be applied to the luminance or value channel without resulting in changes to the hue and saturation of the image.

Consider a discrete grayscale image, and let n_i be the number of occurrences of gray level i. The probability of an occurrence of a pixel of level i in the image is

$$P(x_i) = n_i / n$$

Where, $i \in 0 \ldots L - 1$

L → the total number of gray levels in the image

n → the total number of pixels in the image
P → in fact the image's histogram and Normalized to [0, 1]
x → an occurrence of a pixel in the image

$$C(i) = \Sigma^{i}_{j=0} P(x_j)$$

Where ***C*** *→ the cumulative distribution function corresponding to p,*
We would like to create a transformation of the form that will produce a level Y for each level x in the original image, such that the cumulative probability function of Y will be linearized across the value range. The transformation is defined by:

$$Y_i = T(x_i) = c(i)$$

Y_i → *linearized transformation of pixels*
Notice that the T maps the levels into the domain of 0...1. In order to map the values back into their original domain, the following simple transformation needs to be applied on the result.

$$y'_i = y_i \cdot (Max - min) + Min$$

y'_i →*equalized image*
Max →Maximam value pixel in the image
Min → Minmam value pixel in the image

2.2.3. Multi linear training

we present a neural network-based algorithm to detect upright, frontal views of faces in gray- 1scale images. The algorithm works by applying one or get a representative sample of images which contain faces. Each network is trained to set as training progresses [21]. The algorithms makes the output efficient when number of input data training gets increased and the size of the training set are higher . Thus the Multi linear training results to produce a proper detection that are obtained with significant improvement with good accuracy of the detector.

$$\text{Train1} \rightarrow I_i (N_i)$$

Where, I_i → *Intensity,*
N_i →*Different angle image for single person ,*
Train1 → Multi linear training

2.2.4. Rectangle facial training

The first step in our work is to look for face candidates. We use the feature-based method, and the point of the view is from the characteristics human face, one is intensity; the other is symmetry. One feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The other feature compares the intensities in the eye regions to the intensity across the bridge of the nose. Here we propose a three rectangle features which are used to look for face candidates, which are based on the idea mentioned above. In our method, a rectangular window is scanned on the input image to look for face candidates by using three rectangle features. Fig. 2 shows the rectangle features. The parameter ie. the intensity values are been gathered from these rectangles of the image features and are stored in the database for trainning [7].

$$Rect_{si} = Rect_f (w*h) / n_i$$

$$\text{Train2} \rightarrow I_i (Rect_{si})$$

Where, $Rect_{si}$ →*Small rectangle ,* $Rect_f$ →*Full face Rectangle ,* w → *Rectangle width ,* h → *Rectangle height ,* n_i → *Number of Rectangle , Train2 → Rectangle facial training*

3. Proposed Detection Identification And Tracking Algorithm (PDIT)

In face detection the Skin color provides the major role in extracting good information from the face area. The use of color information can simplify the task of face localization in complex environments [7].Several studies show that the major difference is not intensity but color itself. Many researchers have proposed various skin detection techniques based on different color space models such as HSV, YCbCr and YIQ. Human skin color, though it differs widely from person to person, is distributed over a very small area on the CbCr plane [8, 9]. This model is robust against different types of skin, such as those of people from Europe, Asia and Africa. So we use the YCbCr color model for the detection of skin color in this paper.

3.1 Face Detection

Given as input an arbitrary image, which could be a digitized video signal or a scanned photograph, determine whether or not there are any human faces in the image, and if there are, return an encoding of the location and spatial extent of each human face in the image. [3]

$GRect_{si}$=$GRect_f$ (I_i>equalized value)

$SRect_{si}$ =$GRect_{si}$ (w*h) / n_i

If I_i ($SRect_{si}$) = Train2

This rectangle is in face

Else

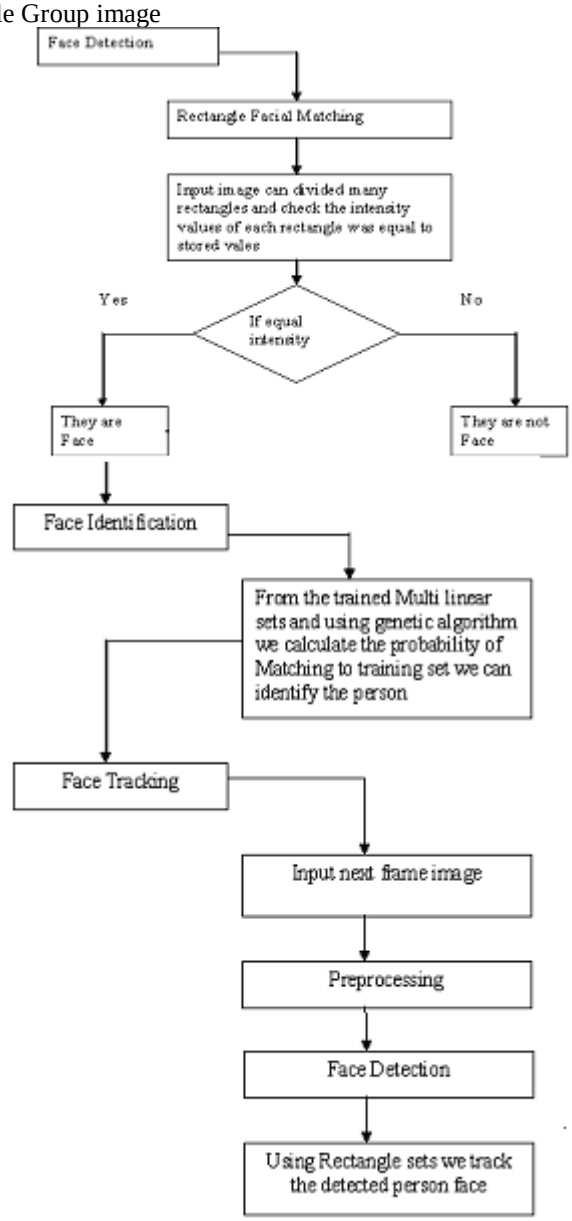
This rectangle not in face

Where

$GRect_{si}$→Face rectangle Group image

$GRect_f$→Full input Group mage Rectangle

$SRect_{si}$ →Small rectangle Group image



3.2 Face Identification

All face images are been splitted into triangle blocks and are given to the neural network. However, this might be still exist some face blocks in the face candidates from the network outputs. We design a simple face verification method to remove these false detection blocks. The main idea of the verification scheme is to consider the distribution of the edge points from the general facial features. Each candidate block is transformed into edge map by the Sobel edge detector method in order to detect the face edges.

The face identification can be used for number of frames or group of frames collecting and compare to training processs.In the training process the face features are been extracted and stored in database, for the face identication and matching. Each image in the database is been matched with the features that are been extracted from the input image and is done by the STRect algorithm. This algorithm stores face rectangle features from different images.

If I_i ($GRect_{si}$) = Train1
Face Identified who is that person
STRect= $GRect_{si}$
Else
No one Identified
Where
STRect→Store the Identified face rectangle

3.3 Face Tracking & Recognisation

In dynamic scenes, tracking is used to follow a face through the sequence. In order to incorporate the face changes over time, in terms of changes in scale, position and to localize the search for the face, it is essential to exploit the temporal correspondence between frames. Tracking exploits the temporal content of image sequences. Face tracking can be divided into two categories 1) head tracking and 2) facial feature tracking. Feature tracking methods track contours and points [12] or follow eyes and mouth [11], and require independent trackers for each feature. Head tracking methods use the information from the entire head and can be region-based [11], color-based [24], or shape-based [5]. Color-based approaches are not robust to lighting changes and approaches that use information from the entire head are, in general, unable to handle s. Tracking involves prediction and update for which filters like Kalman filter and Condensation filter have been used. Tracking approaches can also be model-based, for example, using statistical models or exemplar-based (although only specific features of the face, e.g., lips have been tracked). A combination of feature and head tracking methods, together with filtering, has tried to eliminate the problems of the individual approaches. The tracker of Burchfield simultaneously exploits the elliptical contour fitted to the face and the color information enclosed. This approach can handle out-of plane rotations and occlusions but is unable to handle multiple faces and requires manual initialization. The framework of can be used to track multiple faces, but it does not permit the addition of new faces. Raja et al. combined motion detection with an appearance-based face model. Multiple people tracking were performed using multiple Kalman filters.

If I_i ($NFRect_{si}$) = STRect
Identified person was present in the frame
Else
Identified person was not present in the frame
Where
NF → Next Frame of Group image (applied preprocessing and Face Detection)
$NFRect_{si}$ → Next Frame Face rectangles

4. Experiments and Results

Different real color images containing multiple faces with various sizes and different lighting conditions are employed to test the efficiency of the proposed approach. The appearance of the skin color can be changed due to different lighting conditions. In order to reduce the effect of illumination, we adopt gray

world method [10] to perform light compensation. The corrected red, green and blue color components were then nonlinearly transformed in the YCbCr color space. The skin tone pixels are detected using Cb and Cr components in the YCbCr color space. Let the thresholds be chosen as [Cb1, Cb2] and [Cr1, Cr2], a pixel is classified to have skin tone if the values [Cb, Cr] fall within the thresholds. Each pixel in Cb and Cr components which does not meet a certain threshold range [Cb, Cr] is set to zero. We can see the skin segmented result on a Real color image by YCbCr in Fig. 4.1,4.2,4.3 a

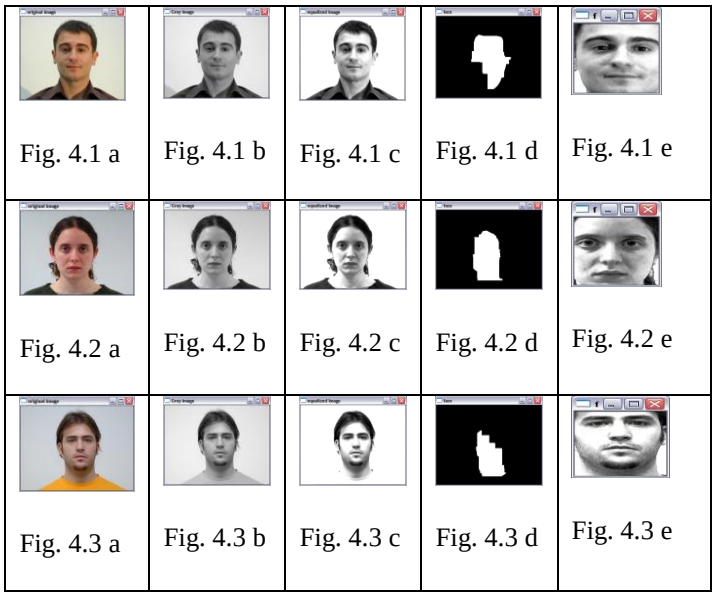


Fig. 4.1 a	Fig. 4.1 b	Fig. 4.1 c	Fig. 4.1 d	Fig. 4.1 e
Fig. 4.2 a	Fig. 4.2 b	Fig. 4.2 c	Fig. 4.2 d	Fig. 4.2 e
Fig. 4.3 a	Fig. 4.3 b	Fig. 4.3 c	Fig. 4.3 d	Fig. 4.3 e

Fig. 4.1,4.2,4.3 Comparison of (a)original (b)Grey Scale (c) Histogram equalization Image (d) Skin Color Extraction (e) Face Extraction

As is evident from these results, our algorithm demonstrates exceptional performance. As the quantitatively assessing the performance in practical application is a complicated issue because of the ideal images and are normally unknown at the receiver end. So here we use the following method for experiments. An original image is applied with Grey scale conversion is represented in Fig. 4.1,4.2,4.3 b and are transformed into the Histogram equalization and is represented in Fig. 4.1,4.2,4.3 c and by taking skin color extraction is done according to proposed algorithm which is described in 4.1,4.2,4.3 d. In this algorithm, In Fig. 4.3 a, we show the results of applying the number of trained input vs. accuracy and in Fig. 4.3 b it describes about the number of trained input vs. speed of detection rate of the face images.

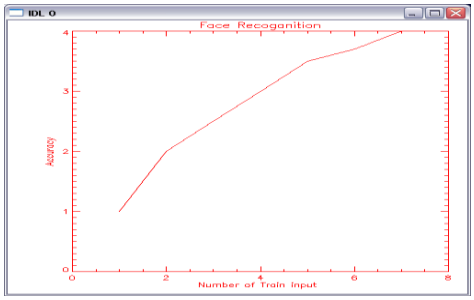


Fig. 4.3 a number of trained input vs. accuracy of detection

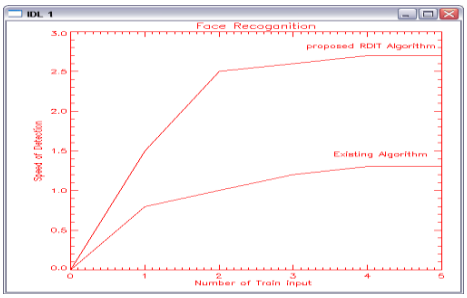


Fig. 4.3 b number of trained input vs. speed

5. Conclusion

In this paper, a new proposed RDIT algorithm is applied to Face detection & tracking and to overcome the existing issue. we can consider the thief detection in the video frame .base on that we can propose the new mechanism named as FACE RDIT (Face Rectangular Detection, Identification and tracking).here we utilize some preprocessing algorithms like gray scale conversion and histogram equalization for image enhancement. Than we derive the new algorithm based on rectangular features of the face and multi linear training .in the proposed technique have four main steps they are training, detection, identification and tracking .that is we can train the thief photo image, than detect the faces in the video frame and than identify the thief face. Finally track that identified thief in next following video frames by our proposed algorithm.

References

1. Yepeng Guan, Lin Yang "unsupervised face detection based on skin color and geometric information"

2. R.-L. Hsu, M. Abdel-Mottaleb, A.K. Jain, "Face detection in color images", IEEE Trans. PAMI, 2002, 24(5), pp. 696–706

3. H. Wang, S.F. Chang, "A highly efficient system for automatic face region detection in MPEG videos", IEEE Trans. Circuit Systems for Video Technology, 1997, 7(4), pp. 615–628

4. C. Garcia, G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis", IEEE Trans. Multimedia, 1999, 1(3), pp. 264–277

5. A. Rizzi, C. Gatta, and D. Marini, "Color correction between gray world and white patch", IS&T/SPIE Electronic Imaging 2002. The human Vision and Electronic Imaging VII Conference, 4662, San Jose, 2002, pp. 367-375

6. A. Guetta, M. Pare, and S. Rajagopal, "Face Detection. EE368 Final Project", 2003

7. I. Pitas, A. Karasaridis, Multichannel transforms for signal/image processing, IEEE Trans. Image Processing, 1996, 5 (10), pp. 1402–1413.

8. C.-C. Chiang, W.-K Tai, et al, "A novel method for detecting lips, eyes and faces in real time", Real-Time Imaging, 2003, 9(4), pp. 277-287.

9. H. Kruppa, B. Schiele, "Using Local Context to Improve Face Detection", In Proc. of the British Machine Vision Conference (BMVC'03), Norwich,

10. P. Viola, M. J. Jones, "Robust real-time face detection", International Journal of Computer Vision, 2004, 57(2), pp. 137-154.

11 L. Mostafa and S. Abdelazeem, "Face Detection based on Skin Color using Neural Networks", in Proc. of the first ICGST International Conference on Graphics, Vision and Image Processing GVIP '05, Cairo, Egypt, pp. 53-58.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-327-331 ISSN: 2249-6645

Simulation and Analysis of SRAM Cell Structures at 90nm Technology

Sapna Singh[1], Neha Arora[2], Prof. B.P. Singh[3]

(Faculty of Engineering and Technology, Mody Institute of Technology and Science, India)

ABSTRACT

SRAM is a most common embedded memory for CMOS ICs and it uses Bistable Latching circuitry to store a bit. This paper represents the simulation of different SRAM cells and their comparative analysis on different parameters such as Power Supply Voltage, Operating Frequency, Temperature and area efficiency etc. All the simulations have been carried out on BSIM 3V3 90nm technology at Tanner EDA tool.

Keywords *– CMOS Logic, Low power, Speed, SRAM and VLSI.*

I. INTRODUCTION

A SRAM cell consist of a latch, therefore the cell data is kept as long as power is turned on and refresh operation is not required for the SRAM cell. SRAM is mainly used for the cache memory in microprocessors, mainframe computers, engineering workstations and memory in hand held devices due to high speed and low power consumption. Each bit in an SRAM is stored on four transistors that form two cross-coupled inverters. This paper compares the different SRAM cells configurations on the basis of the power dissipation, speed, operating frequency range and their temperature dependence with the area efficiency of the circuit.

II. LITERATURE REVIEW OF DIFFERENT SRAM CELLS

2.1 6T SRAM CELL

The schematic diagram of 6T SRAM cell [1] is shown in Fig.1. Access to the cell is enabled by the word line (WL) which controls the two access transistors, in turn, control whether the cell should be connected to the bit lines: BL and BLB. They are used to transfer data for both read and write operations. While it's not strictly necessary to have two bit lines, both the signal and its inverse are typically provided since it improves noise margins.

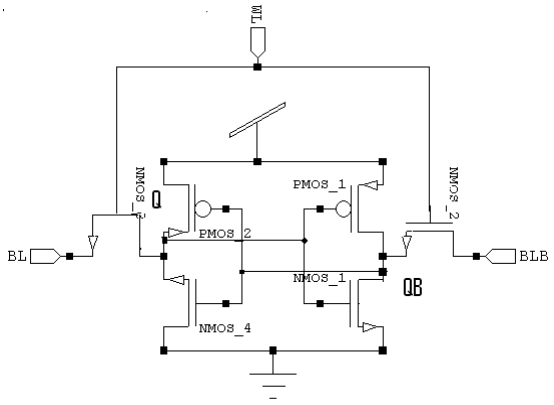


Fig. 1 Schematic of 6T SRAM Cell

2.2 MODIFIED 6T SRAM CELL

Fig. 2 is depicting the circuit diagram of modified 6T [2], [3], [4], [5], [6] SRAM Cell. The transistors NMOS_3, PMOS_4 and NMOS_2, PMOS_1 form cross coupled inverters. Reduction of leakage power is the effective stacking of transistors in the path from supply voltage to ground. This is based on the observation that "a state with more than one transistor OFF in a path from supply voltage to ground is far less leaky than a State with only one transistor OFF in any supply to ground path."

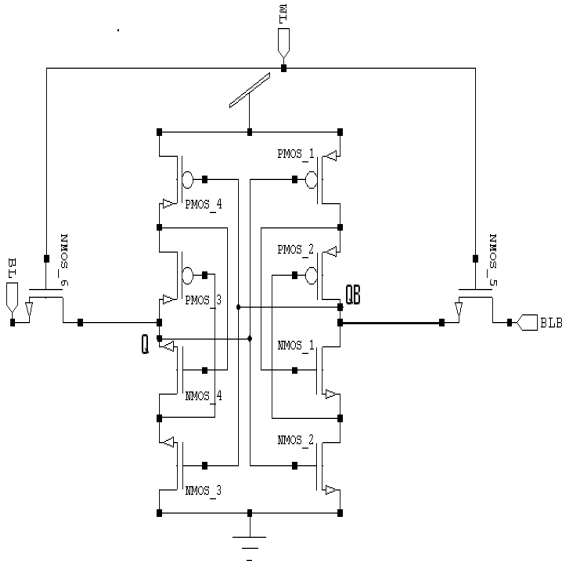


Fig. 2 Schematic of Modified 6T SRAM Cell

2.3 7T SRAM CELL

The 7TSRAM cell [6], [7], [8] uses a novel write mechanism shown in Fig.3. Write mechanism depends only on one of the 2 bit-lines to perform a write operation, which reduces the activity factor of discharging the bit-line pair. The limitation was that area overhead from the conventional 6T SRAM cell

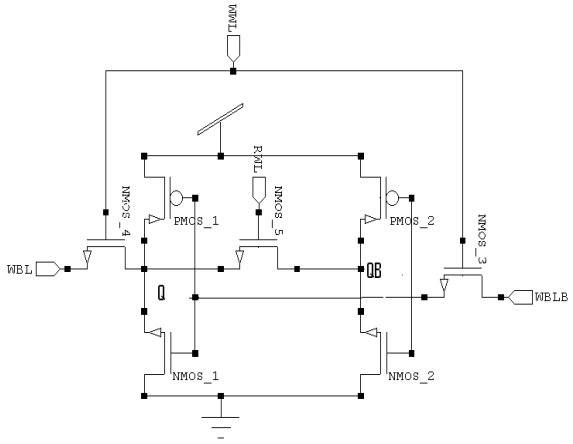


Fig. 3 Schematic of 7T SRAM Cell

2.4 8T SRAM CELL

A dual-port cell (8T-cell) [9], [10], [11] is created by adding two data output transistors to 6T-cell, as shown in Fig. 4. Separation of data retention element and data output element means that there will be no correlation between the read SNM Cell and I Cell. This 8T-cell has 30% more area than a conventional 6T-cell.The 30% area overhead is composed of not only the two added transistors but also of the contact area of the WWL, the word-line for write operations.

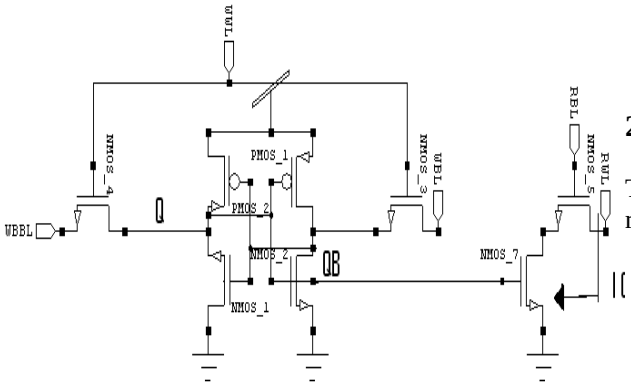


Fig. 4 Schematic of 8T SRAM Cell

2.5 9TSRAM CELL

Schematic of 9T SRAM cell [12] is shown in the Fig. 5. This circuit shows reduced leakage power and enhanced data stability. The 9T SRAM cell completely isolates the data from the bit lines during a read operation. The idle 9T SRAM cells are placed into a super cutoff sleep mode, thereby reducing the leakage power consumption as compared to the standard 6T SRAM cells.

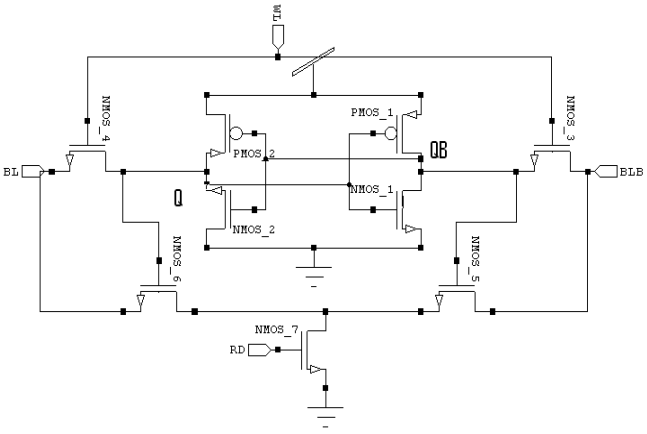


Fig. 5 Schematic of 9T SRAM Cell

2.6 10T SRAM CELL

The dual Port SRAM (10T) [13], [14] as shown in the Fig. 6 has only one read or write can occur per cycle, able to operate the SRAM in Subthreshold region also. The following circuit shows substantial power saving over a low range of power supply voltages.

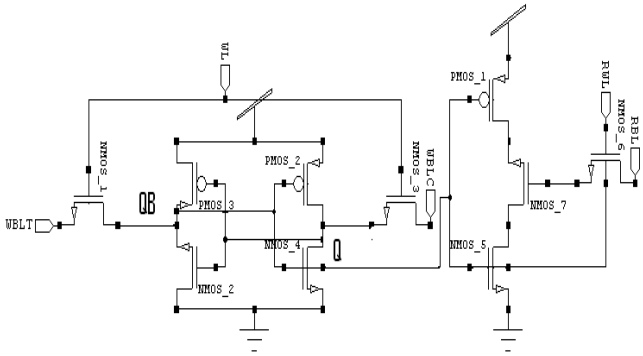


Fig. 6 Schematic of 10T SRAM Cell

2.7 MODIFIED 10T SRAM CELL

Modified 10T SRAM cell [15] is as shown in Fig.7 This circuit shows 10T SRAM Cell with differential read bitlines (BL and BLB).

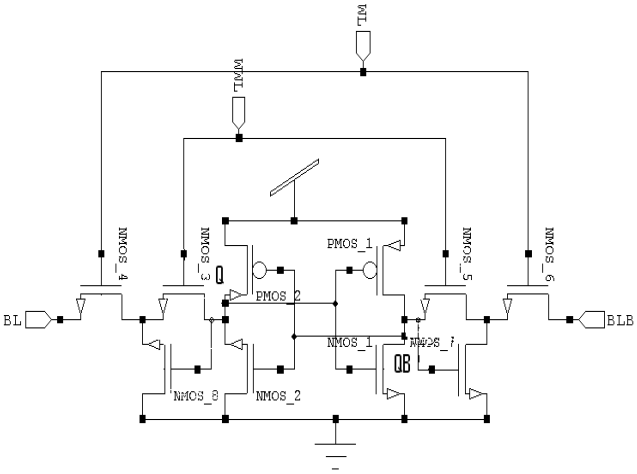


Fig. 7 Schematic of Modified 10T SRAM Cell

Two NMOS transistors (NMOS_4 and NMOS_8) for the RBL and the other additional NMOS transistors (NMOS_6 and NMOS_7) for BLB are appended to the 6T SRAM. As well as the 8T SRAM, precharge circuits must be implemented on the BL and BLB.

2.8 11T SRAM CELL

In Fig. 8 the schematic of the 11T-SRAM cell [16] is shown. Transistors PMOS_3, PMOS_1, NMOS_7, and NMOS_8 are identical to 6T SRAM, but two transistors NMOS_1 and NMOS_2 are downsized to the same size as the PMOS transistors. Minimum size transistors were used for the added 5Tcircuitry, except the access transistor that has a larger size. The most important part of the 11T-SRAM is a boost capacitor (CB) that connects source of NMOS_3 to RDWL.

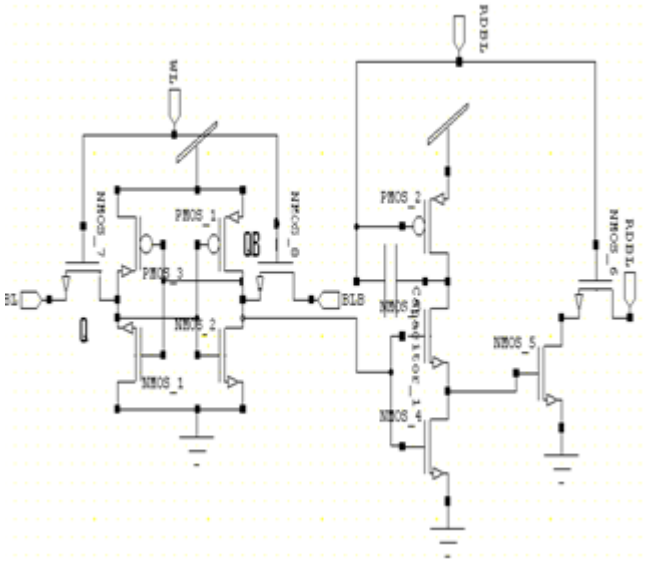


Fig. 8 Schematic of Modified 11T SRAM Cell

III. SIMULATION AND ANALYSIS

3.1 SIMULATION ENVIRONMENT

All the circuits have been simulated using BSIM 3V3 90 nm technology on Tanner EDA tool. To make the impartial testing environment all the circuits has been simulated on the same input patterns. All the simulations has been done on room temperature.

3.2 SIMULATION ANALYSIS

Fig. 9 is depicting the power consumption Vs Vdd for different SRAM cells. Modified 10T SRAM Cell shows the least power consumption over other approaches. Fig. 10 shows delay Vs Vdd for different SRAM cells. The 6T SRAM cell shows least delay among all the other design techniques. The reason for showing maximum speed is the least transistor count in the design approach. Fig. 11 and Fig. 12 shows Power Consumption Vs Operating Frequency and Temperature respectively. Both the above figures depicts shows 10 T Modified SRAM Cell shows always best performance for the range of operating frequency and Temperature among all the other design approaches for SRAM Cell.

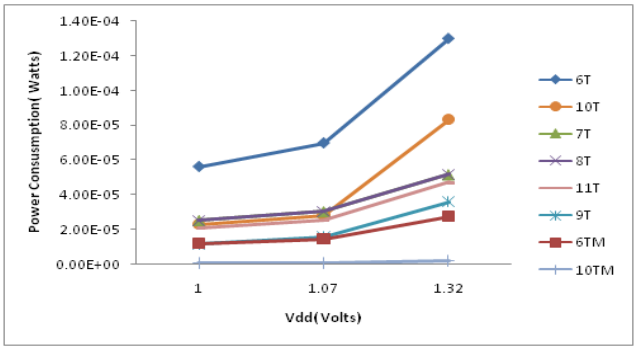


Fig. 9 Power Consumption Vs Vdd for Different SRAM Cells.

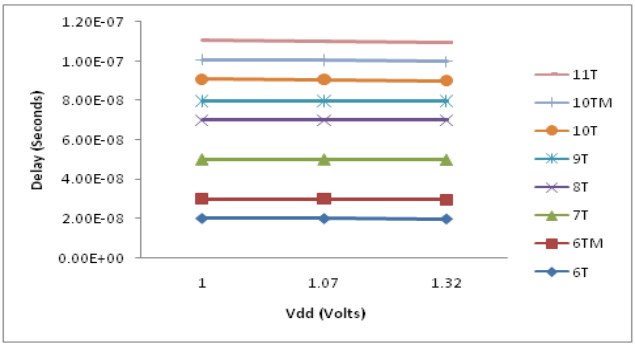


Fig. 10 Delay Vs Vdd for Different SRAM Cells.

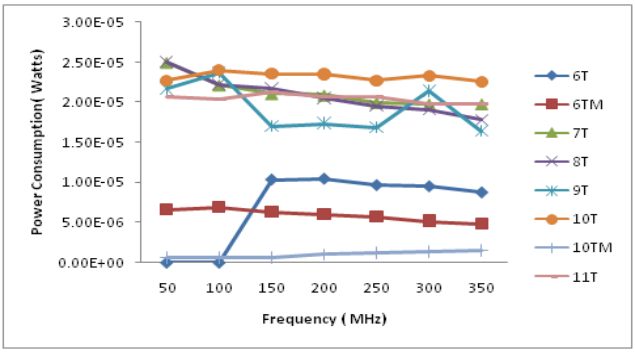


Fig. 11 Power Consumption Vs Operating Frequency for Different SRAM Cells

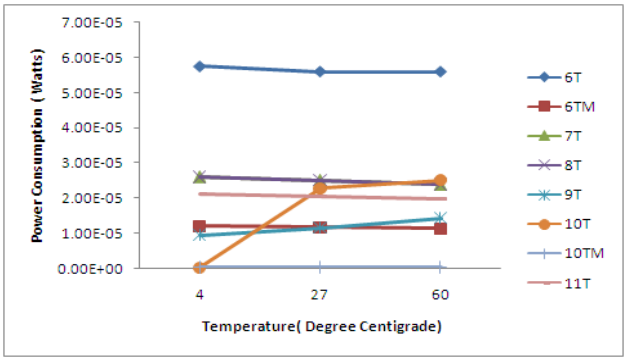


Fig. 12 Power Consumption Vs Operating Temperature for Different SRAM Cells

TABLE 1: Power Delay Product Comparison of Different SRAM Cells

Different SRAM Cells	Power Delay Product (Watt Seconds)		
	Vdd = 1v	Vdd= 1.07v	Vdd=1.32 v
6T	1.12E-12	1.39E-12	2.57E-12
6TM	1.18E-13	1.45E-13	2.69E-13
7T	5.03E-13	6.06E-13	1.05E-12
8T	5.03E-13	6.06E-13	1.05E-12
9T	1.16E-13	1.56E-13	3.49E-13
10T	2.46E-13	3.00E-13	8.60E-13
10TM	**3.04E-15**	**4.85E-15**	**1.78E-14**
11T	2.05E-13	2.52E-13	4.59E-13

TABLE 1 depicts the Power Delay Product over a range of Power Supply voltages and as it is shown in the table that 10 T Modified Designing approach for SRAM Cell shows minimum Power Delay Product.

IV. CONCLUSION

As the battery operated devices are in great demand and to increase their reliability, the life time of battery is a prime concerned but this is done at the cost of speed. But in high speed circuits where speed is the major concern like wireless communications these low-leakage SRAM fails. For low-leakage and high-speed circuits concern should be on both the factors speed and power. This paper tries to find out the solution for SRAM memory cells in both the aspects power consumption and speed or we can say that in terms of power delay product. Modified 10T SRAM Cell shows least power consumption over a range of power supply voltage, operating frequency and operating temperature at the expense of 66.66% area overhead with conventional approach. 6T SRAM Cell shows least delay among all the other design techniques but with the significantly higher power consumption over other approaches. Modified 10 T SRAM Cell shows the least power delay product over a range of supply voltages.

REFERENCES

[1] Jawar Singh, DhirajK.Pradhan et al," A single ended 6T SRAM cell design for ultra low voltage applications",*IEICE Electronic Express*,2008,pp-750-755.

[2] S.S.RathodS.Dasgupt,AshokSaxena , "Investigation of Stack as a Low Power Design Technique for 6-T SRAM cell, *Proceedings IEEE TENCON*, Nov.18-21,Univ.of Hyderabad, 2008,pp 1-5.

[3] Kang, Sung-Mo, Leblebici and Yusuf (1999), "*CMOS Digital Integrated Circuits Analysis and Design*", McGraw-Hill International Editions, Boston, 2nd Edition.

[4] S. Narendra, S. Borkar, V.De, D.Antoniadis, and A.P.Chandrakasan, "Scaling of stack effect and its application for leakage reduction," *Proc. IEEE ISLPLED*, pp. 195-200, Aug. 2001.

[5] C-T. Chu, X. Zhang, L. He and T. Jing, "Temperature aware microprocessor floorplanning considering application dependent power load", in *Proc. of ICCAD*, 2007, pp. 586-589.

[6] Narender Hanchate and Nagarajan Ranganathan, "LECTOR:A Technique for Leakage Reduction in CMOS Circuits," *IEEE Trans., on VLSI Systems*, vol. 12, No.2, Feb 2004.

[7] Aly, R.E. Bayoumi, M.A., "Low-Power Cache Design Using 7T SRAM Cell" *Circuits and Systems II: Express Briefs, IEEE Transactions*, vol. 54 April 2007, Issue: 4, pp. 318-322.

[8] W. Liao, L. He, and K. Lepak, "Temperature-Aware Performance and Power Modeling", *Technical report UCLA Engineering. 04-250*, 2004.

[9] Chang, L. Montoye, R.K. Nakamura, Y.Batson, K.A.Eickemeyer, R.J.Dennard, R.H. Haensch, W.Jamsek, D, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches", *Solid-State Circuits, IEEE Journal* vol. 43, April 2008, Issue 4, pp-956-963.

[10] Benton H. Calhoun Anantha P. Chandrakasan "A 256-kb 65-nm Sub-threshold SRAM Design for Ultra-Low-Voltage Operation",*Solid- State Circuits, IEEE Journa*l vol. 42, March 2007, Issue 3 , pp.680-688.

[11] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power CMOS digital design,' *IEEE J. Solid-State Circ.*, vol. 27, no. 4, pp. 473-484, Apr. 1992.

[12] RajshekharKeerthi,Henry Chen, "Stability and Static Noise margin analysis of low power SRAM"*IEEE International Instrumentation & Measurement Technology Conference*, Victoria Canada, May 2008,pp-1541-1544.

[13] Sherif A.Tawfik, Volkan Kursun, "Stability Enhancement Techniques for Nanoscale SRAM ciruits, International SOC design Conefrence, 2008,pp 113-116.

[14] S. Dutta, S. Nag, K. Roy, "ASAP: A Transistor Sizing tool for speed, area, and power optimization of static CMOS circuits", *IEEE International Symposium on Circuits and Systems*, pp. 61-64, June, 1994.

[15] Hiroki Noguchi et al., "Which is the best dual port SRAM in 45nm process technology?8T,10T single end and 10T differential" *Renesas Technology corporation,2008.*

[16] Farshad Moradi *etal*., "65nm Sub threshold 11 T SRAM for ultra low voltage Application ,*IEEE xplore*,2008,pp-113-117.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-332-337 ISSN: 2249-6645

Secure Cloud by IT Auditing

CHIPURUPALLI SEKHAR[1], U. NANAJI[2]

[1](Department of CSE, St.Theresa Institute of Engg. & Technology, Garividi, Vizayanagaram, (A.P.), India)
[2](HOD, Department of CSE, St.Theresa Institute of Engg. & Technology, Garividi, Vizayanagaram, (A.P.), India)

Abstract
In this paper we discuss the evolvement if cloud computing paradigm and present a framework to provide security to Cloud computing concept through IT Auditing. Our approach is to establish a general framework using several checklists by following data flow and its lifecycle. The lifecycle is based on the cloud deployment models and cloud services models.
The contribution of the paper is to understand the implication of cloud computing and what is meant secure cloud computing via IT. Our approach has strategic value to those who are using or consider using cloud computing because it addresses concerns such as security, privacy and regulations and compliance.

Keywords--- Cloud computing, IT Auditing

1. INTRODUCTION

Cloud computing is Internet ("cloud") based development and use of computer technology ("computing"). It is a style of computing in which dynamically scalable and often virtualized resources are provided aa service over the Internet. Users need not have knowledge of, expertise in, or control over the technology infrastructure "in the cloud" that supports them.

The underlying concept dates back to 1960 when John McCarthy opined that "computation may someday be organized as a public utility"; indeed it shares characteristics with service bureaus which date back to the 1960s. The term cloud had already come into commercial use in the early 1990s to refer to large ATM networks. By the turn of the 21st century, the term "cloud computing" had started to appear, although most of the focus at this time was on Software as a service (SaaS).

A visionary scnario in the cloud is that a thin client interacts with remote cloud operating system to get virtual desktop with a chosen virtual local operating to access virtual data storage and executes applications from anywhere and at anytime. This idea is not new. It can trace back all the way when IBM Watson claimed the world needed only five machine.

But why is it now? At present, IT is reaching a critical Point. Explosion of information is driving 54% growth in storage; large scientific calculation such as weather forecast computation, new medicine, and healthcare informatics is demand-

ing more powerful and faster processing capacity. While in reality, around 85% of computing capacity is idle, average 70% of IT budget is spent on managing IT infra-structure

versus adding new capabilities .On the other hand, technologies like virtual computing, parallel computing, services oriented architecture, autonomic computing are advancing in an unusual pace. In addition,

as the connectivity cost keeps falling, the world is even more flat. Web-based applications over the internet, depicted by cloud, are becoming standard starting applications. People without extensive period of skill training and manual remembering on underline operating systems and basic hardware maintenance can accomplish their work fairly easily. Consumers purchase computing capacity on-demand and are not generally concerned with the underlying technologies used. Computing resources and data being accessed are typically owned and operated by a third-party provider, not necessarily located in nearby. They can be potentially beyond state even country's physical boundary.

In this paper, we address the security issue from information assurance and security point of view. That is, we take holistic view of securing cloud computing by using the IT auditing vehicle.

Types Of Cloud Computing:

Public cloud: An IT capability as a service that providers offer to consumers via the public Internet.

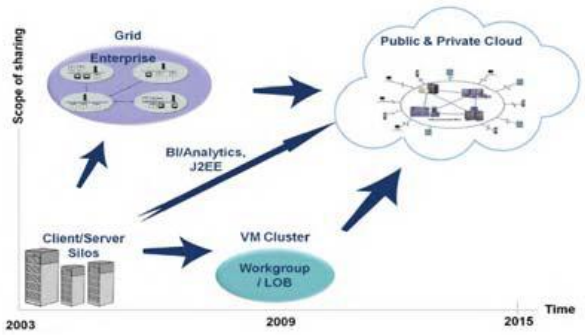
Private cloud: An IT capability as a service that providers offer to a select group of customers.

Internal cloud: An IT capability as a service that an IT organization to its own business (subset of private cloud).

External cloud: An IT capability as a service offered to a business that is not hosted by its own IT organization.

Hybrid cloud: IT capabilities that are spread between internal and external clouds

While there are many roads to cloud computing from existing client-server infrastructures, there are at least three major paths as follows:



IT auditing or in general accounting auditing under cloud computing has added extra role of building strategic plan for the enterprise in addition to the traditional auditing role. We make master checklists as a framework specifically toward cloud computing based on its deployment models and service models. After a section of literature review, we start to make checklists for public cloud, community cloud and private cloud as well as IaaS and SaaS. The last section is devoted to further discussions.

2. Data Life Cycle

Cloud computing makes the world even flatter. Public cloud providers can have their computing resources locally or globally. The sky is the limit in this case. Cloud users do not need to know in theory where the computing resources location is because they are all virtualized. We start with cloud data life cycle. Data and information in this paper are exchangeable terms. In general, the data life cycle includes collection, storage, transferring and destruction. The data collection includes both raw data and derived data. Derived data is also called information that is generated from raw data to deliver intelligence, which is usually not see easily from raw data. Data storage includes active data storage and inactive storage. For example, former employee data can be considered as inactive. Its storage procedure could be different from active employee data. Data processing and data storage are not necessarily under the same location in the cloud environment. Therefore data transferring around the net is more a common activity. Data destruction is to destroy data permanently, no backup should be left somewhere either in the user side or the provider side. We see that cloud data life cycle presents many unique features. Data is crossing different security domain and regulations. Data is constantly moving due to the nature of data storage provided by a third party. Information assurance has a new dimension via contracts among cloud users and cloud providers. They need to establish formal agreement. We can even borrow the term service level agreement (SLA) since we now need to add crossing domain compliance clauses that can be implementable. The following three sections is a framework of checklists for public cloud, community cloud and private cloud. The core is data protection following the data life cycle.

3. Checklist for Public Cloud

Public cloud has its root from Google, Amazon, Microsoft, salesforce, and more. Enterprise uses public cloud to focus on its core business and save cost. Government is keen to use public cloud to take the advantage of cost effective by providing public useful information in cloud. It can also explore cloud concept to integrate various computing resources from different departments and agencies into a manageable pool. Therefore making a connected government is a reality.

IT auditing in public cloud can have different focus based on different service models. We address two popular service models in this paper, Infrastructure as a Service (IaaS) and Software as a Service (SaaS). We will discuss PaaS late.

3.1 IaaS

Infrastructure as a Service (IaaS) is a popular service model that provides computing resources to cloud users who deploy operating systems and run their applications on top of it or use it as a storage or archive. When it comes to IT auditing, location, geopolitics, data owner and regulatory issues are not going to be virtualized. For the public cloud, our check list focuses on the following issues.

1. Data location Aware

Rationale: Cloud computing makes the world even flatter. Public cloud providers can have their computing resources locally or globally. The sky is the limit in this case. Public cloud users do not need to know theoretically where the infrastructure location is because they are all virtualized.

For the IT auditing purpose, public cloud users need to know geometric location of their data storage and their running applications although in general, they do not need to. Public cloud providers o the other hand would like to hide the location information. Knowing the location helps IT auditors understand the applied regulation or study the implication and make proper recommendations and decisions.

What: location aware should include all the history of data location following its life cycle. Pay special attention to those data locate outside legal territory such as in other states or countries.

How: Get these documents from cloud coordinators. Usually it should be in the agreements. Cloud coordinators and IT auditors should talk to cloud providers about location if these documents either not exist or out dated

2. Data ownership aware

Rationale: data owner in public cloud is always a touchy issue between providers and users. We see it happened in the argument among facebook and its users, and a 9th Circuit Court of Appeals ruling stating that providers of hosted e-mail/SMS services may not turn over messages to the company under the Stored Communications Privacy Act

without a warrant. Many cloud users are sure to avoid such situations. So far, no universal legal guidance is established.
Cloud users could assume they are the owners of their data. This assumption should be written in an agreement. When it comes to move data out of the cloud or destroy data, cloud users should know if they data are destroyed completely and how. No backup should be left alone when the data is supposed to discard. The process should be written in the agreement.
What: Clearly stated in the agreement on data ownership on data life cycle. Also included the data destroy and verification process.
How: Discuss with cloud coordinators about the data ownership and data life cycle management. Get written document on data ownership the procedure of data removal.

3. Data protection plan and best practice

Rationale: It is obvious data protection is crucial to cloud users. Detailed data protection plan following data life cycle is important part of agreement among all parties, users, providers and affected stakeholders. In addition to written agreement, actual practice is also important to data protection.
What: Data protection plan should include clear procedure and practice in each phase of data life cycle such as collection, storage, transferring and destruction. The ability of data auditability is an important part of the plan.
How: It auditors needs to understand the classification of essential and non-essential data. With this in mind, they should talk to cloud coordinators and compliance officers to understand what's been done. In addition, IT auditors should suggest various controls like red tape in place for every phase of data life cycle. These controls can report any incidents happened.

4. Data processing isolation

Rationale: Another possibility of data leakage is during the data processing in a shared cloud environment. Data might be stored in a temporary storage accessed by other applications. To isolate data processing and make sure no other applications can access the data during the processing.
What: Processing isolation should have clear procedure to make sure data processing does not leak data
How: IT auditors should not only read document in written but also look for evidence the procedures are followed.

5. Data Lock-in

Rationale: So far, there is no unified cloud user interface to access cloud. Different cloud providers provide different data access method using different format. This will cause an issue when cloud users want to move their data to another provider or back in house. This phenomenon is called data lock-in.
What: To avoid data lock in, cloud users should know the exit strategy and options
How: It auditors should ask such documents that include exit strategy and options.

6. IaaS IT architecture

Rationale: IaaS architecture varies although we see general reference architecture. Knowing the actual architecture, IT auditors can define their work scope and focus easily. Because of the IaaS is new to many management personnel, the IT SaaS architecture could help them visually get the main IT auditing concerns
What: List all the components inside the architecture, not just a general conceptual one. It should include as much detail as possible.
How: Talk to cloud coordinators and cloud providers to get the IT architecture descriptions.

7. Regulatory Compliance

Rationale: many regulatory issues such as HIPAA, GLBA,FISMA, SOX , PCI DSS are new and need to do through investigation when sensitive data are processed and put in cloud. To make compliance in public cloud is a daunting task.
What: Regulatory compliance in terms of public should include privacy, safeguard, security rule, information system Controls, etc.
How: understand the specific needs of compliance for the enterprise by talking to the compliance officers and chief information officers. Collect all the documents and practices.

8. Cloud IT technique

Rationale: IT auditing toward public cloud is challenging because the IT infrastructure basically offered by a third party to which depending on the agreement auditors may not have direct access. Practical IT auditing techniques need to refine to reflect the change.
What: The techniques should include database, data center, wired and wireless connection, cloud operating system like Azure, virtual technology like VMware, hardware dependencies
How: It auditors should find out the agreement using third party cloud provider, how far it can go and test, talk to cloud coordinators what are procedures of reporting any incidents, inspecting specific areas routinely, what kind of tools can use.

9. reporting control

Rationale: cloud control structure should be there with or without cloud presence. It is required by SOX.
Although SOX is for public trading companies, private companies are recommended to do so too. With cloud presence, the reporting and responsible extend to third parties as cloud providers.
What: Reporting structure should be all the way to CIO, CEO or Board of Directors. It includes incidents and response mechanisms involving cloud providers. Usually it is written in an agreement with cloud providers.
How: Ask IT administrators for such documentation. And check if it is in compliance with regulations and best practices

10. Cloud Disaster recovery plan

Rationale: Cloud disaster recovery plan is crucial for business recovers from any disaster. With cloud in place, the disaster recovery process should include them as well.
What: Cloud disaster recovery plan should include how to get crucial data back and how quickly in the case of disaster either on cloud provider side or on the cloud user side. So the plan should include disaster recovery plan from cloud providers.

How: Ask IT administrators for such documentation, and if it is being frequently tested and updated.

11. Cloud business continuity

Rationale: Business glitch is evitable. Damage minimization is carried out by business continuity. Because of cloud computing, the probability being failure is even higher as services are delivered over internet. Business continuity under cloud should include cloud providers' business continuity plans in addition to own business continuity plan.
What: Cloud business continuity should include those foreseeable glitches from inside and outside.
How: Ask IT administrators for such documentation.
Business Continuity plan should be tested frequently.

12. Overall IT projects cost

It is desirable to know the actual cost structure using public cloud and how much saving compared to traditional IT model. It may not related to IT auditing directly. Strategically, the but it tied to IT budget and alignment with

3.2 SaaS

Software as a Service (SaaS) is a popular cloud service model. Applications are accessible various web browsers. It pays for usage. Many checklist items are similar to those from IaaS. We list some of special toward to SaaS.

13. Data activity surrender

Rationale: Some countries require that data and activity from SaaS providers should be kept within the national boundaries so that government agencies can access them when needed. USA has USA Patriot Act that mandates SaaS providers keeps all the customer data that can be accessed under special occasions such as court order. This is not a pleasant outcome many SaaS users want. Therefore SaaS users should ask if there is a possibility that can avoid such intrusion.
What: data activity surrender should include what regulation and laws apply, what information is being surrendered and what option available to avoid such surrender.
How: IT auditors should work with legal and contract team to understand the local law and regulation on data service providers such as phone records, utility bills, library book lending records, etc. T hey should ask documents about what kind of information cloud providers keep and to surrender. This documented policy should be checked on site. It auditors should also ask what option available to avoid surrender

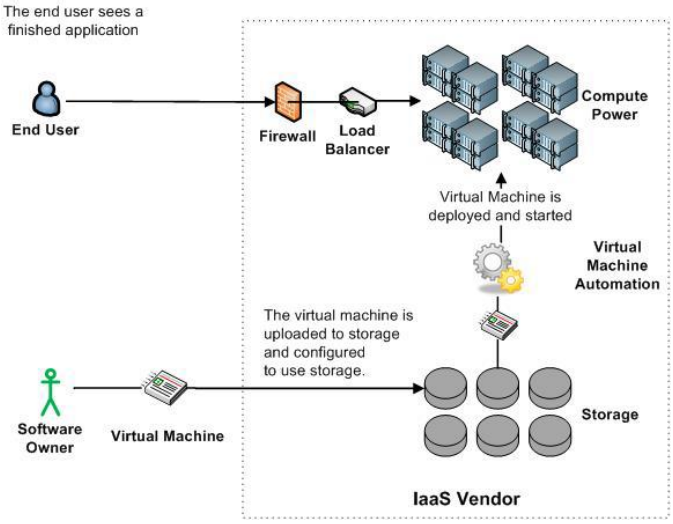


Fig. 2 Architectural diagram of Software as a Service..

14. Data format

Rationale: If data format from specific software can be read by many freely available readers like adobe, work, open office and notepad, SaaS users can avoid pay extra software usage.
What: Check available data format from the software service.
How: It auditors should test out all availability data format and check if these format can be accessed by general reader applications. They should talk to users to find out reasons that specific format being used or not used.

15. Monitoring for availability and performance

Rationale: From SaaS providers, assuring high level of availability and performance is the key for their business success. For SaaS users, monitoring high availability and performance is an important control.
What: The monitoring control should collect availability and performance data and use the data to work with providers to fine tune SaaS service.
How: Talk to cloud coordinator about such type of controls and data for offline analysis.

4. Checklist for Community Cloud

Community cloud is another attractive solution to many large and middle size enterprises that have common business interests within one region. Many corporations are used to regulations within one region or one country. They would like to take the advantages of cloud computing but they do not want to have the complication of cloud crossing their comfortable zones. They want both convenient and control. Hence, community cloud computing is an attractive concept for them. The actual implementation of community cloud varies. It can be built on from scratch. Or it can utilize existing computing resources from the community to form a cloud.
The ownership of the cloud is open to discuss. Ideally it should be owned by the community. The technical challenging is enormous and its implication remains to be seen The IT auditing under community cloud computing is more of under controllable like private cloud. They know where the computing resources located and assigned. It is certainly more work than a pure private cloud as it is owned by a community in which collaboration and competition exist

as always. The key to its success is to make sure agreements are clearly written and the whole community needs to obey the rules and regulation. We have made the following list for IT Auditing specially for community cloud computing

16. Community cloud IT architecture

Rationale: Clear cloud community IT architecture will help the community understand its computing resources and its capacity. It can help to build responsible community.
What: The Community Cloud IT architecture should include core infrastructure, resource layer and service layer .Data life cycle, user identifies and trust are important elements. Use case analysis is proper. If the cloud is built upon the contribution of the community, the relation among own computing resources and the community resource should be marked clearly.
How: Talk to cloud administrators to get a sense of its architecture. If possible, get technical documents that describe the community cloud.

17. Community Cloud management

Rationale: Because community cloud is used by a group of industrials of same regions or of same interests, they are collaborative and in the mean time competitive. Clear community cloud management structure will prevent any future arguments. Community cloud can be managed by a third party or by a technical committee made from the community.
What: It should include clear management structure and responsibility. It should include the procedure in case of argument and disagreement.
How: IT auditors from a specific member of the community should have access the latest documents on management.

18. EXIT Strategy

Rationale: When a member wants to leave the community, the community cloud should have procedure of the departure. Without the procedure and agreement, community cloud could not last long.
What: The Exit strategy should include documents on separation. The document should include procedure to follow and a responsible body for the task. The procedure should clear state what to do about the computing resources the member contributed before.

5. Checklist for Private Cloud

So far, a private cloud is a very practical and attractive option to many security sensitive enterprises. The private cloud gives not only the self control but also the benefits of cloud computing, mainly sharing computing resources including processing power and storage capacity among different departments within an enterprise. Traditionally, department computing resources are not shared due to data sensitivity, self control and different business nature of departments. Private cloud could remove or blur these boundaries. It virtualizes all computing resources from different departments into a computing resource pool. Each department is allocated computing resources from the pool by provisioning need on demand. From the department point of view, the computing resource is unlimited. Therefore achieving a task faster or making a task not achievable before due to computing power constrain. In most cases, a private cloud could cut IT cost down, increase flexibility and scalability, make available 24x7 and even do applications that are impossible before the cloud. Private cloud certainly poses a great management challenging as well as auditing challenging.

19. Private cloud IT architecture

Rationale: Different enterprise implements private cloud differently from actual technology realization. Therefore to understand the cloud IT architecture is vital for meaningful IT auditing
What: IT architecture includes technical details about virtualization, provisioning, workflow, data movement, access control, etc.
How: Talk to cloud administrators to get a sense of its architecture. If possible, get technical documents that describe the cloud.

20. Private cloud reporting control

Rationale: Like public cloud, reporting is required not only by the regulation but also vital to the success of the business. Private cloud reporting is more of internal control. Because of sharing computing resources, private cloud has to make sure that sharing does not hamper security and privacy. Any incidents should be logged and reported immediately. The reporting structure should be established and updated often.
What: Reporting structure should includes incidents and response mechanisms and who is in charge. The escalating reporting structure can guarantee any incident and disaster can be handled properly.
How: Ask IT administrators for such documentation. And check if it is updated.

21. Disaster recovery and continuity plan

Rationale: In the private cloud, disaster recovery plan should also follow the procedure like public cloud except that the cloud is managed by the enterprise. The IT team and management should work together to modify the existing disaster plan to fit the cloud scenario.
What: The disaster recovery plan should include how to get crucial data back and how quickly. The plan should include data different location backup.
How: Ask IT administrators for such documentation, and if it is being frequently tested and updated. Gramm-Leach-Bliley Act (GLBA, the Financial Services Modernization Act),

6. Discussion

In this paper, we discussed a framework of checklist of IT auditing cloud computing to assure secure cloud computing. It is more toward Cloud than a complete list of IT Auditing. IT auditors should refer general requirements for IT auditing. The checklist also gives a reference point to those want to dive into cloud computing wave and a question set to answer if cloud is good for the business in long run. We would like to discuss on PaaS service model in the future work as we are still looking for a feasible PaaS business model.

References

[1] NIST Definition of Cloud Computing v15, accessed on 4/15/2010, http://csrc.nist.gov/groups/SNS/cloudcomputing/cloud-def-v15.doc
[2] Will Forrest, Clearing the Air on Cloud Computing, Discussion Document from McKinsey and Company, March 2009
[3] Luis M Vaquero, et al, A Breaks in the Clouds: Toward the Definitions, ACM SIGCOMM Computer Communication Review, V39 No1, January, 2009, pp 50-55.
[4] open crowd cloud computing taxonomy, http://www.opencrowd.com/views/
[5] NIST Presentation on Effectively and Securely Using the Cloud Computing Paradigm v26, accessed on 4/15/2010, http://csrc.nist.gov/groups/SNS/cloudcomputing/cloud-computing-v26.ppt
[6]FISMA: ttp://csrc.nist.gov/drivers/documents/FISMAfinal.pdf
[7] Gramm-Leach-Bliley Act (GLBA, the Financial Services Modernization Act), http://www.gpo.gov/fdsys/pkg/PLAW-106publ102/contentdetail.html.
[8] HIPAA U.S. Department of Health & Human Services, Office of Civil Rights, HIPAA, http://www.hhs.gov/ocr/hipaa/privacy.html
[9] Cloud Computing: Principles and Paradigms By Benoit Hudzia.

Mr.Ch. Sekhar received the B.Tech Degree from JNTU, Hyderabad in 2005 and He is currently pursuing M.Tech in the Department Of Computer Science and Engineering in St. Theresa Institute Of Engg & Tech Garividi, Vizianagaram, Of JNTUK Affiliation. His research interests include Cloud computing and Computer Networks.

Uppe.Nanaji received the B. Tech degree from JNTU, Hyderabad, India and the M. Tech degree in Computer Science Technology from GITAM College Of Engg Of Andhra University Affiliation in Vishakhapatnam in 2003, and he is currently pursuing the Ph. D in Computer Networks from Andhra University Visakhapatnam. He is working as a Head of the Department for CSE in Saint Theresa College Of Engg & Technology Garividi, Vizianagartam (Dist) India. His research interests include Computer Networks & Data Ware Housing

Robustness of Parity Checker Method against Various Watermarking Attacks

Rajkumar Yadav
Assistant Professor
U.I.E.T, M.D.U, Rohtak

Abstract
In the 21th century, information security has become a geart issue. Steganography and watermarking provide solution to these issues. Watermarking is mainly used for copyright protection. There are many techniques which have been developed for watermarking in the past decade both in the spatial and frequency domain. In this paper, robustness of Parity Checker Method [1] which is a spatial domain technique is checked has been checked against the two watermarking attacks blurring and cropping .By analysis of the result we found that this method provides favorable results.

Keywords: Steganography , Watermarking, Robustness, Parity Checker Method

1. Introduction

In the recent years, with the growth of multimedia system in distributed environment, the problem associated with multimedia security and multimedia copy right protection have become important issues. Also the technology designed to make electronic publishing feasible has also increased the threat of intellectual property threat. Illegal copying and redistribution of digital images, audio or video without any information loss is also a threat to the society. These issues can be solved by using water marking techniques available [2, 3, 4, 5, 6]. The process of digital watermarking involves the modification of original multimedia data to embed a watermark containing the key information as a authentication or copy right codes. The embedding method must leave the original data perceptually unchanged, yet should imposed modification which can be detected by using an appropriative extraction algorithm. A water mark is an imperceptible, robust and secure message embedded directly in digital elements such as image, audio, and sound which uniquely identifies its owner. It should be noted that digital water mark could not itself prevent copying, modification and redistribution of documents [7]. However if encryption and copy protection fails, water marking allows the documents to be traced back to its right owner and prevents unauthorized use. The water mark must be difficult to remove and immune to multimedia data operations. A water mark containing the information regarding owner should be small in size so that it can be easily embedded into images. The water mark can also be embedded as a noise component in image. In general, the watermark can be visible or invisible. A visible watermark typically contains a evidently visible message or a company logo indicating the ownership of the image. The invisible watermark contents appear perceptually identical to the original.

In this paper, the robustness of parity checker method has been checked against the two watermarking attacks. In parity checker method, parity of the the pixel value is checked to insert the watermark bit. The watermark bit inserted at a pixel position according to the parity of the pixel value. The analysis of this technique against the blurring and cropping attack show the favorable results.

The rest of the paper is organized as follows: Section 2 describes the various attacks on the watermarking. Parity Checker Method has been given in the section 3. At last , section 4 gives the result and analysis.

2. Attacks on Watermarking

There may be many attacks on watermarked image namely, blurring, cropping, compression, scaling etc. Here in this paper blurring and cropping have been discussed in section four.

3. Parity Checker Method [1]

In this method, the concept of even and odd parity has been used by using the parity checker. As we already know that even parity means that the pixel value contains even number of 1's and odd parity means that the pixel value contains odd number of 1's. In this method '0' bit is inserted at a pixel value where pixel value has odd parity and if the parity is even then odd parity is made by adding or subtracting '1' to the pixel value. Similarly, '1' is inserted at a pixel value if it had even parity. In case, if even parity is not present at that location then even parity is made over that location by adding or subtracting '1'. In this way '0' or '1' is inserted at any location. The insertion process is shown in figure 1 and 2.

For retrieval of message, again parity chacker is used. If odd parity is present at the selected location then '0' is message bit, else message bit is '1'. Retrieval process is repeated for all locations. In this way, the message bits are retrieved bits from all the locations where these have been inserted. The retrieval process is shown in Figure 3.

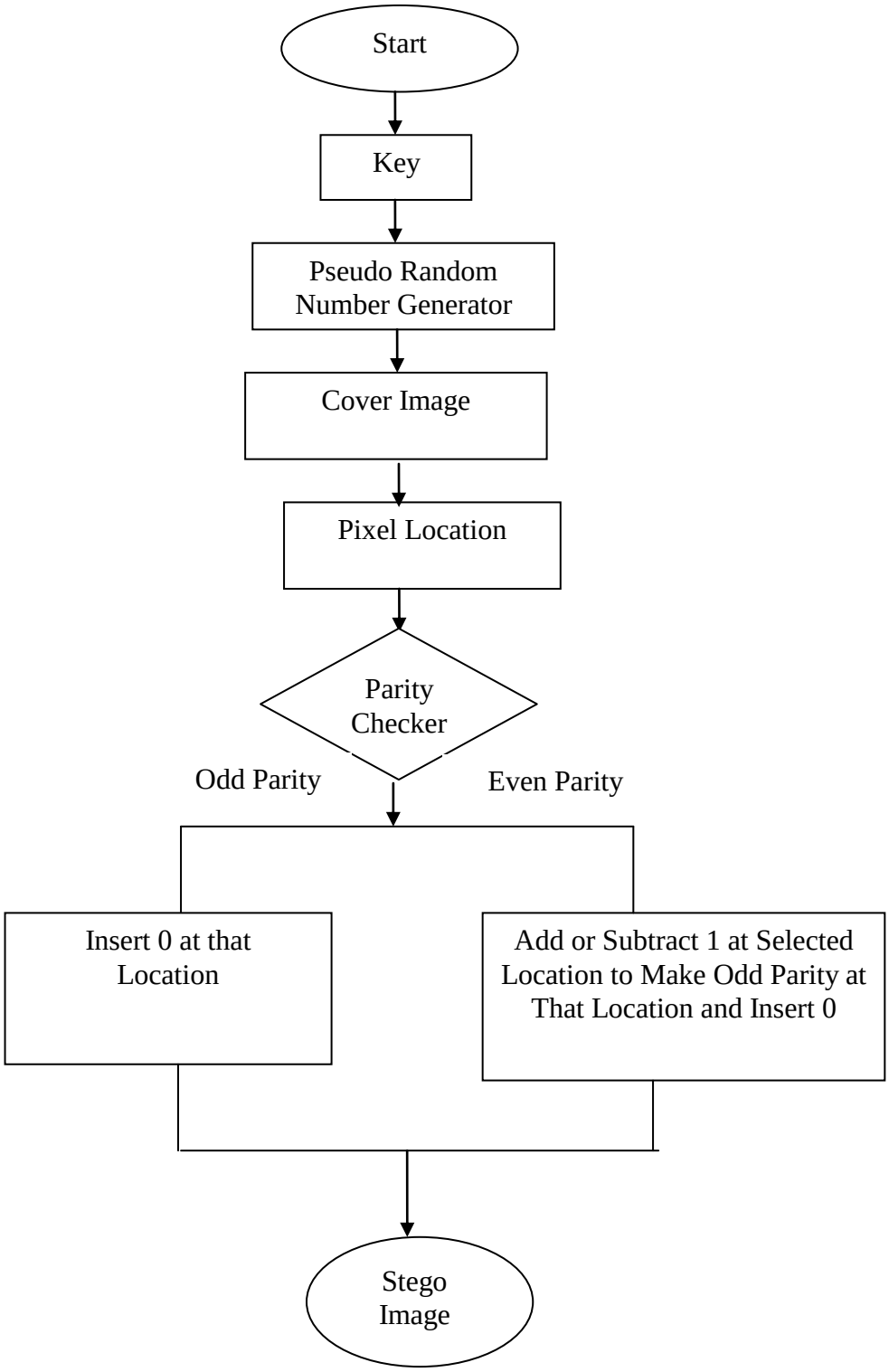


Fig. 1 (Insertion of 0)

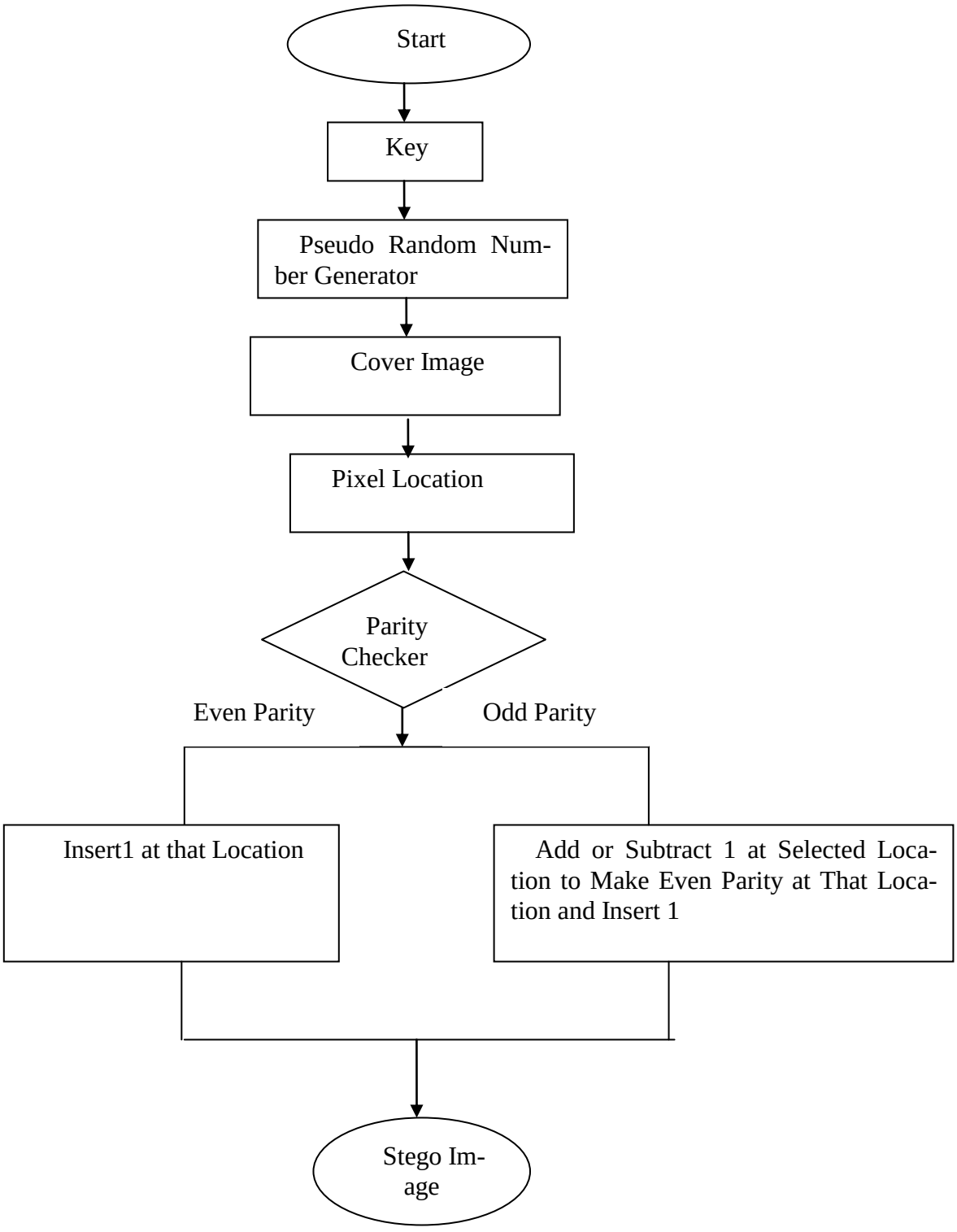


Fig.2 (Insertion of 1)

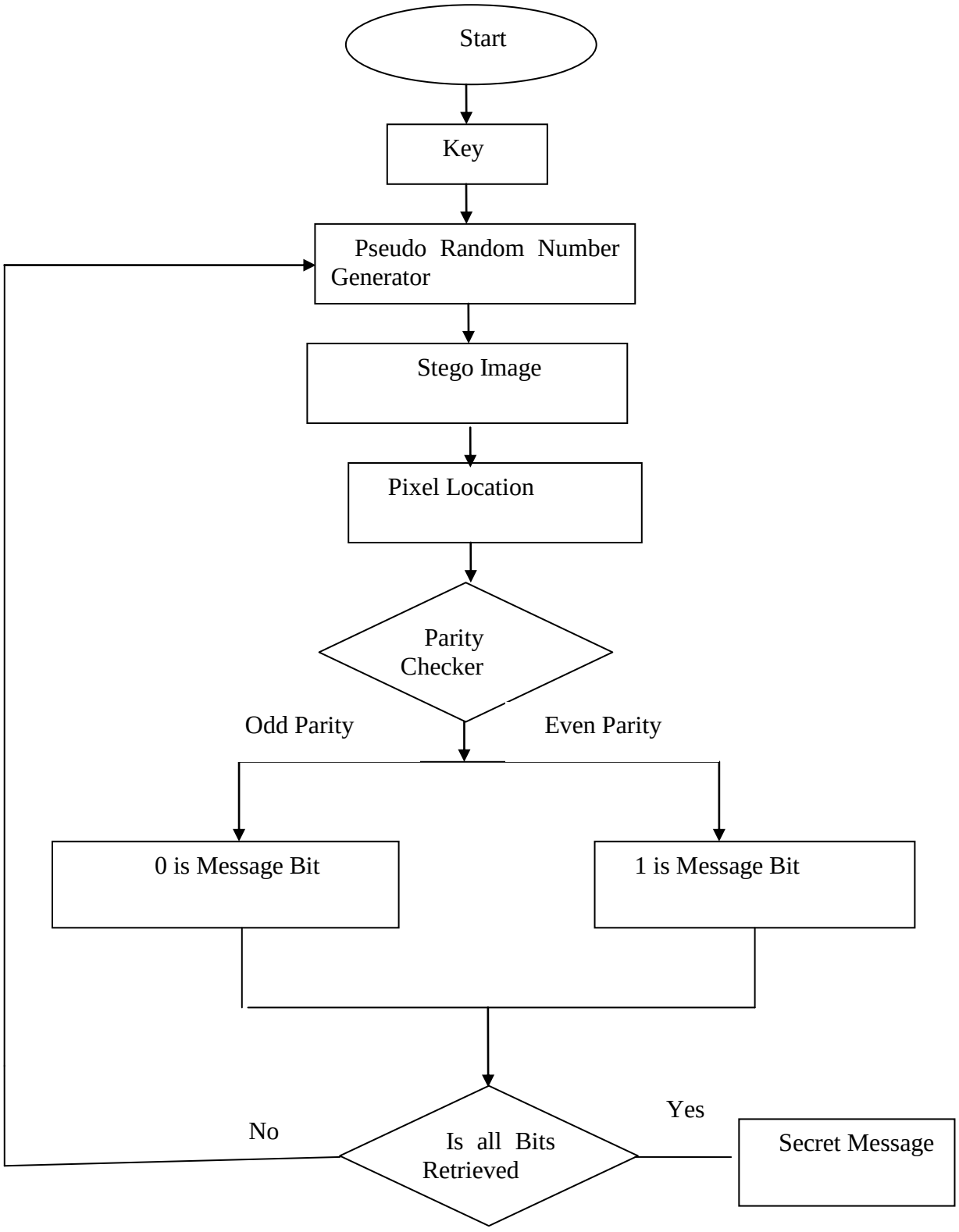


Fig. 3 (Retrieval)

3.4.2 Algorithm [1]

3.4.2.1 Assumption

(i) Sender and recipient agree on the cover object in which message is supposed to be hidden.

(ii) Both sender and recipient agree on the same pseudo–random key to decide the random locations where message is to be inserted.

3.4.2.2 Insertion Algorithm

(i) Find pseudo-random location (L) in cover image from secret key to insert the message bit. (For detail see [Franz et al (1996)] and [Lee and Chen (2000)]).

(ii) If we want to insert 0 then go to step (iv) else go to step (v).

(iii) (a) Check whether at location (L) pixel value is having odd parity. If yes, insert 0 at location 'L' and go to END If no, go to step (b)

(b) Make the parity of pixel value odd by adding or subtracting 1 and then insert 0. Go to END

(v) (a) Check wheather at location 'L' the pixel value is of even parity. If yes, insert 1 at location (L) and go to END. If no, go to step (b).

(b) Make the parity of pixel value by adding or subtracting 1 and then insert 1 and go to END.

(vi) END

3.4.2.3 Retrieval Algorithm

(i) Trace out the location (L) from the same secret key as used for insertion of message.

(iii) Check whether at location (L).

(a) If the parity of pixel value is odd then '0' is the message bit.

(b) If the parity of pixel value is even then '1' is the message bit

(iv) END

4. Results and Analysis

Here in this section robustness of watermarking scheme against two attacks i.e. cropping and blurring of imagehas been analyzede. Figure 4 shows the original image and figure 5 shows the watermarked image with watermark 'Rajkumar' inserted four times in the original image. Figure 6 shows the blurred image and figure 7 shows the cropped image.

Fig.4 Original Image

Fig.5 Watermarked Image

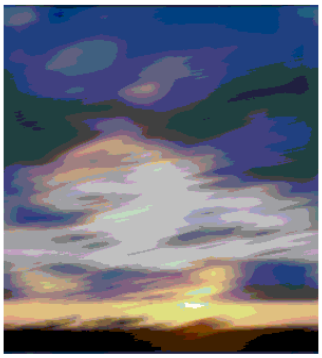


Fig.6 Blurred Image

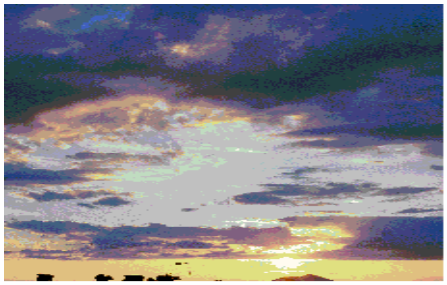


Fig.7 Cropped Image

From this we can easily analyze the above images based on PSNR (Peak signal to noise ratio) in dB by using the software HAYDWT Video watermark. The PSNR values of original image and stego image are given in table 1 Table 2 shows the PSNR values under the various attacks

Table 1

Component	PSNR (Original Image and Stego Image)
Red	**9.59**
Green	**10.03**
Blue	**11.28**

Table 2

Attack	PSNR of Red component (Original Image and Attacked Image)	PSNR of Green component (Original Image and Attacked Image)	PSNR of Blue component (Original Image and Attacked Image)
Blurring	**7.84**	**8.53**	**9.50**
Cropping	**9.67**	**9.62**	**9.51**

After analyzing the results from table 4.3 and 4.4 it is found that PSNR values decrease after applying the various attacks. Under the blurring attack the PSNR value of red, green, and blue component decreases by 1.35dB, 1.54dB, and 1.78dB respectively. Similarly, in cropping the PSNR values of green and blue components decreases by 0.41dB and 1.77 dB respectively. There is slightly in the PSNR values in the red component i.e. 0.08dB. So on the basis of above facts it is concluded that there is very less change in PSNR values under the various attacks which shows the robustness of watermarking technique.

5. References

1. Rajkumar Yadav, Rahul Rishi & Sudhir Batra, "A New Steganography Method for Gray Level Images using Parity Checker", International Journal of Computer Applications (0975-8887) Volume 11-No. 11, December 2010.
2. Su, P.C., Kuo, C.C.J. and Wang, H.J.M. (1999), "Blind digital watermarking for cartoon and map images", In: Security and Watermarking of Multimedia Contents. Wong PW, Delp EJ. (eds.); 3657: 296–306.
3. Andres, F. (2001), "Multimedia and Security", IEEE Multimedia, pp 20-21.
4. Mintzer, F. and Braudaway, G.W. (1999), "If one watermark is good, are more better?", In: International Conference on Acoustics, Speech and Signal Processing. IEEE Signal Processing Society.: 2067–2070. ISBN 1-876346-19-1.
5. Mintzer, F., Braudway, G.W. and Yeung, M.M. (1997), "Effective and Ineffective Digital Watermarks", IEE ICIP vol III, pp 9-12, Santa-Barba, Cal.

6. Peticolas, F.A.P., Anderson, R.J. and Kuhn, M.G. (1999), "Information Hiding: A Survey", Proceedings of IEEE, 87, no. 7, pp 1062-1078.

7. Eager, J.J. and Girod, B. (2001), "Quantization Effect on Digital Watermark", Signal Processing, vol 81, no 2, pp 239-263, EURASIP.

8. I.Cox,J Kilan, " Secure Spread Spectrum Watermarking for Images,Audio and Video" , in Proc. IEEE International Conference on Image Processing ,1996,vol 3,pp. 243-246.

9. S.Craver ,N. Memon , "Resolving Rightful Ownership with Invisible Watermarking Techniques:Limitations,Attacks and Implications",IEEE Trans.,Vol 16,No. 4,pp. 573-586,1998.

10. Chun-Yu-Chang,"The Application of a Full Counterpropagation Neural Network to Image Watermarking", 2005, IEEE

11. H. Y. Gao, The theory and application of audio information hiding, PH.D. dissertation, Beijing university of Posts and Telecommunications, Beijing, China, 2006.

12. J. F. Delaigle, C. Devleeschouwer, B. Macq et al., "Human visual system features enabling Watermarking J," in Proceedings of IEEE International Conference on Multimedia and Expo, pp. 489–492, Lusanne, Switzerland, 2002.

13. S. Katzenbeisser and F. A. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, ArtechHouse Press, Norwood, Mass, USA, 2000.

14. W. Bender,D. Gruhl,N.Morimoto, and A. Lu, "Techniques for data hiding," IBM Systems Journal, vol. 35, no. 3-4, pp. 313–335, 1996.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-347-351 ISSN: 2249-6645

Efficient Query Optimizing System for Searching Using Data Mining Technique

Velmurugan.N
Assistant Professor, Department of MCA,
Saveetha Engineering College,
Thandalam,Chennai-602 105.

Vijayaraj.A
Associate Professor, Department of IT
Saveetha Engineering College,
Thandalam,Chennai-602 105.

ABSTRACT

There is a critical need to design and develop tools that abstract away the fundamental complexity of XML based Web services specifications and toolkits, and provide an elegant, intuitive, simple, and powerful query based invocation system to end users. Web services based tools and standards have been designed to facilitate seamless integration and development for application developers. As a result, current implementations require the end user to have intimate knowledge of Web services and related toolkits, and users often play an informed role in the overall Web services execution process .We employ a set of algorithms and optimizations to match user queries with corresponding operations in Web services, invoke the operations with the correct set of parameters, and present the results to the end user. Our system uses the Semantic Web and Ontologies in the process of automating Web services invocation and execution. Every user has a distinct background and a specific goal when searching for information on the Web. The goal of Web search personalization is to tailor search results to a particular user based on that user's interests and preferences. Effective personalization of information access involves two important challenges: accurately identifying the user context and organizing the information in such a way that matches the particular context. We present an approach to personalized search that involves building models of user context as ontological profiles by assigning implicitly derived interest scores to existing concepts in domain ontology.

Keywords: intuitive, seamless, optimizations, semantic Web, ontology, tailor search.

Introduction

A spreading activation algorithm is used to maintain the interest scores based on the user's ongoing behavior. Our experiments show that re-ranking the search results based on the interest scores and the semantic evidence in an ontological user profile is effective in presenting the most relevant results to the user. With the tremendous growth of information available to end users through the Web, search engines come to play ever a more critical role. Nevertheless, because of their general purpose approach, it is always less uncommon that obtained result sets provide a burden of useless pages. Next generation Web architecture, represented by Semantic Web, provides the layered architecture possibly allowing to overcome this limitation. Several search engines have been proposed, which allow to increase information retrieval accuracy by exploiting a key content of Semantic Web resources, that is relations. However, in order to rank results, most of the existing solutions need to work on the whole annotated knowledge base.

we propose a relation-based page rank algorithm to be used in conjunction with Semantic Web search engines that simply relies on information which could be extracted from user query and annotated resource. Relevance is measured as the probability that retrieved resource actually contains those relations whose existence was assumed by the user at the time of query definition. We address the problem of supporting efficient yet privacy-preserving fuzzy keyword search services over encrypted cloud data. Specifically, we have the following goals: i) to explore new mechanism for constructing storage efficient exact keyword sets; ii) to design efficient and effective fuzzy search scheme based on the constructed keyword sets; iii) to validate the security of the proposed scheme.

Existing System

Searches for the web pages of a person with a given name constitute a notable fraction of queries to Web search engines. A query would normally return web pages related to several namesakes, who happened to have the queried name, leaving the burden of disambiguating and collecting pages relevant to a particular word (from among the namesakes) on the user. Many dynamically generated sites are not indexable by search engines; this phenomenon is known as the invisible web. Some search engines do not order the results by relevance, but rather according to how much money the sites have paid them. Some sites use tricks to manipulate the search engine to display them as the first result returned for some keywords. This can lead to some

search results being polluted, with more relevant links being pushed down in the result list.

Proposed System

We develop web Search approach that clusters web pages based on their association to different people. Our method exploits a variety of semantic information extracted from web pages, such as named entities and hyperlinks, to disambiguate among namesakes referred to on the web pages. We demonstrate the effectiveness of our approach by testing the efficiency of the disambiguation algorithms and its impact on person search. Our system uses word variants or stemming technology, which not only searches for the words present in the user query but also for similar words. This is implemented by domain independent technologies like thesaurus matching as well as by the use of Semantic Web and ontology technologies.

Feasibility study:

A feasibility study is an evaluation of a proposal designed to determine the difficulty in carrying out a designated task. Generally, a feasibility study precedes technical development and project implementation.

Technology and system feasibility:

The assessment is based on an outline design of system requirements in terms of Input, Processes, Output, Fields, Programs, and Procedures. This can be quantified in terms of volumes of data, trends, frequency of updating, etc. in order to estimate whether the new system will perform adequately or not. This means that feasibility is the study of the based in outline.

Economic feasibility:

Economic analysis is the most frequently used method for evaluating the effectiveness of a new system. More commonly known as cost/benefit analysis the procedure is to determine the benefits and savings that are expected from a candidate system and compare them with costs. If benefits outweigh costs, then the decision is made to design and implement the system. An entrepreneur must accurately weigh the cost versus benefits before taking an action. Time Based: Contrast to the manual system management can generate any report just by single click .

Cost Based: No special investment is needed to manage the tool. No specific training is required for employees to use the tool. Investment requires only once at the time of installation. The software used in this project is freeware so the cost of developing the tool is minimal

Legal feasibility:

Determines whether the proposed system conflicts with legal requirements, e.g. a data processing system must comply with the local Data Protection Acts.

Operational feasibility:

Is a measure of how well a proposed system solves the problems, and takes advantages of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.

Schedule feasibility:

A project will fail if it takes too long to be completed before it is useful. Typically this means estimating how long the system will take to develop, and if it can be completed in a given time period using some methods like payback period. Schedule feasibility is a measure of how reasonable the project timetable is. Given our technical expertise, are the project deadlines reasonable? Some projects are initiated with specific deadlines. You need to determine whether the deadlines are mandatory or desirable.

Market and real estate feasibility:

Market Feasibility Study typically involves testing geographic locations for a real estate development project, and usually involves parcels of real estate land. Developers often conduct market studies to determine the best location within a jurisdiction, and to test alternative land uses for a given parcels. Jurisdictions often require developers to complete feasibility studies before they will approve a permit application for retail, commercial, industrial, manufacturing, housing, office or mixed-use project. Market Feasibility takes into account the importance of the business in the selected area.

Resource feasibility:

This involves questions such as how much time is available to build the new system, when it can be built, whether it interferes with normal business operations, type and amount of resources required, dependencies, etc. Contingency and mitigation plans should also be stated here.

SYSTEM DESIGN

OVERVIEW OF DESIGN

Design is multi-step process that focuses on data structure software architecture, procedural details, and interface between modules. Design is the place where quality is fostered in software engineering. Design is the perfect way to accurately translate a customer's requirement in to a finished software product. The design of an information system produces the details that state how a system will meet the requirements identified during analysis. The emphasis is on translating the performance, requirements into design specifications. The various steps

involved in designing the **"Step Construction Using Visual Cryptography Schemes"** are given below.

- First, decide how the output is to be produced in what format.
- Second, the input data can communicate with applications have to be designed based on the requirements.
- Finally, details related to the justification of the system to be presented.

INPUT DESIGN

It is the process of converting input data to the computer-based data. The goal of designing is to make data entry as easier and free from error as possible. Input design determines the format and validation criteria for data entering the system. Personal computers and terminals can place a data at user's finger tips, allowing them to call up specific data and make timely decisions based on the data.

This system contains data collection screen which display heading the defined their purpose. By employing flashing error messages, and providing necessary alerts on the screen, mist entering of data in the system is avoided.

OUTPUT DESIGN

Computer output is the most important and the direct source of information to the user. Efficient and intelligible output design should improve the system relationships with the user and help in decision making.

Major forms and Web Pages of output are hard copy from the printer and the soft copy from the CRT Display. Output is the key tool to evaluate the performance of software so the designing of output should be done with great care. It should be able to satisfy the user's requirements.

CODE DESIGN

A group of characters used to identify and item of data is a code. A major problem encounter in working with a large amount of data is the retrieval of specific dada when it is required. Code facilitated easier identification simplification in handling and retrieval of item by In the developed system a suitable coding is adopted, which can identify each user exactly.

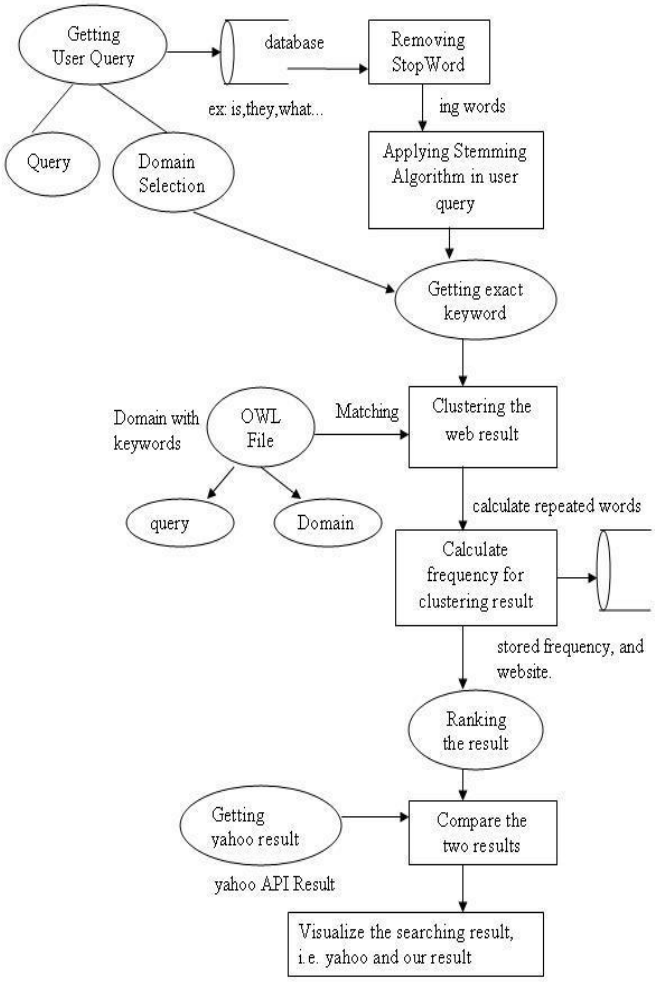


Fig: Data Flow Diagram

Module Description:

- Getting user i/p and stem the keyword
- Web page information retrieval
- Clustering and ranking the web pages
- Implement user profile searching
- Implement relation based searching
- Compare search result

Getting user i/p and stem the keyword

A user submits a query to the middleware via a specialized Web-based interface. User input is involved in various process such as stop word remover, stemming etc. The user query divided in to stopword and addword. Stop word is nothing but which,they,what,where etc… addword is nothing but keywords. The search engine got that keywords and process is performed. Finally we apply the steeming algorithm and get the stem word.This Algorithm attempt to reduce a word to its stem or root form. Thus, the

key terms of a query or document are represented by stems rather than by the original words. This not only means that different variants of a term can be conflated to a single representative form – it also reduces the dictionary size, that is, the number of distinct terms needed for representing a set of documents. A smaller dictionary size results in a saving of storage space and processing time. the words that appear in documents and in queries often have many morphological variants. Thus, pairs of terms such as "computing" and "computation" will not be recognised as equivalent without some form of natural language processing (NLP).

Web page information retrieval

The middleware queries a search engine with this query via the search engine API and retrieves a fixed number (top K) of relevant web pages. The retrieved web pages are preprocessed: . TF/IDF. Preprocessing steps for computing TF/ IDF are carried out. They include stemming, stop word removal, noun phrase identification, inverted index computations, etc. Named entities (NEs) and Web related information is extracted from the web pages. Information Retrieval (IR) is essentially a matter of deciding which documents in a collection should be retrieved to satisfy a user's need for information. The user's information need is represented by a query or profile, and contains one or more search terms, plus perhaps some additional information such importance weights. Hence, the retrieval decision is made by comparing the terms of the query with the index terms (important words or phrases) appearing in the document itself. The decision may be binary (retrieve/reject), or it may involve estimating the degree of relevance that the document has to the query.

Clustering and Ranking the pages:

The clustering algorithm takes the graph, TF/IDF values, and model parameters and disambiguates the set of web pages . The result is a set of clusters of these pages with the aim being to cluster web pages based on association to real person. A set of keywords that represent the web pages within a cluster is computed for each cluster. The goal is that the user should be able to find the person of interest by looking at the sketch. All clusters are ranked by a chosen criterion to be presented in a certain order to the user. Once the user hones in on a particular cluster, the web pages in this cluster are presented in a certain order, computed on this step.

Implement user profile searching:

Personalized web search system, which can learn a user's preference implicitly and then generate the user profile automatically. When the user inputs query keywords, more personalized expansion words are generated by the proposed algorithm, and then these words together with the query keywords are submitted to a popular search engine such as Baidu or Google. These expansion words can help search engines retrieval information for a user according to his/her implicit search intentions, and return different search results to different users who input the same keywords.

Implement relation based searching

With the tremendous growth of information available to end users through the Web, search engines come to play ever a more critical role. Nevertheless, because of their general purpose approach, it is always less uncommon that obtained result sets provide a burden of useless pages. Next generation Web architecture, represented by Semantic Web, provides the layered architecture possibly allowing to overcome this limitation. Several search engines have been proposed, which allow to increase information retrieval accuracy by exploiting a key content of Semantic Web resources, that is relations. However, in order to rank results, most of the existing solutions need to work on the whole annotated knowledge base. In this paper we propose a relation-based page rank algorithm to be used in conjunction with Semantic Web search engines that simply relies on information which could be extracted from user query and annotated resource. Relevance is measured as the probability that retrieved resource actually contains those relations whose existence was assumed by the user at the time of query definition.

Compare search result

Compare existing yahoo result for users given query and our modern relation **s**earch result for users given query and property matching and visualize the results in both search engine based on indexing.

Conclusion

We formalize and solve the problem of supporting efficient yet privacy-preserving fuzzy search for achieving effective utilization of remotely stored encrypted data in Cloud Computing. We design an advanced technique (i.e., wildcard-based technique) to construct the storage-efficient fuzzy keyword sets by exploiting a significant observation on the similarity metric of edit distance.

Future Enhancement

Based on the constructed fuzzy keyword sets, we further propose an efficient fuzzy keyword search scheme. Through rigorous security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search. we will continue to research on security mechanisms that support: 1) search semantics that takes into consideration conjunction of keywords, sequence of keywords, and even the complex natural language semantics to produce highly relevant search results; and 2) search ranking that sorts the searching results according to the relevance criteria.

REFERENCES

- Yi-Hong Chu, Jen-Wei Huang, Kun-Ta Chuang, De-Nian Yang, Member, IEEE, and Ming-Syan Chen, Fellow, IEEE
- C.C. Aggarwal, A. Hinneburg, and D. Keim, "On the Surprising Behavior of Distance Metrics in High Dimensional Space," Proc. Eighth Int'l Conf. Database Theory (ICDT), 2001.
- C.C. Aggarwal and C. Procopiuc, "Fast Algorithms for Projected Clustering," Proc. ACM SIGMOD Int'l Conf. Management of Data, 1999.
- C.C. Aggarwal and P.S. Yu, "The IGrid Index: Reversing the Dimensionality Curse for Similarity Indexing in High Dimensional Space," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2000.
- [R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications," Proc. ACM SIGMOD Int'l Conf.
- Management of Data, 1998.

I. Assent, R. Krieger, E. Muller, and T. Seidl, "DUSC: Dimensionality Unbiased Subspace Clustering," Proc. IEEE Int'l Conf. Data Mining (ICDM), 2007.

- K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is Nearest Neighbors Meaningful?" Proc. Seventh Int'l Conf. Database Theory (ICDT), 1999.
- Blum and P. Langley, "Selection of Relevant Features and Examples in Machine Learning," Artificial Intelligence, vol. 97, pp. 245-271, 1997.
- M.-S. Chen, J. Han, and P.S. Yu, "Data Mining: An Overview from Database Perspective," IEEE Trans. Knowledge and Data Eng., vol. 8, no. 6, pp. 866-883, Dec. 1996.
- C.H. Cheng, A.W. Fu, and Y. Zhang, "Entropy-Based Subspace Clustering for Mining Numerical Data," Proc. Fifth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 1999.
- www.w3schools.com

Authors Profile:

A.Vijayaraj is an Associate Professor in Department of Information Technology at Saveetha Engineering College. He received his Master of Computer Application in Bharathidhasan University, in 1997 and his Master of Engineering in Computer Science and Engineering from Sathyabama University at 2005. He has 12 years of teaching experience from various Engineering Colleges during tenure he was Awarded **Best Teacher Award** twice..He is a Member of, CSI and ISTE. He has Published 2 papers in International journal and 10 Papers in International and National Level conferences. His area of interest includes Operating Systems, Data Structures, Networks and Communication

N.Velmurugan is an Asst.Professor in Department of Computer Applications at Saveetha Engineering College. He received his Master of Computer Application in Bharathidhasan University, in 1999 and his Master of Engineering in Computer Science and Engineering from Sathyabama University at 2011. He has 11 years of teaching experience from various Engineering Colleges during tenure he was Awarded **Best Teacher Award** .He is a Member of ISTE. He has Published 1 papers in International journal and 5 Papers in International and National Level conferences. His area of interest includes Operating Systems, Data Structures, Network Security and Cryptography.

www.ijmer.com Vol.1, Issue.2, pp-352-357 ISSN: 2249-6645

Implementation of Automatic Generation Control of Hydrothermal System Employing Hybrid Genetic-Neural Approach

C.Srinivasa Rao[1]
*(EEE Department, G.Pullaiah College of Engineering and Technology, Kurnool, A.P, India)

ABSTRACT

This paper presents the design of controller based on the combined principle of Genetic Algorithm and Neural networks. The concept of artificial intelligent techniques greatly helps in overcoming the disadvantages posed by the conventional controllers. A hierarchical architecture of three layer feed forward neural network (NN) is proposed for controller design based on back propagation algorithm (BPA). Area Control Error (ACE) is considered as input to the neural network controller and the output of the controller is provided to the governor in each area. The main advantage of neural network is that it can adapt itself from the training data. In order to reduce the complexity of having more training data, Genetic Algorithm (GA) has been incorporated into the neural network in order to obtain optimal values of weights and bias. The proposed controller is tested for a two area hydrothermal system. Simulation results show that the limitations of conventional controller can be overcome by including Hybrid Genetic-Neural concept and thereby the dynamic response of the system with respect to peak time, overshoot and settling time can be improved drastically.

Keywords **- Automatic Generation Control, Genetic Algorithm, Neural Networks, Hydrothermal system**.

I. INTRODUCTION

Large scale power systems are normally composed of control areas or regions representing coherent groups of generators. In a practically interconnected power system, the generation normally comprises of a mix of thermal, hydro, nuclear and gas power generation. However, owing to their high efficiency, nuclear plants are usually kept at base load close to their maximum output with no participation in the system Automatic generation control (ACE). Gas power generation is ideal for meeting the varying load demand. Gas plants are used to meet peak demands only. Thus the natural choice for AGC falls on either thermal or hydro units.

Literature survey shows that most of earlier works in the area of AGC pertain to interconnected thermal systems and relatively lesser attention has been devoted to the AGC of interconnected hydro-thermal system involving thermal and hydro subsystem of widely different characteristics. Concordia and Kirchmayer [1] have studied the AGC of a hydro-thermal system considering non-reheat type thermal system neglecting generation rate constraints. Kothari, Kaul, Nanda [2] have investigated the AGC problem of a hydro-thermal system provided with integral type supplementary controllers. The model uses continuous mode strategy, where both system and controllers are assumed to work in the continuous mode. Perhaps Nanda, Kothari and Satsangi [3] are the first to present comprehensive analysis of AGC of an interconnected hydrothermal system in continuous-discrete mode with classical controllers. It is known that load-frequency control systems include an integral controller as secondary controller in conventional control configurations. The integrator gain is set to a level that compromise between fast transient recovery and low overshoot in dynamic response of the system. Unfortunately, this type of controller is considerably slow. Because of this, the recovery of transients in the power system against to the load perturbations spends very long time.

In recent years intelligent methods such as Fuzzy logic (FL) have been applied to the load frequency control problem [4-7]. The salient feature of these soft computing techniques are that they provide a model-free description of control systems and do not require any model identification. But the main drawbacks of ANN include large number of neurons in the hidden layers for complex function approximation, and very large training time is required. Since artificial neural network configuration will be used to control the system, back propagation algorithm is used as a learning rule to cope with the continuous time dynamics.

GA is a search and optimization method developed by mimicking the evolutionary principles and chromosomal processing in natural genetics. Especially GA is efficient to solve nonlinear multiple-extrema problems [8-9] and is usually applied to optimize controlled parameters and constrained functions. In this study, a step load change in each area is considered. For comparison, the considered power system is controlled by using both conventional integral controller and Hybrid Genetic Algorithm-Neural Network (HGANN) controller for the case mentioned above. The results obtained show that the HGANN configuration using back propagation algorithm applied for AGC of power system gives good dynamic response with respect to conventional controller.

II. DYNAMIC MATHEMATICAL MODEL

Electric power systems are complex, nonlinear dynamic system. The load frequency controller controls the control valves associated with High Pressure (HP) turbine at very small load variations [10]. The system under investigation

has tandem-compound single reheat type thermal system. Each element (Governor, turbine and power system) of the system is represented by first order transfer function at small load variations in according to the IEEE committee report [10]. Two system nonlinearities likely Governor Deadband and Generation Rate Constraint (GRC) are considered here for getting the realistic response. Governor Deadband is defined as the total magnitude of the sustained speed change within which there is no change in the valve position. It is required to avoid excessive operation of the governor. GRC is considered in real power systems because there exits a maximum limit on the rate of change in the generating power. Figure 1 shows the transfer function block diagram of a two area interconnected network .The parameters of two area model are defined in Appendix.

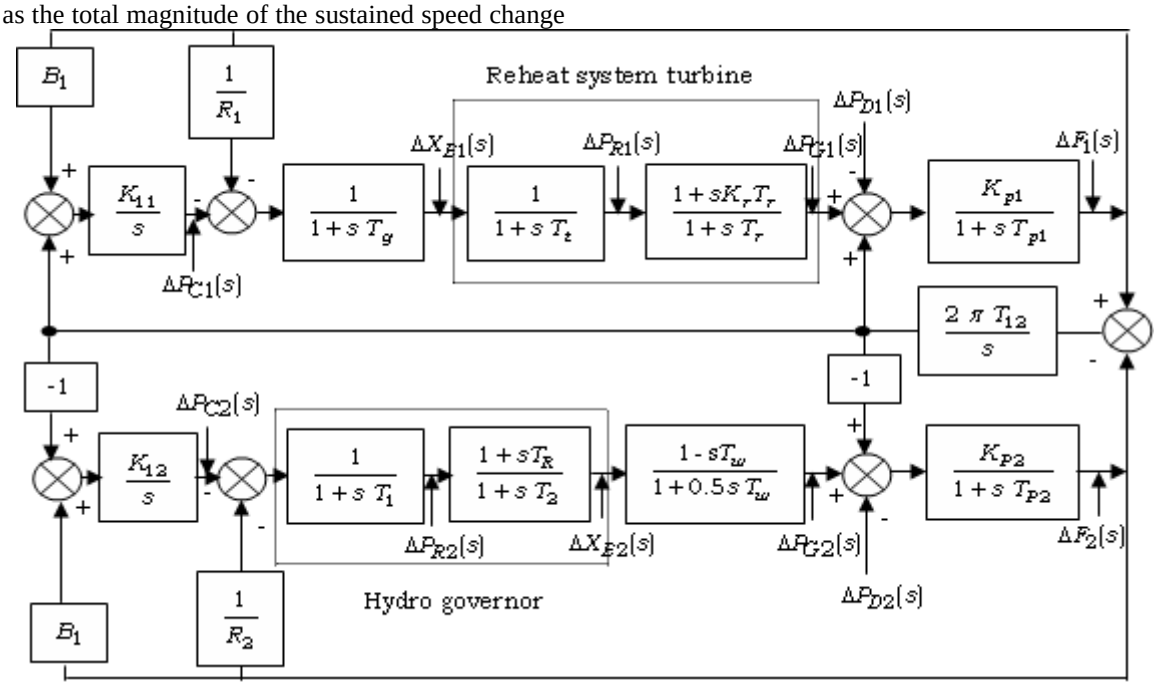


Figure. 1 Two Area Hydrothermal System

III. BACK PROPAGATION ALGORITHM

In the field of electrical engineering, one of the most exciting and potentially profitable recent developments is the increasing use of artificial intelligence techniques like neural networks in the design of various controllers. Artificial neural networks have been applied to many problems, and have demonstrated their superiority over classical methods when dealing with noisy or incomplete data. Neural networks are well suited to this method, as they have the ability to pre-process input patterns to produce simpler patterns with fewer components. A fascinating feature of the brain is that its physical organization reflects the organization of the external stimuli that are presented to it. In view of this back propagation algorithm has been used to design controller. In this back propagation algorithm the weights from input layer-hidden layer-output layer are updated iteratively during the learning phase. The updation of weights in back-propagation algorithm is done as follows: The error signal at the output of neuron j at iteration n is given by $e_j(n) = d_j(n) - y_j(n)$ (1)

The instantaneous value of error for neuron j is $\frac{1}{2}e_j^2(n)$. This instantaneous value $\varepsilon(n)$ of total error is obtained by summing $\frac{1}{2}e_j^2(n)$ of all neurons in output layer

$$\varepsilon(n) = \frac{1}{2}\sum_{j\in c} e_j^2(n) \qquad (2)$$

where c includes all neurons in the output layer. Average squared error is given by

$$\varepsilon_{avg} = \frac{1}{N}\sum_{n=1}^{N}\varepsilon(n) \qquad (3)$$

where N is total number of patterns in training set. So minimization of ε_{avg} is required. So back propagation algorithm is used to update the weights. Induced local field $v_j(n)$ produced at input of activation function is given by

$$v_j(n) = \sum_{i=0}^{m} w_{ji}(n)X_i(n) \qquad (4)$$

where m is the number of inputs applied to neuron j. So the output can be written as

$$y_j(n) = \phi_j(v_j(n)) \qquad (5)$$

The back propagation algorithm applies a correction $\Delta w_{ji}(n)$ to synaptic weights $w_{ji}(n)$ which is proportional to partial derivative $\frac{\partial\varepsilon(n)}{\partial w_{ji}(n)}$, which can be written as

$$\frac{\partial\varepsilon(n)}{\partial w_{ji}(n)} = \frac{\partial\varepsilon(n)}{\partial e_j(n)}\cdot\frac{\partial e_j(n)}{\partial y_j(n)}\cdot\frac{\partial y_j(n)}{\partial v_j(n)}\cdot\frac{\partial v_j(n)}{\partial w_{ji}(n)} \qquad (6)$$

Differentiating the equation (2) with respect to $e_j(n)$

www.ijmer.com Vol.1, Issue.2, pp-352-357 ISSN: 2249-6645

$$\frac{\partial \varepsilon(n)}{\partial e_j(n)} = e_j(n) \quad (7)$$

Differentiating equation (1) with respect to $y_j(n)$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad (8)$$

Differentiating equation (5) we get

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \phi_j'(v_j(n)) \quad (9)$$

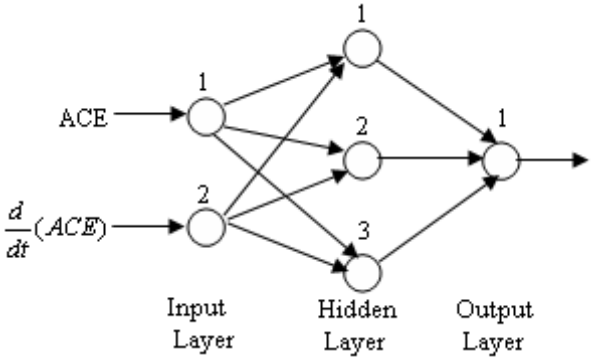


Figure. 2 Architecture of Neural Network Considered

Differentiating equation (4) with respect to $w_{ji}(n)$

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = X_i(n) \quad (10)$$

So using equations (7-10) in equation (6) we get

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = -e_j(n)\phi_j'(v_j(n))X_i(n) \quad (11)$$

The correction $\Delta w_{ji}(n)$ applied to $w_{ji}(n)$ is defined by

$$\Delta w_{ji}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} \quad (12)$$

where η is learning rate parameter. Figure 2 shows the architecture of neural network considered for this work. It can be seen that the Area Control Error (ACE) and rate of change of ACE are considered as inputs in the input layer and ΔP_c is considered as output in the output layer.

IV. GENETIC ALGORITHM

Genetic algorithms are procedures based on the principles of natural selection and natural genetics that have proved to be very efficient in searching for approximations to global optima in large and complex spaces in relatively short time. The basic components of GA are:

- Representation of problem to be solved
- Genetic operators (selection, crossover, mutation);
- Fitness function;
- Initialization procedure.

GA starts by using the initialization procedure to generate the first population. The members of the population are usually strings of symbols (chromosomes) that represent possible solutions to the problem to be solved. Each of the members of the population for the given generation is evaluated and according to its fitness value, it is assigned a probability to be selected for reproduction. Using this probability distribution, the genetic operators select some of the individuals. By applying the operators to them, new individuals are obtained. The mating operator selects two members of the population and combines their respective chromosomes to create offspring. The mutation operator selects a member of the population and changes part of the chromosome

V. HYBRID GENETIC-NEURAL NETWORK

The performance of neural network generally depends upon the values of weights and bias obtained after training. Since there is no clear methodology for determining the number of training data required for proper training of the neural network, hence the weights obtained cannot be seen as the optimum values. So in order to obtain the optimal values of weights and bias, a new hybrid technique involving both GA and NN has been proposed in this paper which uses an evolutionary technique to determine the weights instead of the steepest descent method used in traditional NN.

Normally the GA starts by randomly generating a population of strings. Each string is evaluated and the weights for hidden layer and output layer along with bias are found out as shown in Figure. 3. The above process is

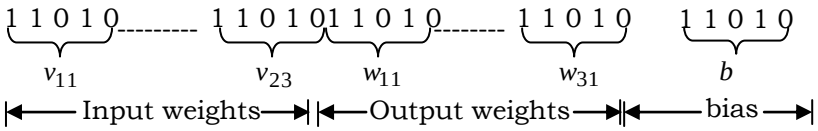


Figure. 3 Allocation of input and output weights of NN using GA

carried out for all the other strings in the population and new offspring's are created using the genetic operators like reproduction, crossover and mutation. This process is carried out for a number of iterations and the winner string is selected based on minimum value of MSE. The final values of input and output weights along with bias are calculated as shown in Fig. 3. The various weights for both input layer and output layer along with the bias are calculated for the winner string and network is built with the help of obtained weights and bias.

VI. RESULTS AND DISCUSSIONS

The proposed system in modeled in MATLAB/SIMULINK environment and the results have been presented. A load change of 0.04 p.u M.W in each area has been considered to study the comparison between HGANN network controller

www.ijmer.com Vol.1, Issue.2, pp-352-357 ISSN: 2249-6645

and integral controller. A value of 0.5 has been considered as the gain of integral controller. A performance index has been considered in this work to compare the performance of proposed methods is given by $J = \int_0^t \left(\alpha \cdot \Delta f_1^2 + \beta \cdot \Delta f_2^2 + \Delta P_{tie12}^2\right)$.

Table 1. Weights between input and hidden nodes.

	Node 1	Node 2	Node 3
Node 1	520.26	1743.7	2678.4
Node 2	509.58	462.64	-381.8

Table 2. Bias Values at hidden nodes.

Node 1	0.6004
Node 2	-0.9900
Node 3	13.648

Table 3. Weights between hidden and output nodes.

	Node 1
Node 1	-2.3068
Node 2	1.9983
Node 3	-8.977

Table 4. Bias Value at output node.

Node 1	2.125

The ISE criterion is used because it weighs large errors heavily and small errors lightly. Even though Δf_1 and Δf_2 have very close resemblance, separate weighing factors i.e., α and β are considered for each of them respectively so as to obtain better performance. The parameters α and β are weighing factors which determine the relative penalty attached to the tie-line power error and frequency error. A value of 0.65 has been considered in this work as the value for both α and β. Table 1-4 shows the values of the weights and bias obtained from the winner string of genetic algorithm.

Table 5 shows the performance of the controllers in both the areas. It can be seen that the performance of the system is greatly improved in the presence of HGANN controller rather than an integral controller.

Table 5. Comparison of performance of controllers .

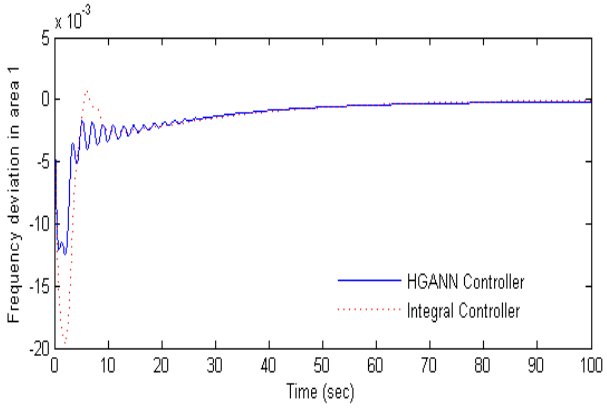
	Thermal Area			Hydro Area		
	Peak Time	Overshoot	Settling Time	Peak Time	Overshoot	Settling Time
With HGANN Controller	1.985	0.012506	21.75	1.185	0.017814	20.54
With Integral Controller	2.055	0.019635	22.745	1.565	0.023684	21.71
% Improvement	3.40	36.30	4.37	24.28	24.78	5.38

Where % improvement = $\left(\frac{\left(|With\ \ Integral\ controller| - |with\ \ HGANN\ \ controller|\right)}{|With\ \ Integral\ controller|}\right) \times 100$

Table 6 shows the comparison of performance index of the system in the presence of both controllers. It can be observed from the table that the system with HGANN controller has less performance index than that of the system with integral controller which demonstrates the superiority of the HGANN controller.

Table 6. Comparison of Performance Index Values.

	Performance Index Value
With HGANN Controller	0.0001295
With Integral Controller	0.0002114



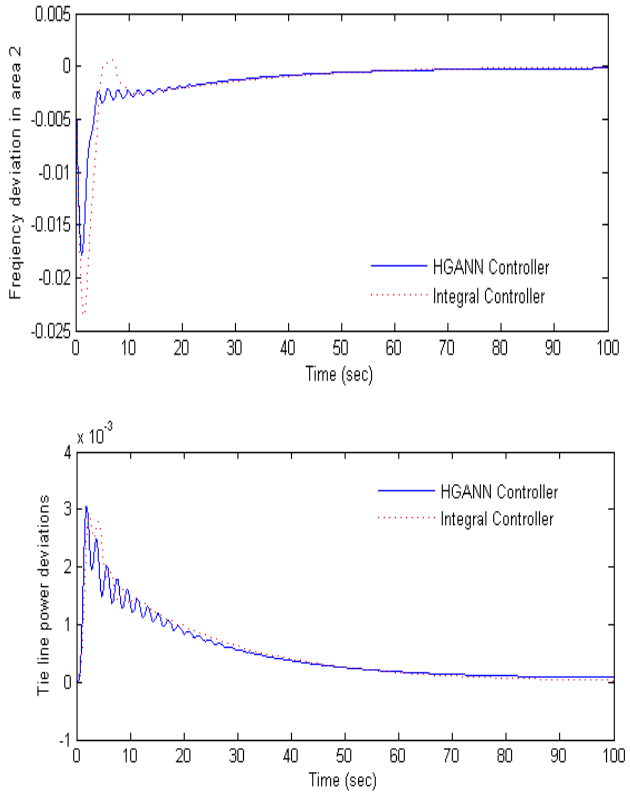


Figure. 4 Frequency and tie line power error deviations in both the areas

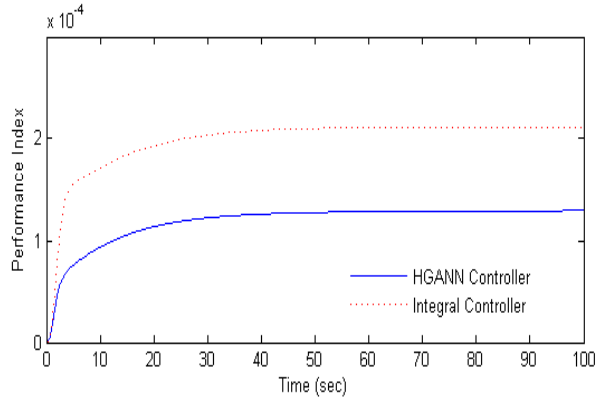


Figure. 5 Comparison of Performance index of system with both controllers

Figure 4 shows the various frequency deviations and tie line power deviations in both the areas during a load change of 0.04p.u MW. It can be observed that HGANN controller is far superior than the integral controller in terms of peak time, overshoot and settling time in both the areas. Figure 5 shows the comparison between both the controllers in terms of performance index.

VII. CONCLUSION

The performance of integral controller and HGANN controller for a two area hydrothermal system has been investigated. It has been observed that the integral is capable of bringing better dynamic response of the system to some extent. But the conventional design approach requires a deep understanding of the system, exact mathematical models and precise numerical values. The basic feature of neural concept is that the process can be controlled with slight knowledge of its underlying dynamics. But the neural network suffers from lack of optimal values of weights. In order to overcome this, an evolutionary technique like GA has been used to obtain the optimal value of weights. The simulation results show the superior performance of the system using HGANN controller.

APPENDIX

R =2.4 Hz/p.u.MW; D =8.33×10^{-3} p.u. MW/Hz; K_g =1; T_g =0.08 sec; K_t =1; T_t =0.3 sec; K_r =0.5; T_r =10 sec; T_1, T_2, T_R =41.6, 0.513, 5 sec; T_w =1 sec; K_p =120 Hz/p.u. MW ; T_p =:20 sec; B =0.425 p.u. MW/Hz

ACKNOWLEDGEMENTS

The author sincerely acknowledges the financial support provided by the management of G.Pullaiah College of Engineering and Technology: Kurnool for carrying out the present work.

REFERENCES

[1] C. Concordia and L.K.Kirchmayer, *Tie-Line Power and Frequency Control of Electric Power System* - Part II, *AIEE Transaction*, vol. 73, Part- 111-A, pp. 133-146, April 1954.

[2] M.L.Kothari, B.L.Kaul and J.Nanda, *Automatic Generation Control of Hydro-Thermal system*", *journal of Institute of Engineers(India)*, vo1.61, pt EL2, pp85-91, Oct 1980.

[3] J.Nanda, M.L.Kothari, P.S.Satsangi, *Automatic Generation Control of an Interconnected hydrothermal system in Continuous and Discrete modes considering Generation Rate Constraints*, *IEE Proc*., vol. 130, pt D, No.*1*, pp 455- 460, Jan. 1983

[4] G.A. Chown and R.C.Hartman. *Design and experience with a fuzzy logic controller for Automatic generation*

www.ijmer.com Vol.1, Issue.2, pp-352-357 ISSN: 2249-6645

control, *IEEE Transactions on power systems*, Vol. 13 ,No. 3, pp.965- 970, August 1998

[5] Jawad Talaq and Fadel Al-Basri. *Adaptive fuzzy gain scheduling for load frequency control, IEEE Transactions on power systems*, Vol. 14, pp.145-150, February 1999.

[6] T.P.Imthias Ahamad,P.S. Nagendra Rao and P.S.Sastry, *A reinforcement learning approach to automatic generation control, Electric power systems research 2002(63)*, pp.9-26.

[7] Y.L.Karnavas and D.P. Papadopoulos, *AGC for autonomous power system using combined intelligent techniques, Electric power systems research 2002(62)*,pp.225-239

[8] David Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, 1989.

[9] J. H. Holland. *Adaptation in nature and artificial systems.* Michigan, 1975.

[10] Dynamic Models for steam and Hydro Turbines in Power system studies, *IEEE committee report. Transactions in Power Apparatus & Systems*, Vol.92, No.6,pp.1904-915,Nov./Dec.1973.

www.ijmer.com Vol.1, Issue.2, pp-352-357 ISSN: 2249-6645

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-358-363 ISSN: 2249-6645

An Introduction to Graphical Processing Unit

Jayshree Ghorpade[1], Jitendra Parande[2], Rohan Kasat[3], Amit Anand[4]
[1](Department of Computer Engineering, MITCOE, Pune University) India
[2](SunGard Global Technologies, Pune) India
[3](Department of Computer Engineering, MITCOE, Pune University) India
[4](Department of Computer Engineering, MITCOE, Pune University) India

ABSTRACT

Today's world requires maximum computing speed. The progress that the CPU has achieved over the past 2 decades, though tremendous, has now reached a point of stagnation. To overcome this, a new highly parallel and multithreading processor optimized for high degree of computation was introduced, which was named as the Graphics Processing Unit (GPU) by NVIDIA or the Visual Processing Unit (VPU). A Graphics Processing Unit (GPU) is a single-chip processor primarily used to manage and boost the performance of video and graphics. This paper talks about the reasons for choosing GPU to accelerate the computation. This paper also states where GPU will work more efficiently than the CPU.

Keywords **– CPU, data parallelism, GFLOPS, GPU, SPMD**

I. INTRODUCTION

CPU frequency growth is now limited by physical matters and high power consumption. Their performance is often raised by increasing the number of cores. Present day processors may contain up to four cores (further growth will not be fast), and they are designed for common applications, they use MIMD (multiple instructions / multiple data). Each core works independently of the others, executing various instructions for various processes

The GPU is a specialized processor efficient at manipulating and displaying computer graphics. The term was defined and popularized by Nvidia as "a single chip processor with integrated transform, lighting, triangle setup/clipping, and rendering engines that is capable of processing a minimum 10 million per seconds"[1]. There are mathematically-intensive tasks, complex algorithms which would put quite a strain on the CPU. GPU lifts this burden from

the CPU and frees up cycles that can be used for other jobs. The highly parallel graphics processing unit (GPU) is rapidly gaining maturity as a powerful engine for computationally demanding applications. GPU hides latency with computation not with cache! The GPU's performance and potential offer a great deal of promise for future computing systems. One of the most important challenges for GPU computing is to connect with the mainstream fields of processor architecture and programming systems, as well as learn from the parallel computing experts of the past.

The GPU is a chip that functions on the same principle as the CPU with the one important difference that it has nothing to do with any part of the system that is not part of the graphics package on the computer. GPU is essentially a CPU that is specifically designed and dedicated to the control of graphics. The end result is an easier to control graphics package and better response time based on computer commands. Games with intensive graphics end up running a lot quicker and multimedia that you find at online websites tend to be a lot better as well. The advantages of having a GPU are therefore quite easy to notice from those outcomes and that is why people are now clamoring to have GPU devices installed into their computers.

When we compare GPUs with CPUs over the last decade in terms of Floating point operations (FLOPs), we see that GPUs appear to be far ahead of the CPUs as shown in Fig.1.

GPUs came into existence with only image and graphics computation in mind. But now GPUs has evolved into an extremely flexible and powerful processor in terms of

- Programmability
- Precision
- Performance

So GPUs are well suited for fast, efficient, non graphical computing too.

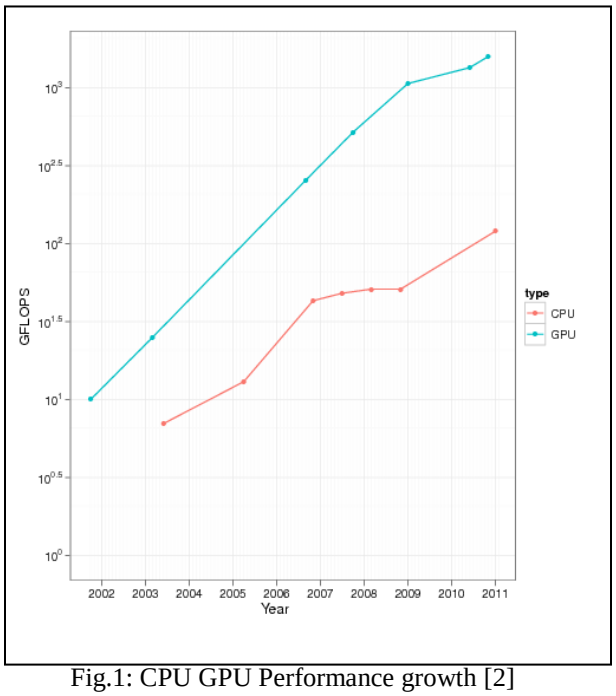


Fig.1: CPU GPU Performance growth [2]

II. CPU GPU COMPARISON

The CPU or Central Processing Unit is where all the program instructions are executed in order to derive the necessary data. The advancement in modern day CPUs have allowed it to crunch more numbers than ever before, but the advancement in software technology meant that CPUs are still trying to catch up. A Graphics Processing Unit or GPU is meant to alleviate the load of the CPU by handling all the advanced computations necessary to project the final display on the monitor.

Originally, CPUs handle all of the computations and instructions in the whole computer, thus the use of the word 'central'. But as technology progressed, it became more advantageous to take out some of the responsibilities from the CPU and have it performed by other microprocessors.

The GPU is a device that is beneficial primarily to people that has intensive graphical functions on their computer. In other words, if you just use Microsoft Office and the e-mail page of your browser when you are on the computer, chances are very good that the GPU will not add that much to your computing experience. However, if you play video games and look at videos on the internet on a frequent basis, what you will discover is that installing a GPU onto your computer will greatly improve the performance you get out of the entire thing. Improving computer performance is always a good thing and this is why the GPU has become very popular in recent years. GPU computing is on the rise and continuing to grow in popularity and that makes the future very friendly for it indeed.

In GPU computing the CPU calculations are replaced by Graphics Processing Units. Migrating large scale algorithms and entire kernel onto the GPU co-processors help in arriving at the answer much faster and thus decreases the processing time. GPUs are never a completed replacement for CPUs but complementary. Parallel operation of CPU and GPU has found to increase the performance. CPUs offload the tasks which are better performed by GPU leading to high performance computing. GPU excel CPUs in certain computational tasks.

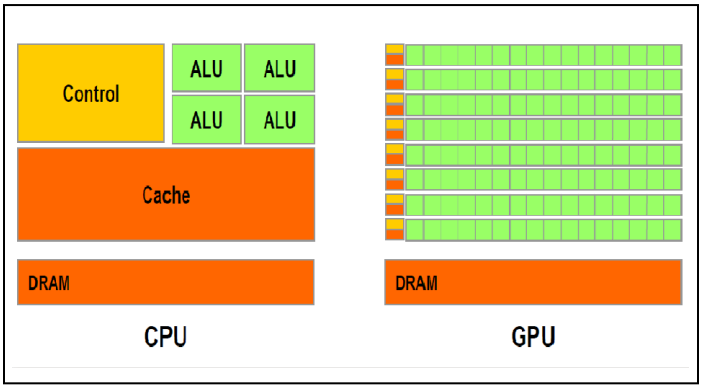


Fig.2: CPU GPU Comparison [3]

Whether it is CPU or GPU every processing unit has its own memory (cache) and shared memory (DRAM). Since it is very hard to transfer data between these structures one should avoid using complex data structures and messaging in their parallel algorithms. As it can be seen in Fig. 2, GPU has many ALU (Arithmetic Logic Unit) as compared to the CPU. So, it is able to perform multiple instructions execution at the same time. This provides the GPU with a high degree of parallelism which aids efficient computation. But it lacks the cache space. Owing to these structure differences we can deduce that an algorithm implemented for parallel structures may still work with better performance on a multi-core CPU when it could not be efficiently parallelized on GPU.

III. WORKING OF GPU

The programmable units of the GPU follow a single program multiple-data (SPMD) programming model. For efficiency, the GPU processes many elements (vertices or fragments) in parallel using the same program. Each element is independent from the other

elements, and in the base programming model, elements cannot communicate with each other. All GPU programs must be structured in this way: Many parallel elements each processed in parallel by a single program. Each element can operate on 32-bit integer or floating point data with a reasonably complete general-purpose instruction set. Elements can read data from a shared global memory and, with the newest GPUs, also write back to arbitrary locations in shared global memory.

A graphics processing unit is a dedicated graphics rendering device for a personal computer, workstation, or game console. Modern GPUs are very efficient at manipulating and displaying computer graphics. But it is also used in general purpose computation in various computation intensive algorithms. These algorithms can harness the high multiplicity of the GPU to improve their performance. Mapping the general purpose computations on the GPU is very much similar to manipulating the computer graphics. GPU computing applications are structured in the following way [4]:

1. The programmer directly defines the computation domain of interest as a structured grid of threads.
2. An SPMD general-purpose program computes the value of each thread.
3. The value for each thread is computed by a combination of math operations and both read accesses from and write accesses to global memory. Unlike in the previous two methods, the same buffer can be used for both reading and writing, allowing more flexible algorithms (for example, in-place algorithms that use less memory).
4. The resulting buffer in global memory can then be used as an input in future computation.

The GPU is organized as multiple SIMD (Single instruction, multiple data) groups. Within one SIMD group, all the processing elements execute the same instruction in synchronization. A set of threads that execute in this way is called a "warp". Branching is allowed, but if threads within a single warp follow different execution paths, there may be some performance loss.

Memory interfaces are wide and achieve highest bandwidth when that access width is fully utilized. For applications that are memory bound, this means that all threads in a warp should access adjacent data elements when possible. For example, neighboring threads in a warp should access neighboring elements in an array. This may require some rearrangement of data layout or data access patterns.

The GPU offers multiple memory spaces that can be used to exploit common data-access patterns: in addition to the global memory, there are constant memory (read-only, cached), texture memory (read-only, cached, optimized for neighboring regions of an array), and per-block shared memory (a fast memory space within each warp processor, managed explicitly by the programmer).

IV. CPU GPU WORK SHARING

For efficient use of the GPU one must find areas in the execution path of the program to send to the GPU, instead of offloading the entire code to GPU. This leads to better resource utilization. In short we must use the CPU for operations involving memory references and logical statements, while the computation intensive part of the code must be sent to the GPU.

Since the GPU is a coprocessor on a separate PCI-Express card, data must first be explicitly copied from the system memory to the memory on the GPU board.

As shown in Fig.3, the CPU has an input data stream, from where it receives the data to be processed. It has a global memory through which it references the tasks to be performed. Whenever a specific task is selected to be sent to the GPU, the CPU checks for an available unit of the GPU to which it can assign the task. On completion of the task the GPU signal the CPU and processing resumes. Since the GPU has multiple such units capable of a high degree of computation, parallelism is achieved and computation speeds up.

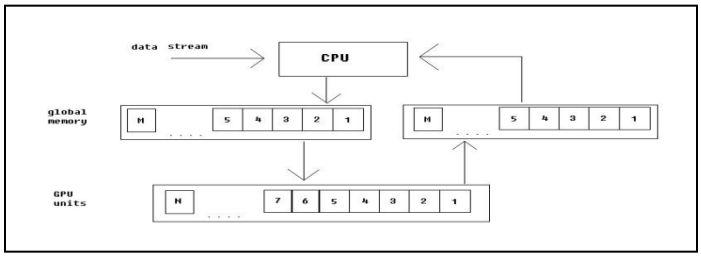


Fig.3: Data transfer between GPU and CPU

V. BENEFITS AND LIMITATIONS

Benefits:

1. Application that require large number of parallel threads work well.
2. Allows use of per-block shared memory.
3. Allows use of "data parallelism" in applications.
4. Easier to calculate reciprocal and reciprocal square root.
5. Can perform large amount of computation per data element.
6. If the synchronization is infrequent then GPU can handle them.

Limitations:

1. Applications with limited concurrency do not fully utilize the potential of the GPU.
2. Even with many threads, if all the threads are doing different work then GPU is not utilized fully.
3. Frequent global synchronization requires an expensive global barrier.
4. If there is high degree of point-to-point synchronization among random threads the GPU does not work well.
5. Frequent communication between CPU and GPU hampers the performance of GPU.
6. Applications which require small amount of computation do not work well on GPU.

VI. COMPARATIVE STUDY OF DIFFERENT GPU'S

There is always a constant and tough struggle for supremacy among the manufacturers of computer components like CPUs, graphics cards, system memory modules, coolers, etc. There is fierce competition in each price category, especially among top-end products. The graphics card market is a vivid example of that. Having the world's fastest graphics card under one's belt is not only prestigious but also profitable because it proves the manufacturer's technical superiority and promotes its sales in other price sectors.

Up to this moment the Nvidia GeForce GTX 580 has been the fastest single-GPU graphics card although AMD could offer its dual-processor Radeon HD 5970 as an alternative. On March 8, 2011, AMD released an even faster dual-GPU product, Radeon HD 6990. NVidia hasn't taken long to respond and has just rolled out its own dual-processor GeForce GTX 590.

Similarities:

They do not differ much in terms of the peak power draw: 375 watts for the Radeon HD 6990 and 365 watts for the GeForce GTX 590. AMD recommends a 750W or higher power supply with two 150W power connectors for its graphics cards. A 1200W or higher power supply is recommended for a CrossFireX tandem built out of two Radeon HD 6990s. NVidia has the following recommendations: 700 and 1000-watt power supplies for a single GeForce GTX 590 and a SLI tandem, respectively. Each card has a single connector for building multi-GPU configurations. It is located in the top front part of the PCB.

Differences:

The Radeon HD 6990 carries two full-featured Cayman GPUs. They are indeed full-featured because dual-processor cards used to be equipped with cut-down versions of GPUs in the past. The GPU frequency of the Radeon HD 6990 is only 50 MHz lower than that of the Radeon HD 6970 and equals 830 MHz. However, there is a high-speed mode you can trigger by means of the abovementioned switch near the CrossFireX connector. The card's GPU frequency is 880 MHz in that mode, but AMD says that turning that switch on will make your warranty void. The card's GPU frequency is lowered to 150 MHz in 2D applications to save power.

As opposed to AMD, NVidia equips its GPUs with such caps. The company didn't disable any subunits in the GPUs of its GeForce GTX 590, either. Each of the card's GPUs has 512 unified shader processors, 64 texture-mapping units and 48 raster operators. In other words, we've got two GeForce GTX 580 processors on a single PCB here. Their frequencies are lowered more than those of the AMD Radeon HD 6990, though. The GeForce GTX 590 clocks its GPUs at 607/1215 MHz, which is 21.4% lower than the clock rates of the GeForce GTX 580 (772/1544 MHz). The reason for this reduction is clear enough. If NVidia used the clock rates of the GTX 580 for the GTX 590, the latter's heat dissipation and power consumption would be beyond all reasonable limits. The GeForce GTX 590 drops its GPU clock rates to 51/101 MHz in 2D mode as a power-saving measure.

The GPU-Z tool is a tool that detects the CPU, RAM, motherboard chipset, and other hardware features of a modern personal computer, and presents the information in one window. It reports about the two cards as shown in fig4 and fig5.

Fig.4: ATI Radeon HD6990 Specifications [5]

Fig.5: NVidia GTX 590 Specifications [5]

The Radeon HD 6990 carries a total of 4 gigabytes of graphics memory (2 gigabytes per each GPU) whereas the GeForce GTX 590 has 1.5 gigabytes of onboard memory for each GPU or 3 gigabytes in total. As usual, AMD installs Hynix chips on its Cayman-based reference cards. These chips have a voltage of 1.5 volts and a rated frequency of 5000 MHz. The card's memory frequency is 5000 MHz, too, but is lowered to 600 MHz in 2D mode. The memory bus is 256 bits wide.

The GeForce GTX 590 comes with Samsung chips that have a rated access time of 0.4 nanoseconds and a rated frequency of 5000 MHz. However, the card clocks them at 3414 MHz only, which is 15% lower than the memory frequency of the GeForce GTX 580 and 10% lower than that of the GTX 570. The memory frequency is lowered to 270 MHz in 2D mode. The memory bus is 384 bits wide.

VII. FUTURE WORK

GPU has gained over CPUs because they are more powerful and cheaper compared to CPUs. The main area of work in future for the GPU will be to reduce the cost and provide a huge multi-processing capability to the users all around. The future of GPU rests in making it as a co-processor which performs much of the computation for the CPU. This, as of now, is difficult to do since the languages and software tools available for combining GPU process with CPU are still in their preliminary stage of development. So, a major scope for future improvement is to develop new languages and software tools specifically to take advantage of high level of parallelism.

The other thing that can be improved with engineering is GPU communication with the CPU or the NIC (Network Interface Card). At present it takes a longer and slower route to copy contents from the CPU memory to the GPU blocks and vice-versa. This uses up a lot of time of computation. But with engineering we can develop better architectures to improve the communication speed of the GPU, reducing altogether the time required for computation.

VIII. CONCLUSION

Thus we can conclude the following advantages of GPU over CPU such as

1. GPU's contain much larger number of dedicated ALU's then CPU.
2. GPU's also contain extensive support of stream processing paradigm. It is related to SIMD processing. [4]
3. Each processing unit on GPU contains local memory that improves data manipulation and reduces fetch time.

The Graphical Processing Unit is visually and visibly changing the course of general purpose computing. The future of the GPU certainly has much more promise on the horizon than the general-purpose CPU. Although the GPU will not overtake the CPU as the main processor, we do think that the GPU has much more potential for expanding our computing experience. The CPU has a large amount of logic dedicated to branch prediction, whereas stream processing does not require as much of this type of logic. The GPU is much better at parallelism than the CPU and as the gap in the transistor rate expansion continues to grow the GPU parallelism performance will also continue to grow. The power of solving these highly parallel problems has immense implications to the scientific community because we are the able to change the evolution of scientific computation from the CPU growth curve that double every 18 months to GPU growth curve that doubles about every 6 months.

REFERENCES

[1] *Graphics Processing Unit – Wikipedia* (http://en.wikipedia.org/wiki/graphics_processing_unit)

[2] *CPU GPU Trends over time*. (http://csgillespie.wordpress.com/2011/01/25/cpu-and-gpu-trends-over-time/)

[3] *GPGPU: OPENCL vs CUDA vs ArBB* (http://www.keremcaliskan.com/gpgpu-opencl-vs-cuda-vs-arbb/)

[4] Peter Zalutski – "*CUDA–Supercomputing for masses*"

[5] Sergey Lepilov –"*Equilibrium: AMD Radeon HD 6990 vs. NVidia GeForce GTX 590*."

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue.2, pp-364-370 ISSN: 2249-6645

A Novel Approach for Solving the Power Flow Equations

Anwesh Chowdary[1], Dr. G. MadhusudhanaRao[2]

[1]Dept.of.EEE, KL University-Guntur, Andhra Pradesh, INDIA

[2]Dept.of.EEE, JJ Group of Institutions-Hyd, Andhra Pradesh, INDIA

ABSTRACT

This paper presents a detailed investigation into the effectiveness of iterative methods in solving the linear system problem in power flow solution process. Previously Newton method employing an LU method, GMRES method has been one of the most widely used power flow solution algorithms A fast Newton-FGMRES method for power flow calculations is proposed in this paper. Three accelerating schemes to speed up the Newton-FGMRES method are proposed.. The simulation result gives the effectiveness of proposing one compared with existing methods.

Index Terms: **Flexible GMRES method, iterative methods, Newton power flow calculation.**

I. INTRODUCTION

A majority of computational effort in the Newton power flow method lies in solving a set of linear equations. The traditional direct LU factorization method has been popular in solving a set of linear equations. In the last 20 years or so, the iterative methods emerged as a vital alternative to the traditional Newton-LU method due to its speed. However, direct methods find the exact solution after a finite number of steps. Iterative methods, on the other hand, successively approximate the solution to a predetermined degree of accuracy based on an initial guess.

For a set of very large linear equations, the use of direct methods is impractical; it simply takes too long. Experience in solving VLSI circuit design problems has confirmed the impracticality of LU factorization for large circuit design problems. The cost of using direct methods to solve a system of linear equations is of the order for dense matrices and to for sparse matrices. Stationary iterative methods bring the cost down to the order of for dense matrices and for sparse matrices.

Non stationary iterative methods, such as Krylov subspace methods [1], [2], converge in at most iterations (assuming no round-off error), where the system size, and preconditioning often significantly is reduces the required number of iterations. The benefits of using iterative methods over direct methods increase with system size. While both iterative methods and direct methods are applicable to solve small systems of linear equations, it is often hard to solve very large linear equations without using an iterative method. Direct methods take longer computation time for large-scale systems and this difficulty can be greatly improved by the use of iterative methods. It is important to note that the distinction between direct and iterative methods is becoming more blurred, as many preconditioning techniques result in methods that are a combination of both iterative and direct solvers. Nevertheless, there is still much to be learned from both methods.

The advantages of iterative solvers over direct methods based on the direct LU factorization method in power system applications have been demonstrated in [3]–[11]. It is now recognized by many researchers that the Newton-GMRES (generalized minimal residual) method can outperform the Newton-LU method when solving large-scale power flow equations. A significant amount of speed-up, for instance 50%, obtained by the Newton-GMRES method over the Newton-LU method has been achieved. Nonstationary/Krylov subspace methods have become more complicated because the operations performed at each step involve iteration dependent coefficients. The oldest Krylov subspace method is the Conjugate Gradient (CG) method for symmetric positive definite (SPD) matrices. Since the discovery of the oldest Krylov subspace method, the Conjugate Gradient method, much work has been done to find similar methods that can be applied to nonsymmetric and/or nondefinite matrices. Some of these newer, more general methods include the GMRES method, the Biconjugate Gradient Stabilized method, and the Quasi-Minimal Residual method. These methods are all termed Krylov subspace methods because they are derived with respect to a Krylov basis.

For iterative Krylov subspace methods, it has been found that preconditioning plays an important

www.ijmer.com Vol.1, Issue.2, pp-364-370 ISSN: 2249-6645

role in the convergence rate of iterative solvers. Several preconditioners developed for power system computations appeared in [5], [7], and [12]. However, these preconditioners for "normalizing" linearized power mismatch equations were fixed at each Newton iteration. Recently, an adaptive preconditioner was proposed for the Jacobian-free Newton-GMRES(m) method in [13]. The proposed preconditioners were updated using a rank-one update algorithm. However the updated preconditioners were only used for the linearized equations of next iterations. The preconditioners were still kept constant while solving the linear equations.

To further improve the iterative methods, a flexible inner outer Krylov subspace method (FGMRES, flexible inner-outer preconditioned GMRES) was developed [14], [15]. Different from the traditional iterative Krylov subspace methods, the preconditioners used in this FGMRES method were allowed to vary with in each iteration. Thus, the FGMRES method has been observed to be more effective than the traditional GMRES in several numerical studies [14].

In this paper, the FGMRES method is applied to solve linear equations arising from the Newton power flow method. To further improve the speed of this Newton-FGMRES method, three accelerating schemes are developed and incorporated into the proposed Newton-FGMRES. This paper compares the convergence characteristics and computational speed of the Newton-FGMRES and fast Newton-FGMRES with the traditional Newton-GMRES on two practical power systems: a 12 000-bus system and a 21 000-bus system. Numerical studies show the advantages of the proposed fast Newton-FGMRES in computational speed and in robustness under different loading conditions. We point out that the traditional direct method (Newton-LU) was used as a benchmark method for both the traditional Newton-GMRES method and the fast Newton-FGMRES method. We have also evaluated the fast decoupled Newton method on the two large-scale power systems. However, the fast decoupled Newton method diverges on both test systems.

2.1. PROBLEM FORMULATION

It is assumed that all control devices remain fixed throughout the Newton solution process. Hence, voltage regulating generators will be considered as PV buses with unlimited reactive capabilities. The power flow Jacobian will be formulated in polar coordinates, as follows A single iteration of the Newton process involves solving equation (1) for the state update **(A@,+)** and then updating the state vector *(0, V).* Traditionally, the linear system (1) is solved via an LW factorization of the Jacobian, a forward elimination and a backward substitution. When solving the power flow equations, the Jacobian is relatively inexpensive to evaluate, since evaluation of the bus power mismatches involves similar calculations. Likewise, the forward elimination and backward substitution procedures are fairly inexpensive due to efficient sparse storage of the matrix factors *L* and *U.* Based on the UNIX run-time profiler output, the most time-consuming procedure of a single Newton iteration is the LU factorization of the Jacobian matrix. For large-scale power systems (e.g., the 3493 bus case studied here), the cost of an LU factorization of the system Jacobian dominates the costs of the other operations, consuming approximately **85%** of the total Newton process execution time.

2.2. NEWTON METHODS:

Nonlinear algebraic systems of equations are usually solved by a Newton method due to the local quadratic convergence. While this local contraction property is desirable, it is often the case that the last step of the Newton method decreases the residual of the nonlinear system well beyond the user specified tolerance. This "over solving" cannot be avoided in an exact Newton method when the linear system is solved directly via an LU factorization. However, an inexact Newton method, such **as** Newton- GMRES, monitors the level of accuracy in the solution by keeping track of the norm of the residual. Hence, an inexact Newton method based on an iterative linear solver can be stopped during the solution of the linear system, if the solution to the linear system has been computed accurately enough. By avoiding the waste of computation spent on over solving, an inexact Newton approach can be a serious competitor to a exact Newton method.

3. INEXACT NEWTON METHODS

An alternative to the direct solution (via LU factorization) of the linear system (1) is an iterative approach. Non stationary iterative methods for the solution of linear equations have received great attention recently from researchers in the field of numerical analysis. A promising technique in the category of Krylov subspace approaches is the Generalized Minimal Residual (GMRES [12]) method, which attempts to solve the linear system

$$Ax = b \qquad (2)$$

by minimizing the residual ***r*** defined by

$$r(x) \stackrel{\text{def}}{=} b - Ax \qquad (3)$$

via Krylov subspace updates to the candidate linear system solution z. GMRES is a member of the family of Krylov subspace iterative methods, which

produces a sequence ***xk*** of approximations to the solution ***2*** = ***A-lb*** of linear system (2). In general, the Krylov subspace iterates are described by

$$x_k \in x_0 + \mathcal{K}_k(r_0, A), \qquad k = 1, 2, \ldots \qquad (4)$$

where z_o is the initial estimate of the solution to (2)

$$\mathcal{K}_k(r_0, A) = \text{span}(r_0, Ar_0, \ldots, A^k r_0). \qquad (5)$$

In particular, GMRES creates a sequence ***2k*** that minimizes the norm of the residual at step ***k*** over the lc^{th} Krylov subspace as follows

$$\|b - Ax_k\|_2 = \min_{x \in x_0 + \mathcal{K}_k(r_0, A)} \|b - Ax\|_2. \qquad (6)$$

At step *k* , GMRES applies the Arnoldi process to a set of *k* ortho normal basis vectors for the k^{th} Krylov subspace to generate the next basis vector. When the norm Fig **1:** The GMRES(m) algorithm (for ***A*** E **PX")** without preconditioning of the newly created basis vector is sufficiently small, GMRES solves the following ***(k*** + **1)** x ***IC*** least squares problem

$$\|g_k - H_k y_k\|_2 = \min_{y \in C^k} \|g_k - H_k y\|_2,$$

where ***Hk*** is a ***(k*** + **1)** x ***k*** upper Hessenberg matrix of full rank ***k*** and ***gk*** = ***Ilrollel*** with standard basis vector **el** Ei ***Rk+l.*** 'Io solve the least squares problem, a Modified Gram- Schmidt procedure is generally used. We have described a restarted GMRES algorithm, following **[7],** in Figure 1. As mentioned in [7], a forward difference approximation can be used to compute the directional derivatives used by GMRES. Since the Jacobian matrix is only used by GMRES in matrix vector multiplications, it is possible to avoid the cost of creating the Jacobian matrix. However, the forward difference approximations to the directional derivatives involve evaluating the nonlinear power flow mismatch function at every GMRES iteration. However problems in large scale power systems present research area in terms of applicability still cannot compete with direct methods because of possible convergence problem.

Figure 1: Standard GMRES algorithm with right preconditioning

1. $r_0 = b - Ax_0$, $k = 0$, $\rho = \|r\|_2$, $v_1 = r_0/\rho$
   $errtol = \max(abstol, reltol\|b\|_2)$
   $g = \rho(1, 0, 0, \ldots) \in R^{m+1}$
2. While $\rho > errtol$ and $k < m$ do
   (a) $k = k + 1$
   (b) $v_{k+1} = Av_k$
   (c) for $j = 1, \ldots, k$
      i. $h_{j,k} = v_{k+1}^T v_j$
      ii. $v_{k+1} = v_{k+1} - h_{j,k} v_j$
   (d) $h_{k+1,k} = \|v_{k+1}\|_2$
   (e) $v_{k+1} = v_{k+1}/h_{k+1,k}$
   (f) Apply and create Givens rotations
      i. If $k > 1$ apply Q_{k-1} to the kth column of H
      ii. $\nu = \sqrt{h_{k,k}^2 + h_{k+1,k}^2}$
      iii. $c_k = h_{k,k}/\nu$; $s_k = h_{k+1,k}/\nu$
      $h_{k,k} = c_k h_{k,k} - s_k h_{k+1,k}$; $h_{k+1,k} = 0$
      iv. $g = G_k(c_k, s_k)g$
   (g) $\rho = |(g)_{k+1}|$
3. Set $r_{i,j} = h_{i,j}$ for $1 \le i, j \le k$
   Set $(w)_i = (g)_i$ for $1 \le i \le k$
   Solve the upper triangular system $Ry_k = w$
4. $x = x_0 + V_k y_k$

4. FAST NEWTON-FGMRES

The proposed fast Newton-FGMRES method is composed of the 1) Newton method, 2) FGMRES method for solving the linear equations, and 3) the three accelerating schemes including a hybrid scheme, a partial preconditioner update scheme, and an adaptive tolerance control scheme. The hybrid scheme generates the preconditioners for the inner iterations of the FGMRES method based on the complete LU factorization of the coefficient matrices. Of course, the complete LU factors can also be used to solve the corresponding linear equations.

When the dimension of the coefficient matrix changes, the preconditioner can be fast updated from the previous one by using the partial preconditioner update scheme. Using the adaptive tolerance control scheme, the stopping criterion used by FGMRES is based on the residuals from the previous Newton iterations. We are now in a position to present the fast Newton-FGMRES power flow method.

Step 1) Input the data of the power system to be studied.

Step 2) Initialization

i. Set the initial value for bus voltage.

ii. Construct the admittance matrix.

www.ijmer.com Vol.1, Issue.2, pp-364-370 ISSN: 2249-6645

iii. Save the initial bus state information (PV bus or PQ bus) as flag
iv. Define a threshold value for preconditioner updates.
Step 3) Construct the Jacobian matrix J and evaluate the power mismatch ds
Step 4) Solve the power mismatch equation $J\Delta x = dS$. If L and U have not been formed, then
i. Set flag0=flag
ii. Factorize J and save the factors L,U: .
iii. Solve $J\Delta x = dS$by a forward elimination and a backward substitution using L and U.
iv. Go to Step 5.
Else
i. If, Flag0≠Flag update the preconditioner by the partial preconditioner update scheme.
ii. Solve $J\Delta x = dS$by the FGMRES method.
iii. If FGMRES converges to the tolerance of , go to Step 5. Otherwise, clear L and U do this step again.
Step 5) Update the bus voltage value.
$x = x + \Delta x$
Step 6) Check the reactive generation constraints and save the current bus state information as Flag.
Step 7) Decide whether a new preconditioner is needed.
i. Compare Flag with Flag0 and evaluate the number of buses whose states have been considerably changed m, .
ii. If,m≥m0then a new preconditioner is required. Clear L and U. Otherwise, keep L
and U.
Step 8) Stopping criterion: If , $\|\Delta x\|_2 \geq \varepsilon$then go to step 3; otherwise, power flow calculation stops.

5. NUMERICAL RESULTS

The proposed Newton-FGMRES and fast Newton-FGMRES methods are evaluated on the following two practical power systems in North America: a 12 000-bus system and a 21 000-bus system. The traditional Newton-LU method, the Newton-GMRES method, the proposed Newton-FGMRES method, and the proposed fast Newton-FGMRES method are compared in terms of convergence characteristics and computation time. The initial guess for the iterative solver was selected to be a flat start. The convergence criterion was set to 10^{-5} and the maximum iteration number for GMRES and FGMRES was set to be 10. ILU- preconditioners were used in the Newton-GMRES method and the Newton-FGMRES method. Different parameters for ILU- were also considered: k=15in (a) and k=25 in (b). We also use the approximate minimum degree ordering as the sparse ordering scheme. We first considered the cases with unlimited Q-generation and then

TABLE I
COMPUTATION TIME FOR NEWTON POWER FLOW CALCULATION IN SECONDS FOR A 12 000 BUS SYSTEM BY DIFFERENT METHODS

	Newton-GMRES	Newton-FGMRES		Fast Newton-FGMRES	
		Time	Improvement	Time	Improvement
(a)	0.186	0.170	+8.6%	0.111	+40.3%
(b)	0.152	0.141	+7.2%		+27.0%

TABLE II
COMPUTATION TIME FOR NEWTON POWER FLOW CALCULATION IN SECONDS FOR A 21 000 BUS SYSTEM BY DIFFERENT METHODS

	Newton-GMRES	Newton-FGMRES		Fast Newton-FGMRES	
		Time	Improvement	Time	Improvement
(a)	0.518	0.489	+5.6%	0.422	+18.5%
(b)	0.472	0.461	+2.3%		+10.6%

considered the cases with limited Q-generation. In these numerical studies, the control actions of ULTC and phase-shifters are neglected. The computer used for the tests is described as follows: CPU: 1.83 GHz, the number of cores in CPU: 1, main memory: 1G, programming language: Fortran. All computation times shown in the tables are the average time. Note that we are only concerned with the computation time of the iterative process. For the fast Newton-FGMRES method, the computation times listed in the tables correspond to those required from Step 3 to Step 8 of the flowchart of the method.

We have observed that both the Newton-FGMRES method and the fast Newton-FGMRES method converge faster than the Newton-GMRES method on the 12 000-bus system and the 21 000-bus system, respectively. The total computation time required by the three methods is summarized in Tables I and II. The proposed fast Newton-FGMRES method is generally faster than the Newton-FGMRES method and the Newton GMRES method. The difference in required computation time can be significant. For the 12 000-bus system, the fast Newton-FGMRES method can be 40.3% faster than the traditional Newton-GMRES method. For the 21 000-bus system, the fast Newton-FGMRES method can be 18.5% faster than the traditional Newton-GMRES method. This reveals that the three accelerating schemes are effective in improving the performance of the Newton-FGMRES. Figs. 1 and 2 shows the convergence characteristics of the three methods on the 12 000-bus system and the 21 000-bus system, respectively. It can be observed that by

the use of FGMRES, the Newton iteration can be reduced by at least two times the iterations using GMRES. Therefore both the Newton-FGMRES method and the fast Newton-FGMRES method converge faster than the traditional Newton-GMRES method. We next considered the cases with limited Q-generation. The upper and lower limits of the reactive power generation were checked during the solution process. The proposed fast Newton-FGMRES method was compared with the traditional Newton- GMRES method in terms of computation time. The test results are summarized in Table III The traditional direct method (Newton-LU) was used as a benchmark method for both the traditional Newton-GMRES method and the fast Newton FGMRES method. The fast refactorization method used in this Newton-LU method is described as follows. The factorization is divided into two steps: the symbol decomposition and the numerical decomposition. In the symbol decomposition, the positions of nonzero fill-ins are identified

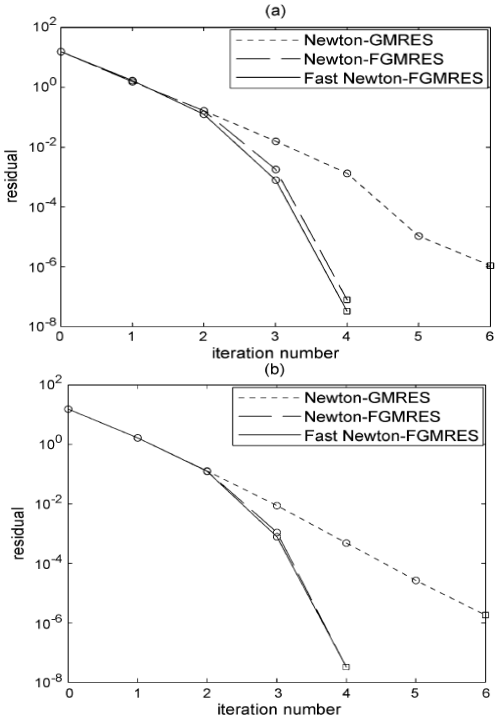


Fig. 2. Convergence characteristics of different methods on the 12 000-bus system.

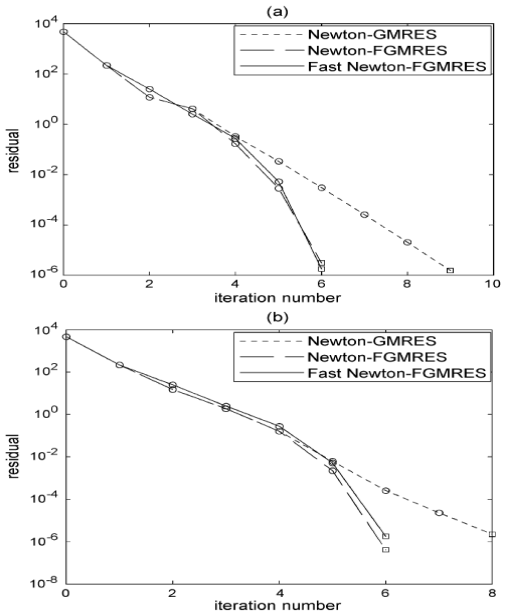


Fig 3. Convergence characteristics of different methods on the 21 000-bus

Our numerical studies reveal that the Newton-GMRES method has no superiority in speed over the traditional Newton-LU method on these two large-scale power systems. This can be explained from the viewpoint of the precondition: the dimension of the Jacobian changes during the iterative process and the repeating construction of preconditioners damages the advantage of the traditional Newton-GMRES method. However, the proposed fast Newton-FGMRES method can still be faster than the Newton-LU power flow method by 16.6% on the 12 000-bus system and 26.2% on the 21 000-bus system. In light of numerical evaluations of these two large-scale power systems, it may be concluded that the proposed fast Newton-GMRES method is considerably faster than the traditional Newton-LU method system.

TABLE III
COMPUTATION TIME OF NEWTON POWER FLOW IN SECONDS BY DIFFERENT SOLVERS

Bus Number	Newton -LU	Newton-GMRES		Fast Newton-FGMRES	
		Time	Improvement	Time	Improvement
12000	0.307	0.322	-4.9%	0.256	+16.6%
21000	0.836	0.844	-1.0%	0.617	+26.2%

Different Loading Conditions:

TABLE IV
COMPUTATION TIME OF THE NEWTON POWER FLOW IN SECONDS AT DIFFERENT LOADING CONDITIONS FOR A 21 000 BUS SYSTEM BY THE NEWTON LU AND THE FAST NEWTON-FGMRES

Loading Condition	Newton-LU	Fast Newton-FGMRES	
		Time(s)	Improvement
0.5	0.642	0.436	+32.0%
0.6	0.642	0.418	+34.9%
0.7	0.642	0.418	+34.8%
0.8	0.642	0.418	+34.9%
0.9	0.643	0.418	+34.9%
1.0	0.777	0.422	+45.7%
1.1	0.771	0.448	+41.9%
1.2	diverge	diverge	----

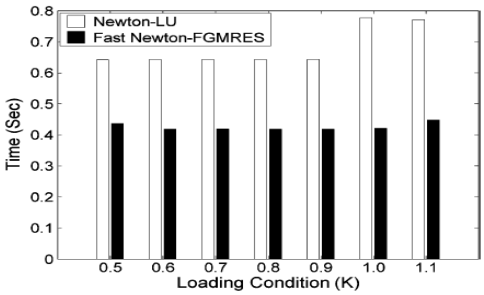


Fig4.Comparison of the computation time required by the Newton-LU and the fast Newton-FGMRES under different loading conditions on the 21 000 bus system.

TABLE V
MAXIMUM AND MINIMUM COMPUTATION TIME REQUIRED BY THE NEWTON-LU AND FAST NEWTON-FGMRES ON THE 21 000-BUS SYSTEM

	Newton-LU		Fast Newton-FGMRES	
	K	Time(Sec)	K	Time(Sec)
Maximum Time	1	0.777	1.1	0.448
Minimum Time	0.5	0.642	0.8	0.418
Difference	21.0%		7.2%	

VI. CONCLUSIONS

In this paper, we have proposed a Newton-FGMRES method for solving power flow equations. From a computational viewpoint, Newton-FGMRES is a slight extension of the existing Newton-GMRES method. However, we have explored the numerical characteristics of power flow equations and developed three accelerating schemes including a hybrid scheme, The proposed fast Newton-FGMRES solver has been evaluated on two practical large-scale power systems, one with 12 000 buses and another with 21 000 buses. Numerical results show the advantages of the proposed fast Newton-FGMRES method as opposed to the traditional Newton-GMRES method in terms of the convergence characteristics and the computation time. Even though the Newton-GMRES method has no superiority in speed over the traditional Newton-LU method on these two large-scale power flow equations, the proposed fast method consistently outperforms both the traditional Newton-LU and Newton-GMRES in terms of computational speed.

REFERENCES

[1] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. Philadelphia, PA: SIAM, 1994.

[2] Y. Saad and M. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM J. Sci. Stat. Comput.*, vol. 7, no. 3, pp. 856–869, 1986.

[3] F. D. Galiana, H. Javidi, and S. McFee, "On the application of a preconditioned conjugate gradient algorithm to power network analysis," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 629–636, May 1994.

[4] H. Mori, H. Tanaka, and J. Kanno, "A preconditioned fast decoupled power flow method for contingency screening," *IEEE Trans. Power Syst.*, vol. 11, no. 1, pp. 357–363, Feb. 1996.

[5] A. Semlyen, "Fundamental concepts of a Krylov subspace power flow methodology," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1528–1534, Aug. 1996.

[6] M. A. Pai and H. Dag, "Iterative solver techniques in large scale power system computation," in *Proc. 36th IEEE Conf. Decision and Control*, San Diego, CA, 1997, pp. 3861–3866.

[7] A. J. Flueck and H. D. Chiang, "Solving the nonlinear power flowequations with an inexact Newton method using GMRES," *IEEE Trans. Power Syst.*, vol. 13, no. 2, pp. 267–273, May 1998.

[8] D. Chaniotis and M. A. Pai, "Iterative solver techniques in the dynamic simulation of power systems," in *Proc. IEEE Power Eng. Soc. Summer Meeting*, 2000, pp. 609–613.

[9] M. A. Pai, P. W. Sauer, and A. Y. Kulkarni, "A preconditioned iterative solver for dynamic simulation of power systems," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1995, pp. 1279–1282.
[10] H. Dag and A. Semlyen, "A new preconditioned conjugate gradient power flow," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1248–1255, Nov. 2003.
[11] F. de Leon and A. Semlyen, "Iterative solvers in the newton power flow problem: Preconditioners, inexact solutions and partial Jacobian updates," *Proc. Inst. Elect. Eng., Gen., Transm., Distrib.*, vol. 149, no. 4, pp. 479–484, Jul. 2002.
[12] A. B. Alves, E. N. Asada, and A. Monticelli, "Critical evaluation of direct and iterative methods for solving systems in power flow calculations and contingency analysis," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 702–708, Nov. 1999.
[13] Y. Chen and C. Shen, "A Jacobian-free Newton-GMRES(m) method with adaptive preconditioner and its application for power flow calculations," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1096–1103, Aug. 2006.
[14] Y. Saad, "A flexible inner-outer preconditioned GMRES algorithm," *SIAM J. Sci. Stat. Comput.*, vol. 14, no. 2, pp. 461–469, 1993.
[15] V. Simoncini and D. B. Szyld, "Flexible inner-outer Krylov subspace methods," *SIAM J. Numer. Anal.*, vol. 40, no. 6, pp. 2219–2239, 2003.

AUTHOR'S BIOGRAPHY

[1]**Dr.G.Madhusudhana Rao**, Professor in JJ Group of Institutions, Hyderabad. Ph.D from JNT University Hyderabad and completed M.Tech from JNT University-Hyderabad. He has Published 12 research papers in International Journals and more than 15 International conference papers and more than 10 national conference papers.

[2]**Anwesh chowdary,** Asst.Professor in K L University Guntur. He completed his M.Tech from K L University-Guntur. B.Tech from JNTU Hyderabad. He has Published 2 research paper in International Journals and 1 International conference paper and 1 national conference paper.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-371-384 ISSN: 2249-6645

THE EFFECT OF INTERNAL ACOUSTIC EXCITATION ON THE AERODYNAMIC CHARACTERISTICS OF AIRFOIL AT HIGH ANGLE OF ATTACKE

Dr. Mohammed W. Khadim
Mechanical Engineering Dept.
Iraq Karbalaa University

Abstract:

The effect of internal acoustic excitation on the aerodynamic characteristics of NACA 23015 airfoil have been investigated experimentally and numerically (computational and ready package Fluent (6.1)) in the present work, as a function of excitation frequency and excitation location from the leading edge. The solution of the flow equations are presented for an angle of attack range (14, 16, 18, 20 and 24) degrees, at excitation frequency values (100, 150 and 200) Hz, with the two-excitation location from the leading edge (6.5% and 11.5%) of chord, at Reynolds number based on chord of $3.4x10^5$. The experimental tests are separately conducted in two suction, open-typed wind tunnels at the Reynolds number based on chord of $3.4x10^5$ for the measurements and $1x10^4$ for the visualization. The results indicate that the excitation frequency and location are the key parameters for controlling the separated flow, and the acoustic excitation technique is able to alter the flow properties and thus to improve the aerodynamic performance. The most effective excitation frequency is found to be equal 150 Hz, which leads to increase the lift coefficient at 45% at the excitation location 6.5% chord and 35% at the excitation location of 11.5% chord, especially at the poststall region of angle of attack (16-20) degree, with 10% increasing for the Lift/Drag coefficient.

Nomenclatures

Character	Description	Units
A	Area	m^2
$a_{e,w,n,s}$	Coefficients in Discretized Equations	
C	Chord Length	M
C	Speed of Sound	m/s
Cp	Pressure Coefficient	
Ca	Non-Dimensional Coefficient for the Axial Force	
$C_\mu, C_{\varepsilon 1}, C_{\varepsilon 2}, \sigma_k, \sigma_\varepsilon$	$k-\varepsilon$ Constants in model	1/s
E	Sound Energy Density	$W.s/m^3$
F	Sound Frequency	Hz
g_{ij}	Metric Tensor Element	
G1,G2,G3	Contraveriant Velocity Components	
h1,h2	Geometric Quantities	
I	Sound Intensity	W/m^2
K	Turbulent Kinetic Energy	m^2/s^2
P	Pressure	N/m^2
Pac	Sound Power	Watt
P_k	Production Rate of Turbulent Kinetic Energy	
S	Source Term	
St	$f\ c/U_\infty$ Strouhal number =	
S_ϕ	ϕ Linearized Source Term for	
t_{ij}	Viscous Stress Tensor	
U,V	Mean Velocity Components	m/s
U_∞	Free-Stream Velocity	m/s
u,v	Cartesian Velocity Components	m/s
$\vec{u}$	Velocity Vector	m/s
u_i	Velocity in Tensor Notation	m/s
u_ξ, u_η	Covariant Velocity Components	m/s
x, y	Cartesian Coordinate	
xu, yu	Upper Surface Camber Line Coordinates in X-Y Axis	
xl, yl	Lower Surface Camber Line Coordinates in X-Y Axis	
Z	Acoustic Impedance	$N.s/m^3$
Subscripts		
e, w, n, s	Faces of the control volume	
∞	Free stream condition	
ϕ	Dependent variable	
I	Covariant components i=1, 2, 3, ...	
ξ, η	Derivative with respect to curvilinear components	
ε	ε equation Refers to the source term of	
k	k equation Refers to the source term of	

Introduction:

During the course of combat, take- off and landing of an aircraft, a wing- stall phenomenon may occur when the aircraft is flying at a high angle of attack (AOA). This is due primarily to the occurrence of flow separation over a large portion of the wing surface. As a result, the aerodynamic lift is lost and the drag is increased dramatically. In order to keep the flow attached to the surface, the boundary layer must have sufficient energy to overcome the adverse pressure gradients when it occurred. Thus, the basic idea of the flow control herewith is to energize the boundary layer and therefore suppress the flow separation.

The goal is to be able to control the separated flow over an airfoil. For that matter, are should clarify the difference between separated flow control (strong) and flow separation control (weak). Both imply the condition of a detached flow, but in a different flow scale. **[1]** makes a clear distinction between these two flow fields as shown in figure (1):-

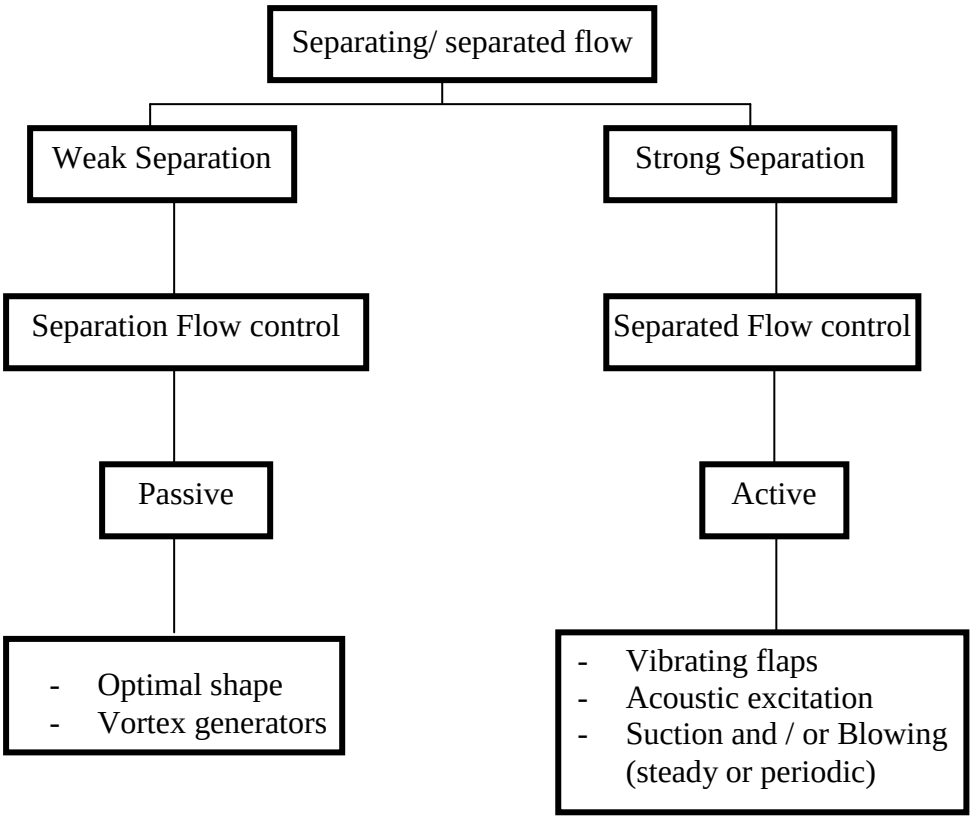


Fig. (1) Flow Field Classification

Summarized below are some ideas and experiments proven to be effective on suppressing the massive separation **[2]**.

- Airfoil Performance: flaps +Slats,
- Standford airfoils,
- Wall transpiration/ suction,
- Momentum injection,
- Moving Walls,
- Turbulators, and
- Wall heating/ cooling.

It was found that sound at particular frequencies and intensities could change the transition process of boundary layer **[3]**. The flow field exhibits different characteristics, and the momentum exchange is enhanced due to the introduction of the acoustic waves. In accordance with this observation, the control of flow with vertical structures using acoustic excitation techniques has been studied extensively in recent years.

One of the techniques is called the external acoustic excitation, in which the sound is radiated onto the wall from a source outside the flow system. This technique has been applied by several

researchers, first observed that the application of the external acoustic excitation could change the lift on the airfoil. Further studies of the interaction between the external acoustics and the separated flow were carried out, they pointed out that the entrainment was enhanced by the sound-induced velocity rather than the sound pressure.

However, it was found that the external excitation became effective only when the excitation frequency was close to the tunnel resonance frequency. Furthermore, the external excitation requires high sound pressure levels in order to achieve satisfactory results **[4]**. Hence, the external acoustic excitation appears impractical for actual applications.

To overcome those drawbacks, an internal excitation technique is used, in which the sound is emanated from a narrow opening on the wall surface. Some researchers mentioned the use of sound emission from small holes on the suction surface, but no quantitative results were presented.

Peterka, J.A., and Peter, P.D. **[5]** using a circular cylinder in cross – flow with a transverse standing demonstrated the first practical application of the external acoustic excitation for the boundary layer control, sound field imposed simultaneously. In the presence of sound field having its frequency matched sufficiently closely to that occurring naturally in the shear layer, the growth of the instability in the shear layer is enhanced, and the heat transfer from a body under separated flow can be increased.

Ahuja and Jones 1983 studied the effects of the external acoustic excitation on the turbulent boundary layer characteristics over an airfoil as a function of excitation frequency and level and flow velocity. The experiments successfully demonstrated that separation of turbulent boundary layer flows can be controlled by sound in both pre- and post- stall regions.

Zaman, et.al, **[4]** carried out wind- tunnel measurements of lift, drag and wake velocity spectra under external acoustic excitation for a smooth symmetrical (airfoil) at Reynolds numbers based on chord range of (4*104 – 1.4*105). Excitation frequencies in the range St ≤5 are found to be effective at Rec= 4*104. Amplitudes only a few dB above the background level being sufficient to produce the effect. For α≥ 18°, during post- stall, significant increase in CL (and CL/ CD) is also achieved.

Ahmed, N. A. and Archer, R. D., **[6]** studies experimentally the post stall behavior of a wing under externally imposed sound. At α = 19 degree, the affect of sound is observed at a smaller frequency range of 200 <f < 1200Hz, and the improvement are less pronounced with up to 15% increase in lift and 10% decrease in drag being observed.

Ishii, et.al, **[7]** studied the effect of acoustic waves with different frequencies on the flow over a wing with a NACA 0012 section at angle of attack α = 12◦ for two Reynolds numbers Re = 5 × 104 and 1 × 105. For these Reynolds numbers the flow separates from the leading edge. It is shown that the acoustic waves with appropriate frequencies make time-averaged lift coefficients higher. In the effective frequency range, the maximum vorticity in the laminar boundary layer of the airfoil becomes larger.

Yarusevych, et.al, **[8]** studied experimentally the boundary layer separation and wake structure of a NACA 0025 airfoil and the effect of external excitations in presence of structural vibrations on airfoil performance. The results establish that external acoustic excitation at a particular frequency and appropriate amplitude suppresses or reduces the separation region and decreases the airfoil wake. The acoustic excitation also alters characteristics of the vertical structures in the wake, decreasing the vortex length scale and coherency.

For the internal acoustic excitation technique, Collins, F. G., **[9]** examined experimentally the effect of sound emitted from periodically spaced holes near the wing leading edge, upon the flow over two low- speed wings, with camber (NACA 2142) and (NACA 0015). This technique found to have a beneficial effect upon the aerodynamic properties of these airfoils. It could be used to improve the low-speed lift and stall performance of light aircraft during take off and landing and could be used for stall/ flutter suppression on rotor and propeller blades.

Hsiao, F. B., and Shyu, R. N., **[10]** explored the control of a wall- separated flow on a five-digit NACA airfoil and a circular cylinder by using the internal acoustic excitation technique. Throughout the experiments, the sound pressure level was always kept at the value of 95dB measured at the slot exit with the effective frequencies ranging from 100 to 400 Hz. Data indicated that the excitation frequency and the forcing location are the key parameters for controlling the separated flow, and the forcing level is the least- effective parameter.

Hsiao, F. B., and Shyu, J. Y., **[11]** studied the separated flow properties and corresponding aerodynamic behaviors of a high AOA, 63- 018 NACA airfoil under internal acoustically pulsing excitation in a subsonic wind tunnel. The experimental results show the following.

1. The shear layer instability frequency, which, increases with increasing Reynolds number, is easily excited by a periodic pulsing fluctuation at the same frequency.
2. For the low post- stall angle airfoil performance (AOA= 18- 24), the leading edge flow separation is suppressed by excitation of a frequency near the shear layer instability.

3. The most effective forcing frequency for improving the aerodynamics properties is to match the vortex shedding frequency in the wake.

Khuder, N. A., **[12]** studied experimentally the influence of internal acoustic excitation with changing the excitation position on the aerodynamic coefficients for the five digit (NACA 23015) airfoil. For the two-excitation position (6% chord and 11.5% chord) and a certain value of Reynolds number ($3x10^5$), at angles of attack values (3^o, 6^o, 9^o and 12^o) tests were done. The tests showed that the internal acoustic excitation at a certain frequency (150 Hz) improving aerodynamic performance.

The present study investigates the effectiveness of the internal acoustic excitation technique and the position of the excitation on the separated flow properties and its relevant aerodynamic performance on an airfoil. NACA 23015 airfoil have been investigated experimentally and numerically. The solution of the flow equations are presented for an angle of attack range (14, 16, 18, 20 and 24) degrees, at excitation frequency values (100, 150 and 200) Hz, with the two-excitation location from the leading edge (6.5% and 11.5%) of chord, at Reynolds number based on chord of $3.4x10^5$.

Mathematical and Numerical Formulation:

In order to analyze the flow field around airfoils with acoustic excitation, a solution of Navier-stokes equations is required. Because of the complexity of airfoils configurations and the strong viscous effects, it is impossible to obtain an analytical solution of the Navier- stokes equation for practical configurations. Thus, numerical techniques have to be used to solve those equations. The need for the full Navier- stokes simulation of complex fluid flows arises in numerous engineering problems. The five digits NACA, which is used in the present work (see figure (4)) is defined completely by the formula below; the thickness distribution is**:**-

$$T(x) = \tau\ c\left[1.4845\sqrt{\frac{x}{c}} - 0.6300\ \frac{x}{c} - 1.7580\left(\frac{x}{c}\right)^2 + 1.4215\left(\frac{x}{c}\right)^3 - 0.5075\left(\frac{x}{c}\right)^4\right] \quad ...(1)$$

Where, c is the airfoil chord and x is the distance along the chord line from the leading edge. The parameter τ is the thickness ratio of the airfoil (maximum thickness/chord).

Flow equations (momentum, continuity and turbulence model equations) for steady two-dimensional flow are solved at the present work for the NACA 23015 airfoil model with the effect of internal acoustic excitation.

-Assumptions:

In the present work, the working fluid is air and the flow characteristics are assumed to be as follows,

- Steady state,
- Newtonian,
- Incompressible,
- Two dimensional and
- Isentropic flow.

The general partial differential equation (i.e. sometimes called transport equations) for continuity, momentum and ($\kappa - \varepsilon$) model, have the form **[13]**:-

$$\left(\rho\ u\ \phi\right)_x + \left(\rho\ v\ \phi\right)_y = \left(\ \Gamma\phi_x\right)_x + \left(\ \Gamma\phi_y\right)_y + S_{x,y} \qquad \text{..... (2)}$$

The arrangement (ϕ) identifies the dependent variable, (Γ) is the exchange coefficient for variable (ϕ) and ($S_{x,y}$) is the source term.

In order to solve the governing equations of motion in the computational space, a transformation of the equation (2) expressed in the Cartesian coordinate system (x,y) from physical space into computational space (ξ,η) is required, which can be written as :-

$$\left(\rho\ G1\ \phi\right)_\xi + \left(\rho\ G2\ \phi\right)_\eta = \left(\Gamma\ J\ a1\ \phi_\xi\right)_\xi + \left(\Gamma\ J\ a2\ \phi_\eta\right)_\eta + S_{new} \qquad \text{..... (3)}$$

Computational solutions are obtained in the present work on staggered grid. This implies that different dependent variables are evaluated at different grid points. Peyret, R. and Taylor, T. D., **[14]** compare various staggered grid for the treatment of the pressure.

The simplest grid generation technique is the algebraic method which is used in the FLUENT 6.1 for the present work.

-Implementation of Boundary Conditions:
-Inlet Boundary Conditions:

An approximation for the inlet distribution for κ and ε can be obtained from turbulent intensity (Ti), typically 1-6 %, and a characteristic length (L) by means of the following simple assumed forms.

$$\kappa = \frac{3}{2}(uTi)^2 \quad \text{... (4)}$$

$$\varepsilon = \frac{C_\mu^{3/9} k^{3/2}}{l} \quad \text{... (5)}$$

$$l = 0.07L \quad \text{... (6)}$$

The fluid properties at the inlet were the atmospheric air properties with Reynolds number value based on the chord of (3.4x105).

-The Boundary Conditions at the Wall:

The sound pressure is the input condition at the wall for the present work, which is calculated according to the corresponding acoustic excitation frequency used. See table (1)

Table (1) Sound Pressure Values According To Sound Frequencies

Frequency (Hz)	Pressure (Pa)
50	0.14278
100	0.14278
150	0.14266
200	0.14253
250	0.14227
300	0.14227
350	0.14266
400	0.14253
450	0.14189
500	0.14227

Apparatus, Experimental Set up and Method of Investigation:

-Experimental Apparatus:

- Subsonic Wind tunnel
- Smoke Wind tunnel

-Instrumentations of local flow:

- Pitot - static tube
- Static Tube
- Multi-Tube Manometer

-Sound Excitation Cycle (see figure (2)):

- Function Generator
- Power Amplifier
- Microphone
- Frequency Meter
- Speaker in an Isolated Wood Box. See figure (3).

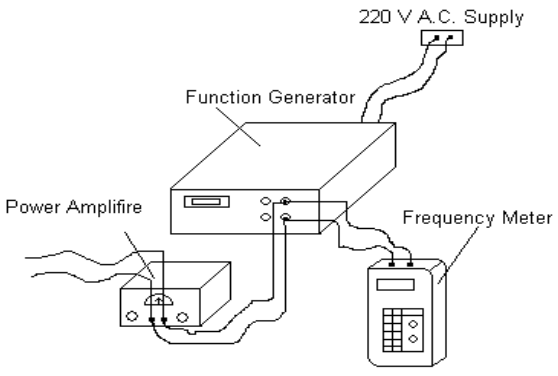


Fig. (2) Excitation Cycle

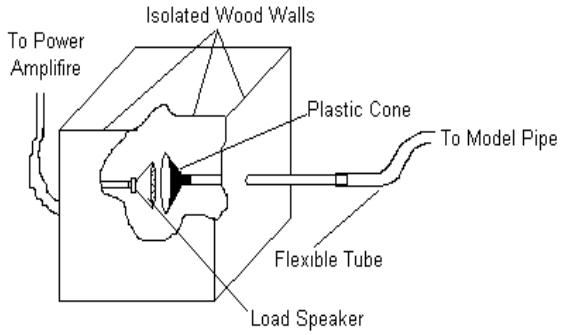


Fig. (3) Excitation box

The experimental work is performed on models A and B for the open loop wind tunnel tests, and on models C and D for the flow visualization tests see figure (4).

Fig. (4) Model A, B, C, and D Airfoils.

For the wind tunnel tests the experiments are conducted to measure the pressure distribution at the (16) upper and (8) lower taps of the models A and B.

First, the sound pressure level (SPL) was measured at the exit of the forcing slot by a CEL-254 digital sound level meter, with no flow blowing. SPL is always kept at the value of (78) dB measured at the slot exit (it is found that the variation of the pressure distribution due to the sound pressure level change is not so significant when compared to that of the frequency changes **[10]**).

There are three main testing conditions:

(1) Taped slot and no excitation.
(2) Open slot with no excitation.
(3) Open slot with different excitation frequencies.

The first two cases provided a baseline for comparisons.

For the above three testing conditions wind tunnel tests are carried out at Reynolds number ($3.4x10^{5)}$ based on chord. Over models A and B, the angle of attack was varied from 14^o to 24^o, by a step 3^o, and the excitation frequency for each angle was varied from 50 Hz to 500 Hz, by a step 50 Hz.

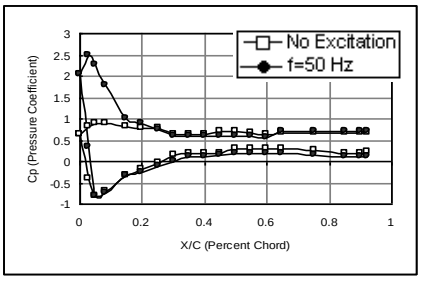
The flow visualization was done for the three conditions mentioned with model C and D, and the flow pattern was photographed. Tests are done for a certain value of Reynolds number (1×10^4) based on the chord, and constant sound pressure level (78 dB).

Results and Discussions:

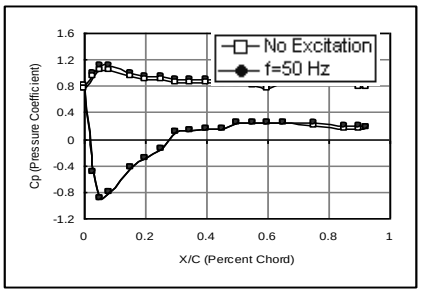
- Experimental Results:

The comparison of the surface pressure coefficients distributions without excitation and with excitation at frequencies (50, 150, and 300) Hz for the excitation locations (11.5%, and 6.5%) of chord are presented in figures (5a-5b) and (6a-6b), respectively.

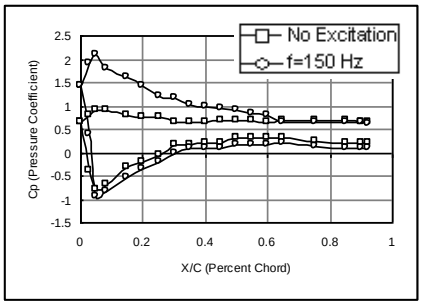
(5a)



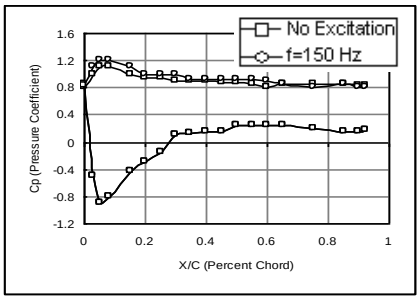
(6a)



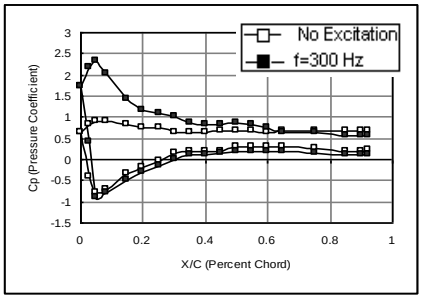
(5b)



(6b)



(5c)



(6c)

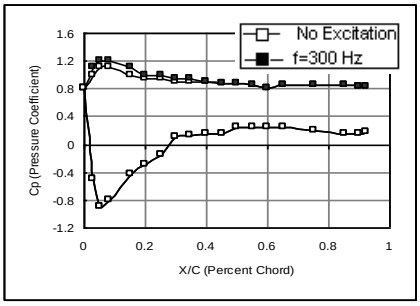


Fig. (5)
Experimental Pressure Coefficients Distribution 11.5% chord 14 Deg.

Fig. (6)
Experimental Pressure Coefficients Distribution

The larger suction peak area will result in a substantial contribution to the lift for the enhancement of the flow mixing and momentum transport due to internal excitation produces a suction peak at the leading edge of the upper surface of the airfoil. The suction peaks results in an increase of lift.

Figure (7) depicts the dependence of lift on the excitation location at angle of attack range (0-24) degrees

It can be seen that effectiveness of the boundary layer control with internal excitation strongly depends on the excitation location, and excitation at a location close to the separation point is the most effective, especially in the post stalled region.

Figures (8a-8c) show the typical flow patterns at the Reynolds number based on chord of $1x10^4$ and angle of an attack (14, 16, and 20) degrees respectively, with 150 Hz excitation frequency at excitation locations (6.5%, and 11.5%) of chord and without excitation for taping slot or not.

The separated flow at the leading edge is clearly revealed when the flow is unexcited. It would cause a severe deterioration in lift.

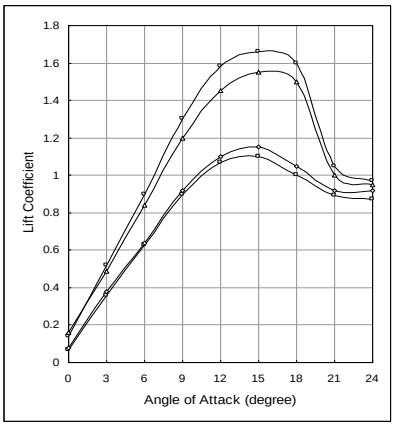


Fig. (7)
Comparison of Experimental Lift Coefficient Curves

After the flow is internally excited by the acoustic waves at the excitation frequency of 150 Hz at the stalled region (effectively for 6.5% chord excitation location), the separated boundary layer is then reattached to the boundary of the airfoil.

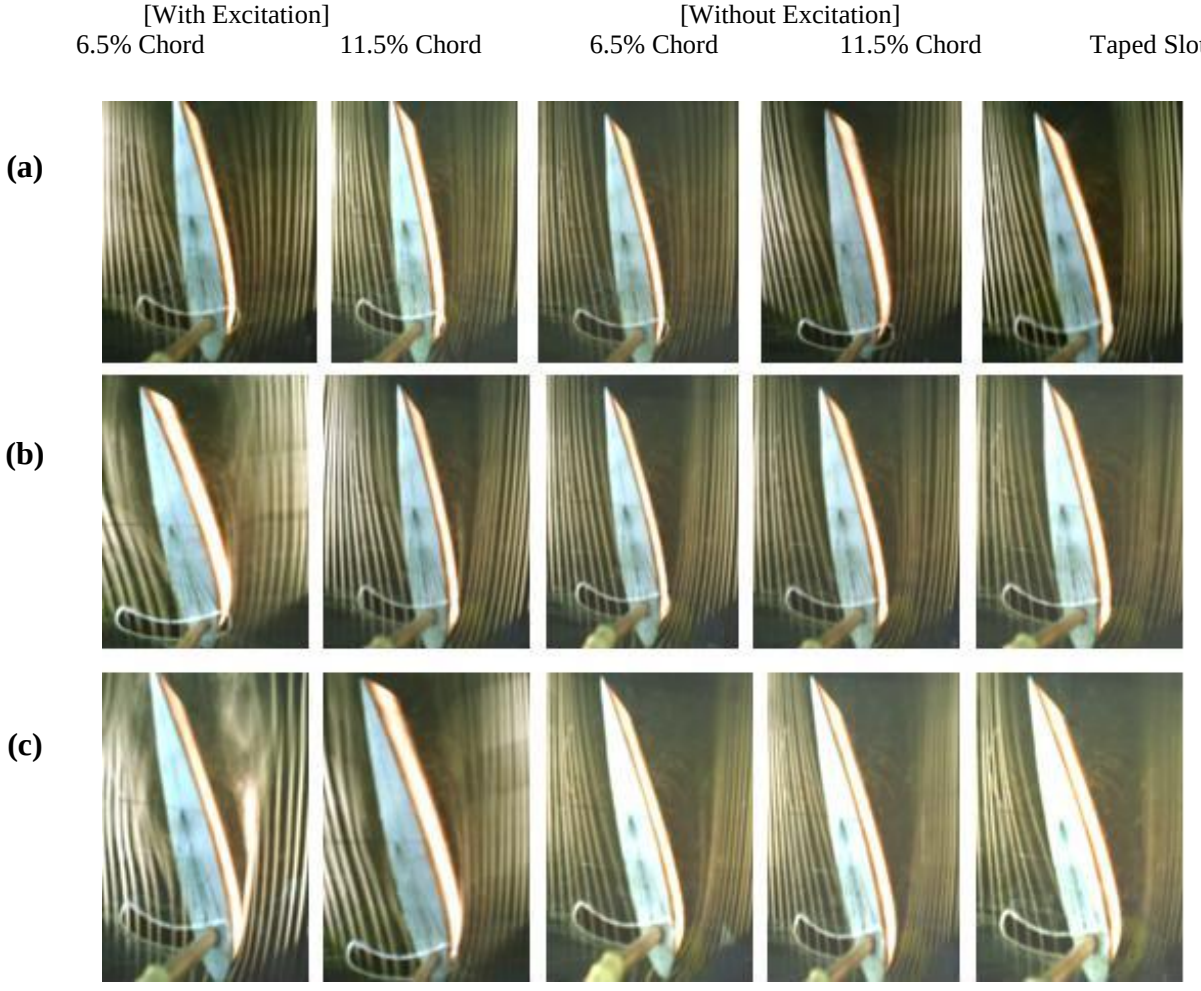


Fig. (8)
Comparison of Flow Patterns for a 150 Hz Excitation Frequency

The separated flow at the leading edge is clearly revealed when the flow is unexcited. It would cause a severe deterioration in lift. After the flow is internally excited by the acoustic waves at the excitation frequency of 150 Hz at the stalled region (effectively for 6.5% chord excitation location), the separated boundary layer is then reattached to the boundary of the airfoil. The reattached boundary layer will certainly ensure the lift recovery. In addition, since the wake region is narrowed due to the boundary-layer reattachment, the drag will be reduced accordingly. A narrower wake with a smaller profile defect indicates a less momentum loss, which insures a smaller drag coefficient.

The typical patterns of velocity vector and streamlines are presented in figure (9) for the case of without excitation and with excitation for two excitation locations (6.5%, and 11.5%) of chord at excitation frequency 150 Hz.. At the prestalled region, the boundary layer remains attached over the entire lower surface of the airfoil but it separates somewhere near the rear surface of the upper surface. At the poststalled region, the fluid particles are forced outwards from the wall and form a separated region. In general the fluid particles behind the point of separation follow the pressure gradient and move against the direction of the main flow (the appearance of vorticity in fluids). The flow separation occurs over a major portion of the upper surface of the airfoil which is around (60-70) percentage of the chord (high wake region). Therefore, controlling of the boundary layer is the supplying of additional energy to the boundary layer by an effective excitation frequency (150 Hz), thus enabling the boundary layer to proceed further against an adverse pressure gradient (delay of separation point). Narrower wake region can be seen at excitation location (6.5% of chord).

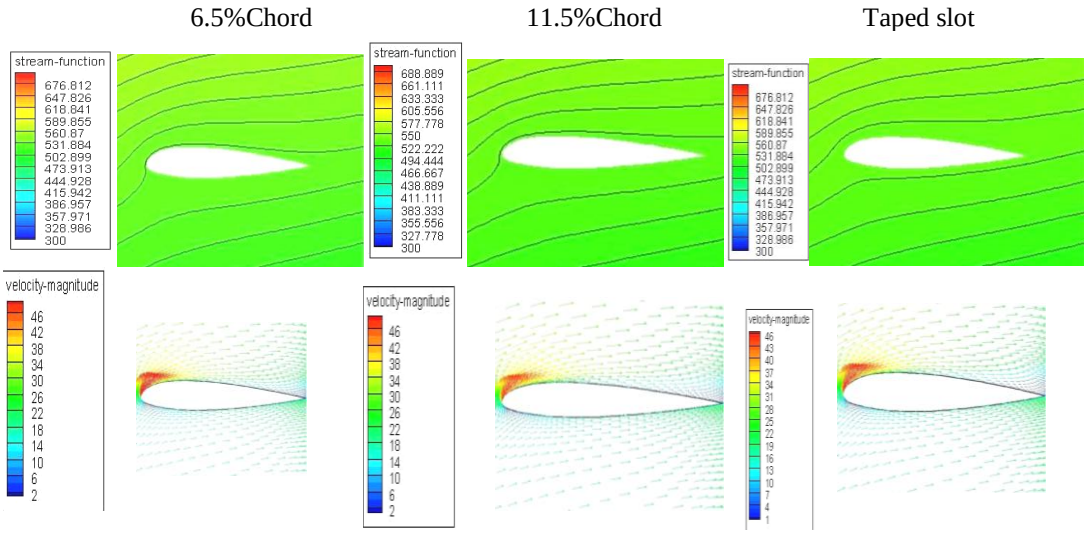


Fig. (9)
Numerical Flow Patterns (Velocity Vectors and Streamlines) for 23015 NACA Airfoil at Angle of Attack 21^o without Excitation and at 150 Hz Excitation Frequency for (11.5%, 6.5%) Chord Excitation Location

The history of the lift coefficient with the angle of attack is showed by figure (10). The results in the case of no excitation and those for acoustic excitation with different frequencies (100, 150 and 200 Hz) are presented. The effective frequency value is (150 Hz) where, the increase in lift coefficient as compared with the non-excitation case exceeds 35-45% for the 6.5% chord excitation location and 30-35% for 11.5% chord position. One can conclude that , by applying the acoustic excitation internally for the flow at the shear layer instability frequency "the double advantages" that is higher lift and less drag, will leads the higher value of the Lift-to-Drag coefficient ratio as shown in figure (11). This in turn ensure the high performance of the present excitation method.

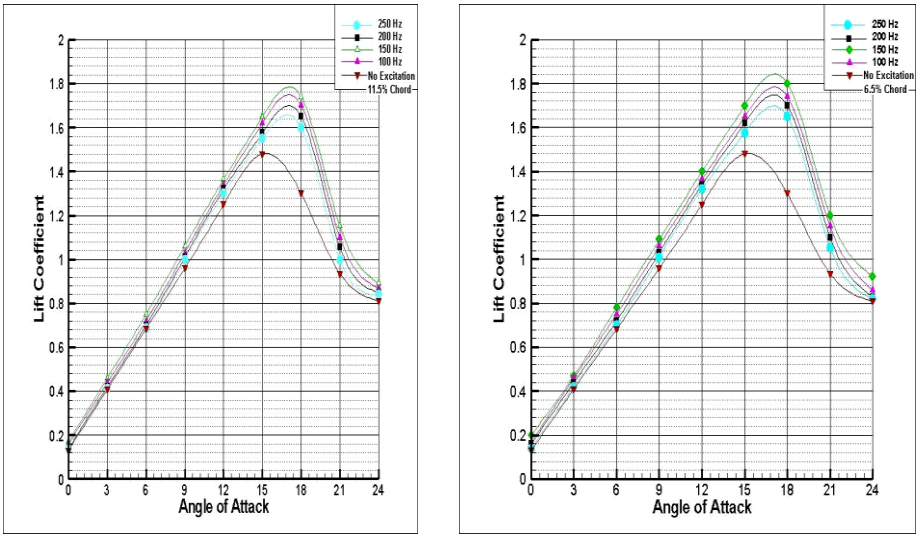


Fig. (10)
Numerical Comparison of Lift Coefficient Curves for 6.5, 11.5 % Chord Excitation Location.

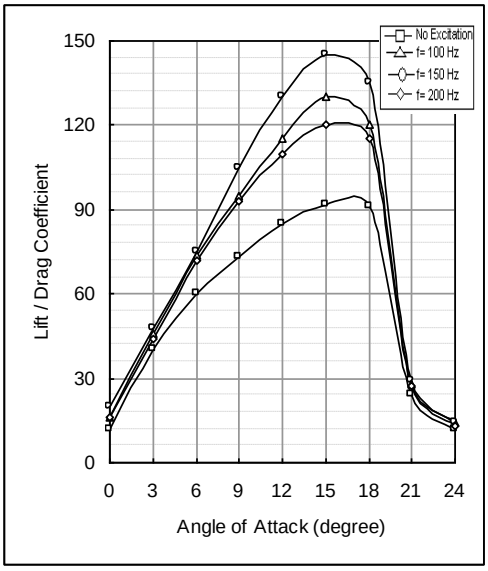
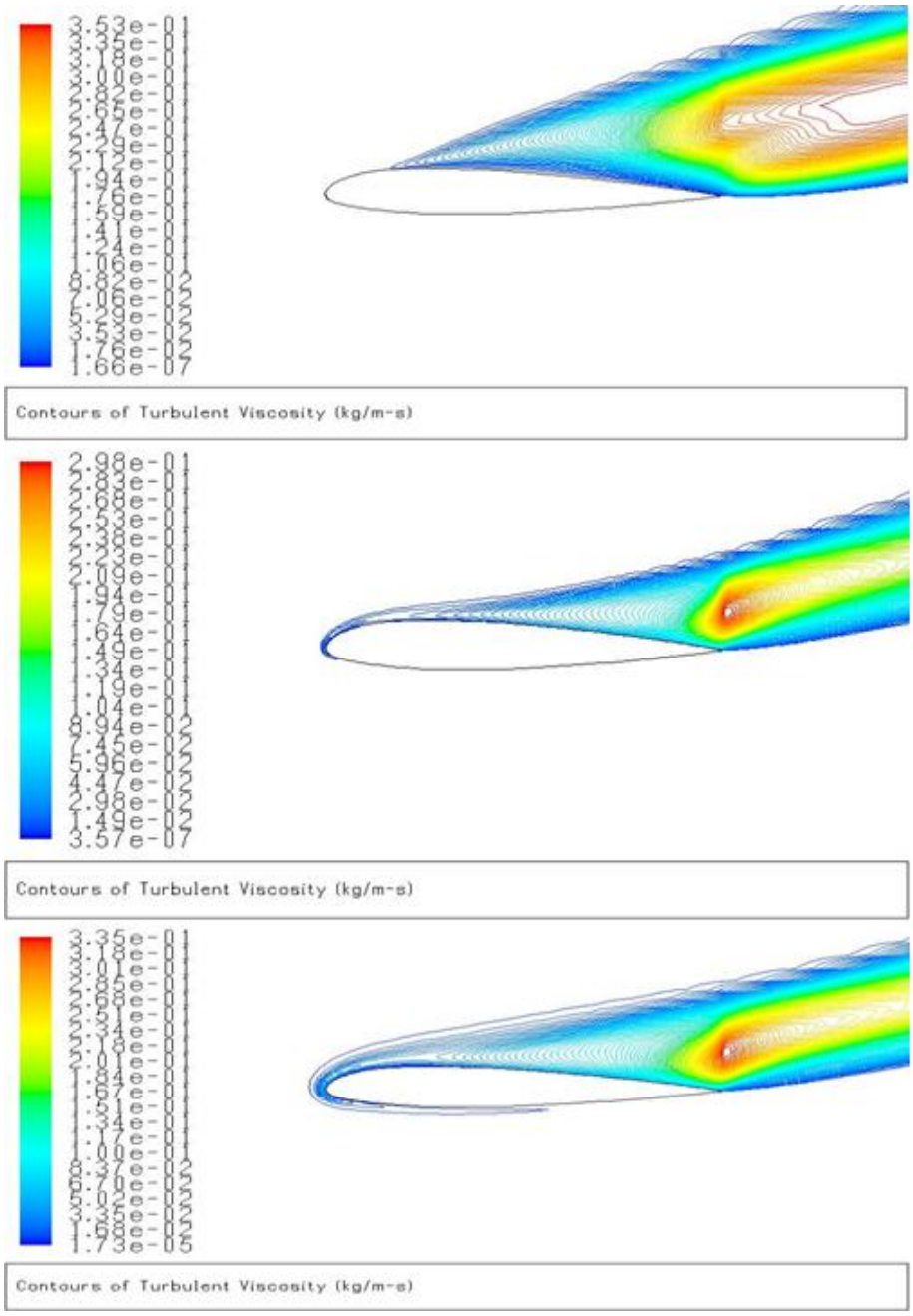


Fig. (11)
Numerical Comparison of Lift/Drag Coefficient (6.5% Chord Excitation Location) With Different Excitation Frequencies.

There are two factors of flow affecting separation, the adverse pressure gradient and viscosity. The control of separation can be achieved by changing or maintaining the structure of viscous flow so that these two governing factors prevent or delay the separation. Figure (12) shows the effect of the internal acoustic excitation on the boundary layer growth where, it causes the reattachment to move the suction peak at the leading edge downstream, thus reducing the pressure gradient.



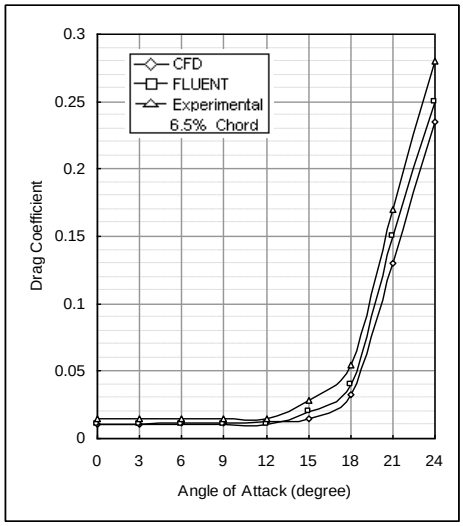
Taped Slot

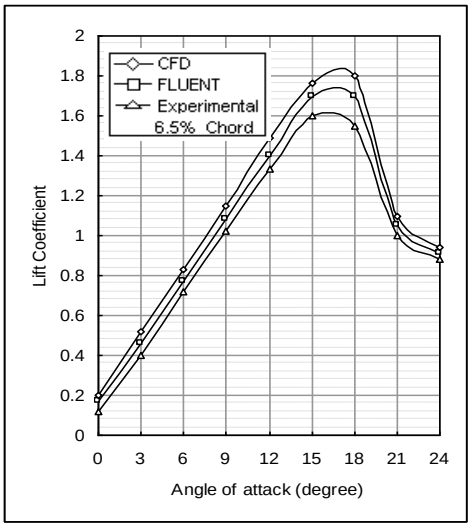
11.5% chord

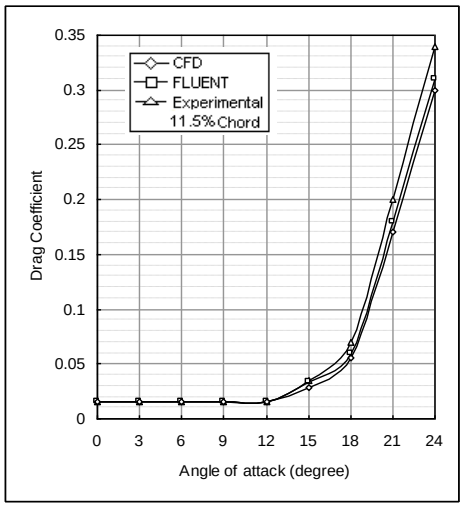
6.5% chord

Fig. (12)
Turbulent Viscosity Contours for 23015 NACA Airfoil at Angle of Attack 21^{o} and 150 Hz Excitation Frequency

Comparisons between the results of the experimental, numerical and Fluent (6.1), of lift and drag coefficients for the case of 6.5% and 11.5% of chord excitation location with (150 Hz) excitation frequency are shown in figures (13) and (14). It is found that the results of computational work agree well with the results obtained by Fluent (6.1), but the experimental results are around 20% lower than the numerical results at the maximum lift point. This difference may be due to errors in measurements and the environmental conditions at the libratory (temperature, humidity, air movement and noise), they all change continuously during the test time and this should affect the results. In addition, the numerical work has been done for the free stream conditions while; the experimental work was bounded by the wind tunnel conditions.







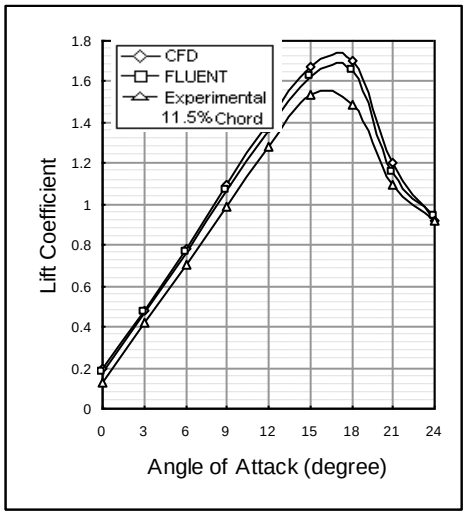


Fig. (14)
Comparison of Lift Coefficient for 6.5% Chord Excitation Location and 150 Hz Excitation Frequency for Numerical, Fluent (6.1) and Experimental Results

Fig. (13)
Comparison of Lift Coefficient for 6.5% Chord Excitation Location and 150 Hz Excitation Frequency for Numerical, Fluent (6.1) and Experimental Results

Conclusions and Suggestions:

1. The enhancement of the flow mixing and momentum transport due to internal acoustic excitation produces a suction peak at the leading edge of the upper surface of the airfoil. The suction peak results in an increase of lift and narrower wake.
2. By the flow visualization, it is found that the locally introduced unsteady vorticity causes the separated boundary layer to be reattached to the surface.
3. The internal acoustic excitation energizes the boundary layer, this leads to decrease the turbulent kinetic energy at the upper surface of the airfoil.
4. The results suggested that there is a critical excitation frequency (150 Hz).
5. The excitation location is the most affected parameter on the internal acoustic excitation technique and the results indicated that, the excitation location close to the leading edge is the more efficient. Internal acoustic excitation at 6.5% of chord lead to increase lift by 45% while, the 11.5% of chord excitation location gives only 35% increase.

References:

1- Miranda, S., "**Active control of separated flow over a circular-arc airfoil,**" M. Sc. Theses, Blacksburg, Virginia, May 8, 2000.

2- Thomas, T. G., and Tutty, O. R., "**Flow Control**," AA407/SE603 Flow Control 24 Lectures in Semester 1, 2003.

3- Schubauer, G. B., and Skramstad, H. K., "**Laminar boundary layer transition on a flat plate**," NACA report, 909, 1948.

4- Zaman, K. B. M. Q., Bar-Sever, A., and Mangalarn, S. M., **"Effect of acoustic excitation on the flow over a low-Re airfoil,"** J. Fluid Mech., vol. 182, pp. 127-148. , 1987

5- Peterka, J.A., and Peter, P.D., **"Effect of sound on separated flows,"** J. Fluid mech., vol. 37, part 2, pp.265-287, (1968).

6- Ahmed, N. A. and Archer, R. D., **" Poststall behavior of a wing under externally imposed sound, "** J. Aircraft, vol. 38, No. 5, pp.961, (2001).

7- Ishii, K., Suzuki, S., and Adachi, S., **"Effect of weak sound on separated flow over airfoil,"** Fluid Dynamic Research 33, 357-371, (2003).

8- Yarusevych, S., Kawall, J. G., and Sullivan, P. E., **"Airfoil performance at low Reynolds numbers in the presence of periodic disturbances,"** Journal of fluids engineering, vol. 128/587, May (2006).

9- Collins, F. G., **"Boundary-Layer control on wings using sound and leading-edge serrations,"** AIAA Journal, vol. 19, No. 2, pp. 129. 1981.

10- Hsiao, F. B., and shyu, R. N., **"The effect of acoustic on flow passing a high-AOA airfoil,"** Journal of sound and vibration 199 (2), 177-188, (1997).

11- Hsiao, F. B., and Shyu, J. Y., **"Control of wall-separated flow by internal acoustic excitation,"** AIAA Journal, vol.28, No.8, pp.1440, 1989.

12- Khuder, N. A., **"Improvement of the aerodynamic performances using the internal acoustic excitation,"** M. Sc. Theses, Mechanical Department, Univ. of Tech., Iraq (2009).

13- Roache, P. J., **"Computational Fluid Dynamics,"** Hermosa Publishers, Albuquerque, 1972.

14- Peyret, R. and Taylor, T. D., **"Computational methods for fluid flow,"** Springer, New York, 1983.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-385-390 ISSN: 2249-6645

Dual-Hop Network with Cooperative Communication

H. N. Pratihari
Department of Electronics & Telecommunication
Orissa Engineering College, Bhubaneswar

***Abstract*-The paper deals to proposed a cooperative communications to exploit spatial diversity gains inherent in multi-user wireless systems without the need of multiple antennas at each node. Transmit diversity generally requires more than one antenna at the transmitter. However, many wireless devices are limited by size or hardware complexity to one antenna. Which communication channels are in use and which are not, and instantly move into vacant channels while avoiding occupied ones. The proposed work optimizes the use of available radio-frequency (RF) spectrum while minimizing interference to other users. In its most basic form, CR is a hybrid technology involving software defined radio (SDR) as applied to spread spectrum communications.**

I. INTRODUCTION

The mobile wireless channel suffers from fading, meaning that the signal attenuation can vary significantly over the course of a given transmission. Transmitting independent copies of the signal generates diversity and helps in combating signal fading due to multi-path propagation in wireless medium. For transmitting independent copies of a signal, that is to achieve transmit diversity, multiple antennas will be required at the source terminal which may not be practical for certain scenarios due to size, cost, hardware limitations. Examples include most handsets (size), wireless sensor networks (size, power).Spatial diversity is generated by transmitting signals from different locations, thus allowing independently faded versions of the signal at the receiver [1]. Cooperative communication generates this diversity in a new and interesting way. Cooperative communication involves two main ideas: Use relays to provide spatial diversity in a fading environment [1, 2]. Envision a collaborative scheme where the relay also has its own information to send, so both terminals help each other to communicate by acting as relays for each other. One can think of a cooperative system as a virtual antenna array, where each antenna in the array corresponds to one of the partners. The partners can overhear each other's transmissions through the wireless medium, process this information and retransmit to collaborate. This provides extra observations of the source signals at the destinations, the observations which are dispersed in space and usually discarded by current implementations of cellular or ad-hoc systems. With cooperation, users that experience a deep fade in their link towards destination can utilize quality channels provided by their partners to achieve the desired Quality of Service.

II.COOPERATIVE COMMUNICATION

The term cooperative communications typically refers to a system where users share and coordinate their resources to enhance the transmission quality. This idea is particularly attractive in wireless environments due to the diverse channel quality and the limited energy and bandwidth resources. With cooperation, users that experience a deep fade in their link towards the destination can utilize quality channels provided by their partners to achieve the desired quality of service (QoS). This is also known as the spatial diversity gain. In a cooperative communication system, each wireless user is assumed to transmit data as well as act as a cooperative agent for another user. Two features differentiate cooperative transmission schemes from conventional non-cooperative systems:

1) the use of multiple users' resources to transmit the data of a single source,
2) a proper combination of signals from multiple cooperating users at the destination.

A canonical example is shown in Fig. 1, where we have two users transmitting their local messages to the destination over independent fading channels. Suppose that the transmission fails when the channel enters a deep fade, i.e., when the signal-to-noise ratio (SNR) of the received signal falls below a certain threshold, as indicated with the grey region in Fig. 1. If the two users cooperate by relaying each others' messages and the

inter-user channel is sufficiently reliable, the communication outage occurs only when both users experience poor channels simultaneously.

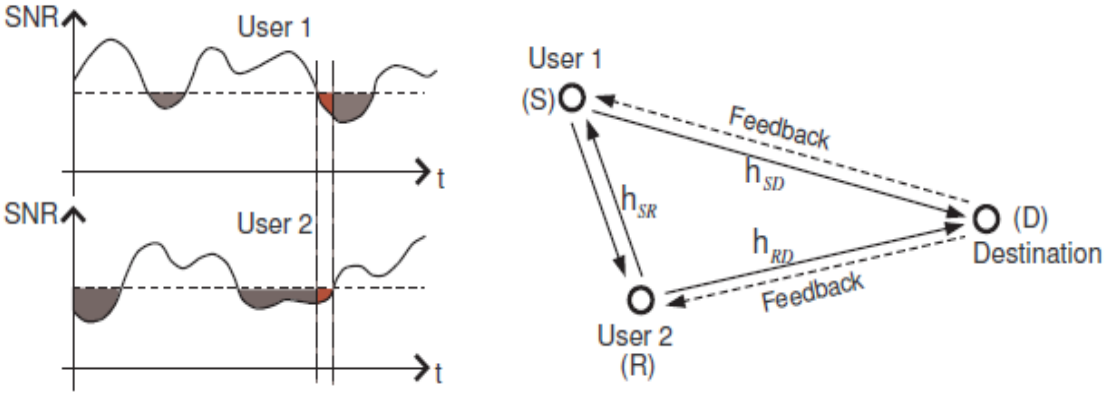


Fig.1. A three node cooperative network model

Cooperative techniques utilize the broadcast nature of wireless signals by observing that a source signal intended for a particular destination can be "overheard" at neighboring nodes. These nodes, called relays, partners, or helpers, process the signals they overhear and transmit towards the destination. The relay operations can consist of repetition of the overheard signal (obtained, for example, by decoding and then re-encoding the information or by simply amplifying the received signal and then forwarding), or can involve more sophisticated strategies such as forwarding only part of the information, compressing the overheard signal, and then forwarding. The destination combines the signals coming from the source and the relays, enabling higher transmission rates and robustness against channel variations due to fading.

III.COOPERATIVE COMMUNICATION SIGNALLING TECHNIQUES

At each time instance, one user acts as the source node while the other user serves as the relay node as shown in Fig. 1. Each user has the right to serve as the source node in a typical cooperative system [1, 2]. At first, the source, e.g. user 1, broadcasts its message to both the relay node and the destination. The relay node can then employ any one of the following cooperative communication signalling techniques to forward the message to the destination. The main and most popular cooperative signaling techniques based on the concept of relaying are:

A) Decode and Forward Method
B) Amplify and Forward Method

A. Decode and Forward Method:

If the relay node employs the Decode and Forward (DF) scheme, it will decode the message received from the source, re-encodes it and then forwards the message to the destination subsequently as shown in Fig.2. When the regenerated message is encoded to provide additional error protection to the original message, it is also referred to as coded cooperation. At the destination, signals from both the source and the relay paths are then combined for detection. This signalling has the advantage simplicity and adaptability to channel conditions.

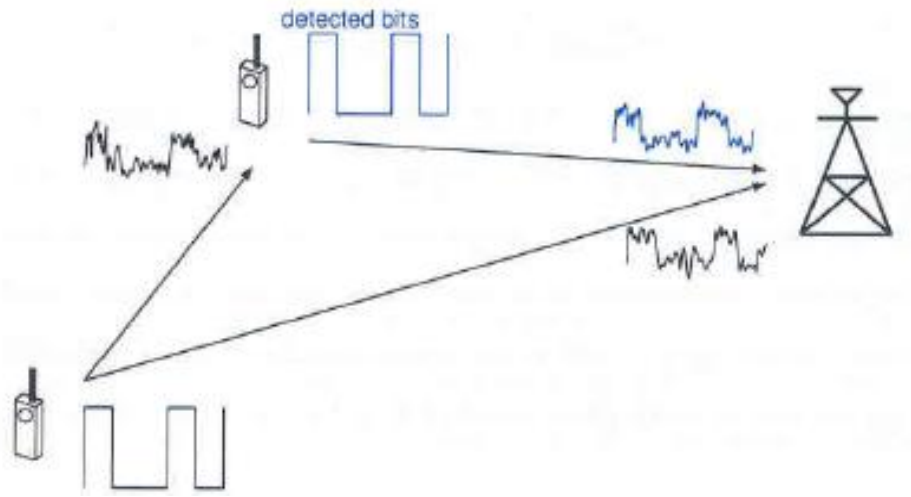


Fig.2 Decode and Forward Method

It is possible that the detection by the partner is unsuccessful, in which case cooperation can be detrimental to the eventual detection of the bits at the destination. To avoid the problem of error propagation, hybrid decode and forward method can be used where, at times when the fading channel has high instantaneous signal to noise ratio, users detect and forward their partners data, but when the channel has low SNR, users revert to a non-cooperative mode.

B. Amplify and Forward Method:

If the Amplify and forward (AF) scheme is employed, the relay node simply amplifies the received signal and forwards it directly to the destination without decoding the message.

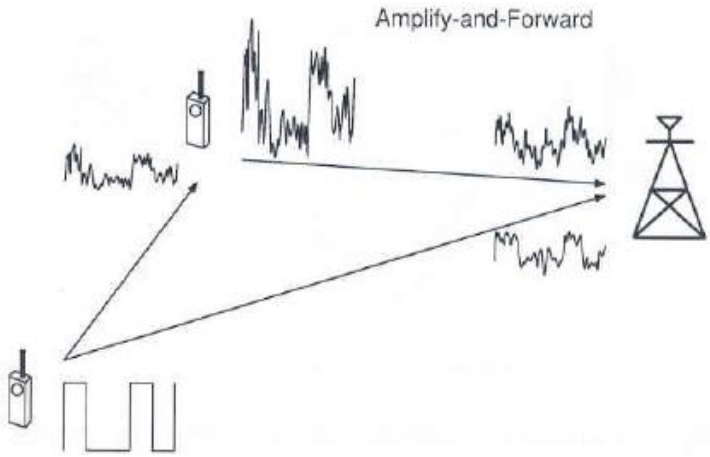


Fig.3. Amplify and Forward Method

Each user in this method receives a noisy version of the signal transmitted by its partner. As the name implies, the user then amplifies and retransmits this noisy version. The base station combines the information sent by the user and partner, and makes a final decision on the transmitted bit (Fig.3). Although noise is amplified by co-operation, the base station receives two independently faded versions of the signal and can make better decisions on the detection of information. In amplify and forward it is assumed that the base station knows the interuser channel coefficients to do optimal decoding, so some mechanism of exchanging or estimating this information must be incorporated into any implementation. The advantages of relay cooperation often rely on sufficiently reliable interuser channels [3]. For example, in the DF scheme, a node is able to relay the message only if it is able to receive from the source reliably while, in the AF scheme, the

quality of the relayed signal is limited by the quality of the source-relay link since both the signal and noise are amplified at relays. Therefore, relays should be adopted only if the source-relay channel is sufficiently reliable. This observation leads to the selective relaying (SR) cooperation scheme where relays are selected to retransmit the source message only if the quality of the transmission over the inter-user channel meets a certain criterion.

IV. ADVANTAGES OF COOPERATIVE COMMUNICATION

The main advantages of cooperative communications are:

A. Higher Spatial Diversity:

As a simple example, Fig. 4a shows a small net-work of four mobile nodes. If the channel quality between mobile nodes S and D degrades severely (e.g., due to shadow or small-scale fading), a direct transmission between these two nodes may experience an intolerable error rate, which in turn leads to retransmissions. Alternatively, S can exploit spatial diversity by having a relay R1 overhear the transmissions and then forward the packet to D as discussed above. The source S may resort to yet another terminal R2 for help in forwarding the information, or use R1 and R2 simultaneously. Similar ideas apply to larger networks as well. Therefore, compared with direct transmission, the cooperative approach enjoys a higher successful transmission probability [4, 5]. We note here that cooperative communications has the ability to adapt and to mitigate the effects of shadow fading.

B. Higher Throughput-Lower Delay:

In Fig.4a, if Rate2 and Rate3 are higher than Rate1 such that the total transmission time for the two-hop case through R2 is smaller than that of the direct transmission, cooperation readily outperforms the legacy direct transmission, in terms of both throughput and delay perceived by the source S. Furthermore, for relays such as R1 and R2, it turns out that their own individual self-interest can be best served by helping others [4,6].

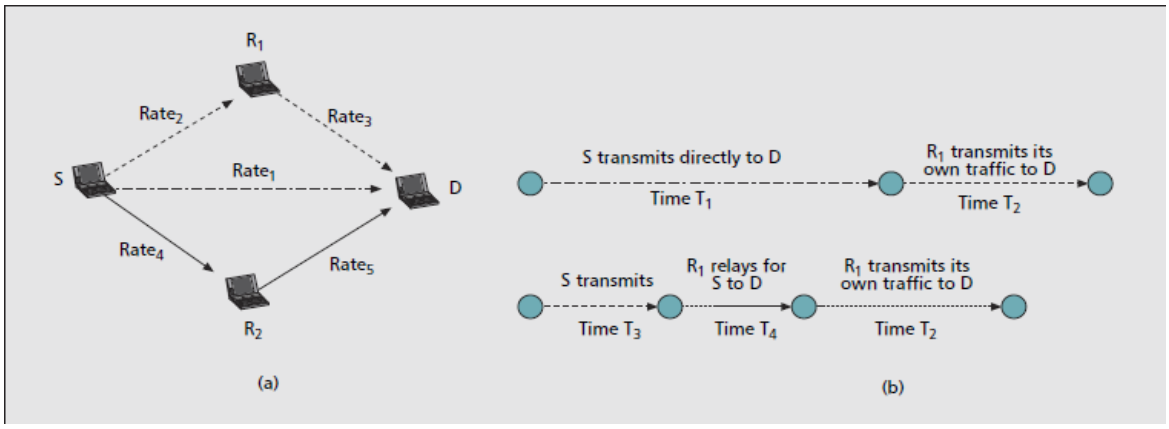


Fig.4. a) Cooperation in a network; b) illustration of the delay and throughput improvement achieved by cooperation in the time domain

As further illustrated in Fig.4b, the intermediate node R1 that cooperates enjoys the benefit of lower channel-access delay, which in turn can be translated into higher throughput.

V.COOPERATIVE COMMUNICATION IN DUAL HOP NETWORK

A. The Model:

The model for the received signal and the channel for a link between any pair of nodes i and j is given by

$$y_j = h_{ij}x_i + n_j$$

Where xi is the signal transmitted by node i, $h_{ij} \rightarrow C\mathcal{N}(0,\Omega_{ij})$ is the complex channel gain over the link i → j, $n_j \rightarrow C\mathcal{N}(0,N0)$ is additive white Gaussian noise at node j. The channel gains, noise, and transmitted signals are

independent. The channel gain h_{ij} captures the effects of fading as well as path loss by setting $\Omega_{ij} = d^{-\alpha}{}_{ij}$, where d_{ij} denotes the distance between node i and node j, and α is the path loss exponent.

B.Dual Hop Relay Network:

Consider N relay nodes as shown in Fig.5, denoted by R_k, k = 1, ⋅ ⋅ ⋅ , N, and let h_{Sk} and h_{kD} denote the complex channel coefficients from the source S to the relay R_k and from R_k to destination node D, respectively. The source node can transmit information to the destination node directly, or transmit information to the destination node via a relay. The relays operate in DF mode, whereby relays are selected proactively to forward the information. The use of relays results in a division of the transmission time into two slots:

i) the first slot for the transmission from the source i.e., S broadcasts its message X_S in the first stage to the destination and to the relays. The received signals at the relay and the destination can be expressed as $X_k = h_{Sk} \cdot X_S + W_R$ and $X_{D1} = h_{SD} \cdot X_S + W_{D1,}$ where h_{Sk} and h_{SD} are the channel coefficients for the S-R_k and the S-D link and W_R and W_{D1} denote the additive channel noise.

ii)The second slot for the transmission from the relay ie, the set of relays $\{R_k, k = 1, \cdot, N\}$transmits symbol $U_k = f(X_k)$ as a function of the received signal X_k simultaneously to the destination in the second stage. Consequently, the signal received at D

$$Z = \sum_{k=1}^{N} h_{kD}U_k + W_D$$

Where W_D is the AWGN with unit variance and N is the total number of relay nodes in the network.

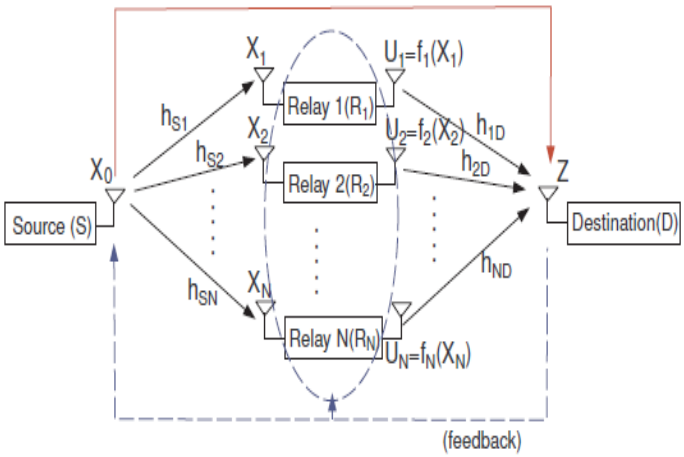


Fig.5. Dual Hop Relay Network

VII .CONCLUSION

In cooperative communication, a user can share the resources of other users to convey their message to the destination. In this system two or more active users in a network can share their information and jointly transmit their messages, either at different times or simultaneously. This results in advantages like spatial diversity, greater reliability and reduction in cost. We have studied the different cooperative signalling techniques for maintaining the data integrity at the destination. We identified the similarities between a dual hop network and a wireless sensor network application and showed how cooperative communication technique can be used by the sensors for reliable communication.

REFERENCES

[1] Yao-Win Hong, Wan-Jen Huang, Fu-Hsuan Chiu and C. Jay Kuo, "Cooperative Communications in Resource-Constrained Wireless Networks," Draft,25 January 2007.

[2] Aria Nosratinia, Todd E. Hunter, Ahmadreza Hedayat, " Cooperative Communication in Wireless Networks" , IEEE Communications Magazine, Vol. 42, Issue 10, Oct. 2004, Page(s): 74-80.

[3] Van Der Meulen, E.C.; "Three-terminal communication channels", Adv. Appl. Prob. 3, 1971.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity–Part I: System Description" and "User Cooperation Diversity–Part II: Implementation Aspects and Performance Analysis," IEEE Trans. Commun., vol. 51, no. 11, Nov. 2003.

[5] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Ander-son, "Wireless sensor networks for habitat monitoring," in Proc. Of ACM Inter. Workshop on Wireless Sensor Networks and Applications (WSNA'02), Atlanta GA, USA, 2002.

[6] H.Scarf, "On the existence of cooperative solutions to a class of N-Person games", Game-theoretic models of cooperation and conflict, Westview Press, Boulder, San Francisco, Oxford, 1992.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-391-395 ISSN: 2249-6645

A Behavioral Study of AODV with and without Blackhole Attack in MANET

Arti Sharma and Satendra Jain
SATI, Vidisha, India

Abstract- Wireless mobile ad-hoc networks are those networks which has no physical links between the nodes. Due to the mobility of nodes, interference, multipath propagation and path loss there is no fixed topology in this network. Hence some routing protocol is needed to function properly for these networks. Many Routing protocols have been proposed and developed for accomplishing this task. The intent of this paper is to analyze the performance of ad-hoc routing protocol AODV with and without black hole attack in wireless network. This paper concentrates evaluating the performance of routing protocol when black hole attacks involve in wireless network and when black hole attack not involve in wireless network. The performance analysis for above protocol is based on variation in speed of nodes in a network with 50 nodes. All simulation is carried out with QualNet 5.0 simulator.

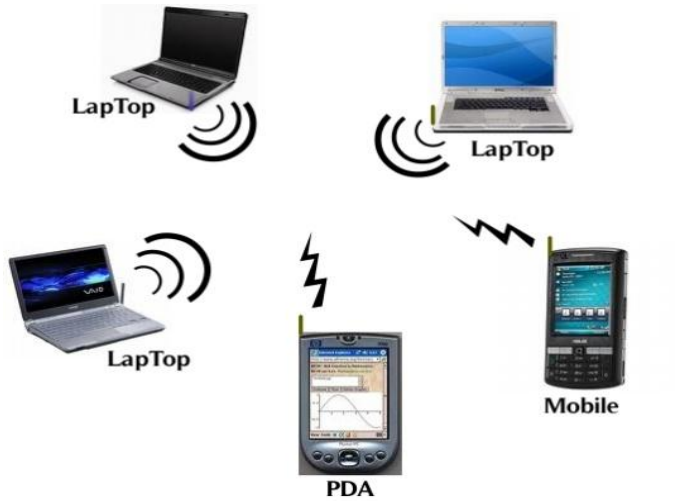
Keywords: *Ad Hoc Networks, routing protocol, Black hole attack, AODV.*

I. INTRODUCTION

Mobile ad hoc networks (MANETs) [1,2] are collections of mobile nodes, which are Dynamically form a temporary network without pre-existing network infrastructure or any centralized administration. These nodes can be arbitrarily located and are free to move randomly at any given time. Every mobile node acts itself as a router. Since there is no centralized administration, so MANET is oftenly called autonomous. MANET implies that the topology may be dynamic - and that routing of traffic through a multi-hop path is necessary if all nodes are to be able to communicate. A key issue in MANETs is the necessity that the routing protocols must be able to respond rapidly to topological changes in the network. At the same time due to the limited bandwidth available through mobile radio interfaces it is imperative that the amount of control traffic generated by the routing protocols is kept at a minimum. Several protocols have been addressed these problems of routing in mobile ad-hoc networks. These protocols were divided into two classes: depending upon the type of requirement and the available resources, when a node acquires a route to a destination.

Proactive protocols [3] are characterized by all nodes maintaining routes to all destinations in the network at all times. Thus using a proactive protocol a node is immediately able to route (or drop) a packet. Examples of proactive protocols include the "FISHEYE". [25], the "Optimized Link State Routing Protocol" (OLSR) [9] and the "Source Tree Adaptive Routing" (STAR) [6]. *Hybrid* protocols [3, 4] are those protocols which have characteristics of both reactive and proactive. Example of hybrid protocol includes "Dynamic MANET On-demand routing protocol" (DYMO) [27].

Reactive protocols [3] are characterized by nodes acquiring and maintaining routes ON-demand. In general, when a route to an unknown destination is required by a node, then the route request is flooded onto the network and replies, containing possible routes to the destination, are returned. Examples of reactive protocols include the "Ad Hoc on Demand Distance Vector Routing Protocol" (AODV) [27] and "Dynamic Source Routing" (DSR) [5].



In this paper, the analysis of routing protocol AODV is presented against black hole attack. The performance of this protocol is analyzed with varying speed of nodes in network. The network contains 50 wireless nodes in which 10 nodes are in black hole attack. These nodes either stop packet forwarding or send wrong and unusual information to other nodes which affects packet drop and lesser throughput.

The organization of this paper is as follows. Section 2 briefly describes the routing protocols AODV. Section 3 briefly describes the affects of black hole attack in network. Section 4 presents experimental configuration. Section 5 focused on results and analysis of the work and Section 6 represents a conclusion of the paper.

II. ROUTING PROTOCOLS

The nature of mobile ad hoc networks makes simulation modeling an invaluable tool for understanding the operation of these networks. In Ad-hoc network multiple routing protocols have been developed during the last years, to find optimized routes from a source to some destination. To establish a data transmission between two nodes, typically multiple hops are required due to the limited transmission range. Mobility of the different nodes makes the situation even more complicated.

The protocols to be used in the Ad Hoc networks should have the following features:

- The protocol should adapt quickly to topology changes.
- The protocol should provide Loop free routing.
- The protocol should provide multiple routes from the source to destination and this will solve the problems of congestion to some extent.
- The protocol should have minimal control message overhead due to exchange of Routing information when topology changes occurs.
- The protocol should allow for quick establishment of routes so that they can be used before they become invalid.

Ad hoc On-Demand Distance Vector (AODV) [27]

The Ad hoc On-Demand Distance Vector (AODV) routing protocol is intended for use by mobile nodes in an ad hoc network. It offers quick adaptation to dynamic link conditions, low processing and memory overhead, low network utilization, and unicast route determination to destinations within the ad hoc network. It uses destination sequence numbers to ensure loop freedom at all times (even in the face of anomalous delivery of routing control messages), avoiding problems (such as "counting to infinity") associated with classical distance vector protocols.
The primary objectives of AODV protocol are [27]:

- To broadcast discovery packets only when necessary,
- To distinguishes between local connectivity management (neighborhood detection) and general topology maintenance and
- To disseminate information about changes in local connectivity to those neighboring mobile nodes those are likely to need the information. AODV decreases the control overhead by minimizing the number of broadcasts using a pure on-demand route acquisition method. AODV uses only symmetric links between neighboring nodes.

III. BLACK HOLE ATTACK

In Blackhole attack all networks traffics are redirected to a specific node which does not exist at all.Because traffics disppear into the special node as the matter disappears into Blackhole in universe .So the specific node is named as a Blackhole.A Blackhole has two properties.First, the node exploits the ad hoc routing protocol,such as AODV , to advertise itself as having a valid route to a destination node, even though the route is spurious,with the intention of intercepting packets. Second, the node consumes the intercepted packets.

IV. EXPERIMENT CONFIGURATION

All the simulation work is performed in QualNet wireless network simulator version 5.0 [3]. Initially number of nodes are 50, simulation time was taken 180 seconds . All the scenarios have been designed with a terrain 1500m x 1500m. Mobility model used is Random Way Point [26] (RWP). In this model a mobile node is initially placed in a random location in the simulation area. For simulation, speed of node is varying from 10mps to 50mps. All the simulation works were carried out using routing protocol AODV with varying speed of node. Network traffic load is provided by constant bit rate (CBR) application. A

CBR traffic source provides a constant stream of packets throughout the whole simulation, thus further stressing the routing task. There are four measurements in our experiments were defined as follows:

1) *Throughput (bits/s):-* Throughput [26] is the measure of the number of packets successfully transmitted to their finaldestination per unit time.

2) *Total Packets received:* - Packet delivery ratio **[27]** is calculated by dividing the number of packets received by the destination through the number of packets originated by the application layer of the source (i.e. CBR source).

***3) End-to-end delay*:** Average End to End Delay **[27]** signifies the average time taken by packets to reach one end to another end (Source to Destination)**.**

4) Average Jitter Effect: Signifies the Packets from the source will reach the destination with different delays [5]. A packet's delay varies with its position in the queues of the routers along the path between source and destination and this position can vary unpredictably.

V. SIMULATION RESULTS & ANALYSIS

a. Throughput is the measure of the number of packets successfully transmitted to their final destination per unit time. It is the ratio between the numbers of sent packets vs. received packets.

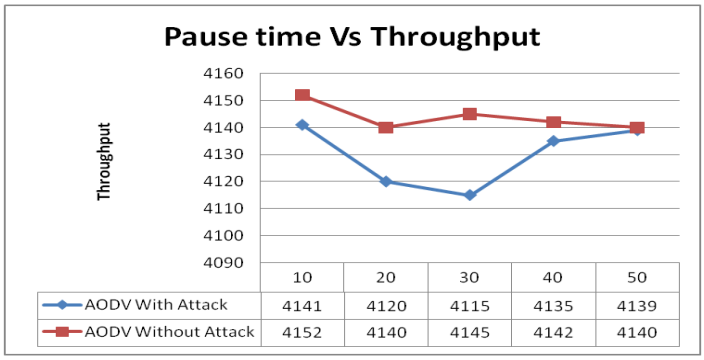


Figure1:- Pause Time Vs Throughput

Figure 1 shows throughput of AODV in presence and without presence of black hole attack with variation of pause time. It is observed that throughput of AODV is rises without presence of attack. It can also be observed that throughput of AODV in both conditions are same at pause time 50s.

b. Average End to End Delay signifies the average time taken by packets to reach one end to another end (Source to Destination).

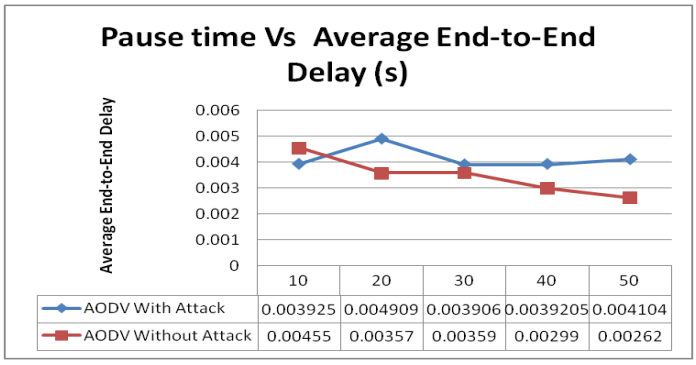


Figure2:- Pause Time Vs Average End-to-End delay

Figure 2 shows end to end delay of AODV in presence and without presence of black hole attack with variation of pause time. It can observe that end to end delay is goes down when AODV works without black hole attack in network. But it end to end delay in presence of black hole at pause time 10 is less then without presence of attack.

c. Total packets received are no. of packets received when sent from source to destination

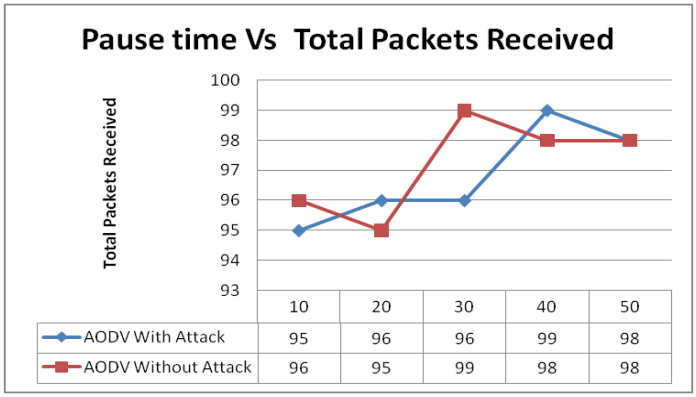


Figure3:- Pause Time Vs Total Packets Received

Figure 3 shows total packet received of AODV in presence and without presence of black hole attack with variation of pause time. It can be observed that performance of AODV without attack performs well. Receiver can receive packet due to better routing technique and route caching. It is also observed that there is fewer packets have received when pause time is 20s. The reason behind it is the signal coverage or mobility of nodes.

d. Average Jitter effect signifies the Packets from the source will reach the destination with different delays. A packet's delay varies with its position in the queues of the routers along the path between source and destination and this position can vary unpredictably.

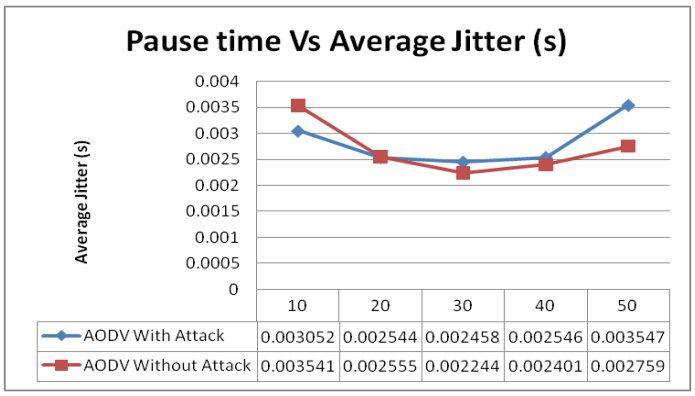


Figure4:- Pause Time Vs Average Jitter

Figure 4 shows average jitter of AODV in presence and without presence of black hole attack with variation of pause time. It is observed that Avg. jitter effect in AODV without attack and AODV with attack changes by increasing or decreasing the pause time. The Jitter effect decreases as the pause time increases. But when it becomes 50s average jitter increases for each protocol.

e. Throughput is the measure of the number of packets successfully transmitted to their final destination per unit time. It is the ratio between the numbers of sent packets vs. received packets.

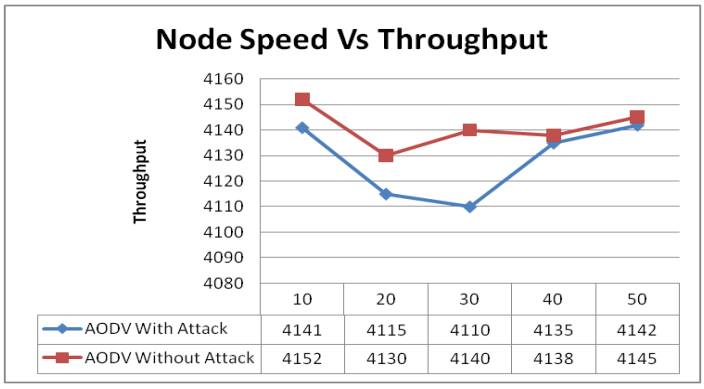


Figure5:- Node Speed Vs Throughput

Figure 5 shows throughput of AODV in presence and without presence of black hole attack with variation of speed of node in network. It can be observed that when node speed is 50m/sec then throughput for AODV without attack is similar to throughput of AODV with attack. It can be observed that throughput of protocol are decreases when nodes in network moving with speed of 20m/sec, and it also varies with different node speeds.

f. Average End to End Delay signifies the average time taken by packets to reach from one end to another end (Source to Destination).

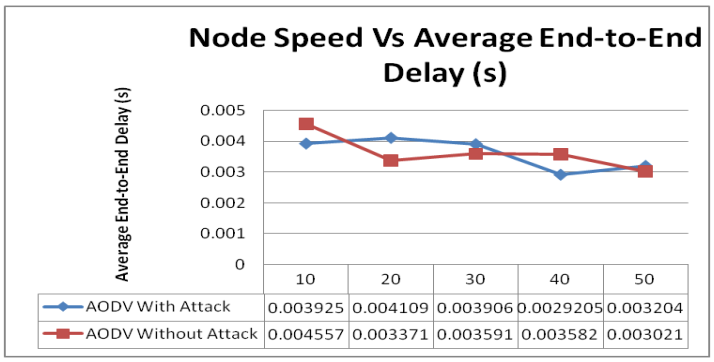


Figure6:- Node Speed Vs Average End-to-End delay

Figure 6 shows average end-to-end delay of AODV in presence and without presence of black hole attack with variation of speed of node in network. It can be observed that the end to end delay in both conditions is varying. AODV can perform in both situation and there is very less effect of node speed in performance.

g. A maximum packet received is the Ratio of received packets that may have been received in the network to the total number of packet sent.

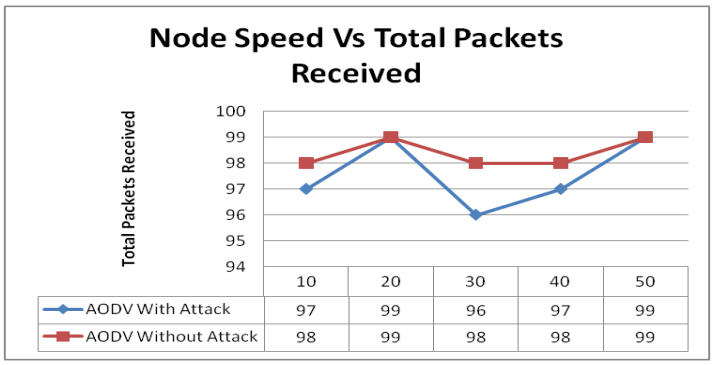


Figure7:- Node Speed Vs Total Packet Received

Figure 7 shows total packet received of AODV in presence and without presence of black hole attack with variation of speed of node in network. It has observed that nodes can receive more packets when network uses AODV without attack. There is much variation in AODV with attack, because nodes in network are move with different speed. Receiver received minimum packets in presence of attack when nodes in network are moving with speed of 30 m/s.

h. Average Jitter effect signifies the Packets from the source will reach the destination with different delays. A packet's delay varies with its position in the queues of the routers along the path between source and destination and this position can vary unpredictably.

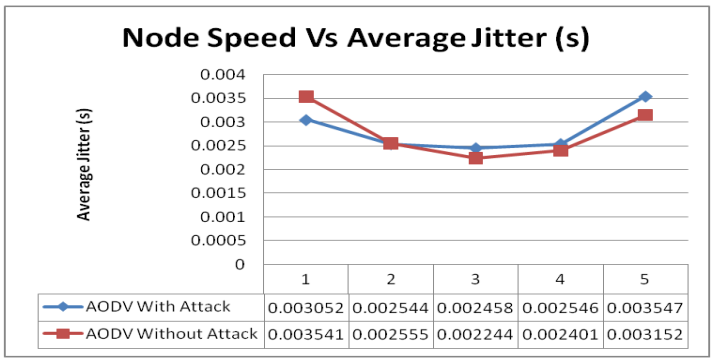


Figure8:- Node Speed Vs Average Jitter

Figure 8 shows average jitter of AODV in presence and without presence of black hole attack with variation of speed of node in network. It can observe that average jitter by AODV is similar in both situations. But most of the time AODV in without attacking situation has less jitter.

VI. CONCLUSION

This paper presents a presents an analysis of AODV routing with and without black hole attack in different scenario in ad hoc network. By different analysis it can be observed in results that AODV can perform better without presence of black hole attack in all situations. If we cannot find similar results as AODV produce without black hole attack than we can predict that there may be an attack on network.

REFERENCES

[1] M. Frodigh, P. Johansson, and P. Larsson. "Wireless ad hoc networking: the art of networking without a network", Ericsson Review, No.4, 2000, pp. 248-263.
[2] G.V.S. Raju and G. Hernandez, "Routing in Ad hoc networks," in proceedings of the IEEE– SMC International Conference, October 2002.
[3] Qualnet Simulator Documentation. "Qualnet 5.0 User`s Guide", Scalable Network Technologies, Inc., Los Angeles, CA 90045, 2006.
[4] A.Boomarani Malany 1, V.R.Sarma Dhulipala 2, and RM.Chandrasekaran 3 "Throughput
and Delay Comparison of MANET Routing Protocols", Int. J. Open Problems Compt. Math.,
Vol. 2, No. 3, September 2009 ISSN 1998-6262; Copyright ©ICSRS Publication, 2009 www.icsrs.
org
[5] Yi-Chun Hu, Adrian Perrig, "A Survey of Secure Wireless Ad Hoc Routing", IEEE Security
and Privacy, 1540- 7993/04/$20.00 © 2004 IEEE, May/June 2004.

[6] Existing MANET Routing Protocols and Metrics used Towards the Efficiency and Reliability- An Overview Shafinaz Buruhanudeen, Mohamed Othman, Mazliza Othman, Borhanuddin Mohd Ali Proceedings of the 2007 IEEE International Conference on Telecommunications and Malaysia International Conference on Communications, 14-17 May 2007, Penang, Malaysia 1-4244-1094-
0/07©2007 IEEE.
[7] Daniel Lang , "On the Evaluation and Classification of Routing Protocols for Mobile Ad Hoc Networks " 2006.
[8] "Security and Privacy in Location Based MANETs/VANETs" OPNETWORK,
www.ics.uci.edu/~keldefra/manet.htm 2005
[9] Julian Hsu Julian Hsu, Sameer Bhatia, Mineo Takai, Rajive Bagrodia Performance of mobile ad hoc networking protocol in realistic scenario 2005
[10] A Performance Comparison of Routing Protocols for Large-Scale Wireless Mobile Ad Hoc Networks Ioannis Broustis Gentian Jakllari Thomas Repantis Mart Molle Department of Computer Science & Engineering University of California, Riverside 2004
[11] Performance Comparison of MANET Routing Protocols in Different Network Sizes Computer Science Project David Oliver Jörg, 2003
[12] C. Cheng , R. Riley , S. P. R. Kumar , J. J. Garcia-Luna-Aceves, " A loop-free extended Bellman-Ford routing protocol without bouncing effect", Symposium proceedings on Communications architectures & protocols, p.224-236, September 25-27, 1989, Austin, Texas,
United States.
[13] Xin Yu, "Distributed Cache Updating for the Dynamic Source Routing Protocol," IEEE Transactions on Mobile Computing, vol. 5, no. 6, pp. 609-626, Jun., 2006
[14] M. K. Marina, S. R. Das "Routing performance in the Presence of Unidirectional Links in Multihop Wireless Networks," Proc. of the 3rd ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC), Jun. 2002
[15] Charles E.Perkins. Ad hoc Networking, Addison-Wedey, 2001
[16] T. S. Rappaport. Wireless Communications: Principles and Practice. Prentice-Hall, 1996.
[17] A. Lindgren, "Infrastructured Ad Hoc Networks," Proc. 2002 Int'l Workshop on Ad Hoc Networking, Vancouver, August 2002.
[18] J. Broch, D. Maltz, D. Johnson, Y.-C. Hu, and J. Jetcheva, "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols," Proc. Fourth ACM MobiCom, pp. 85- 97, 1998.
[19] Satoshi Kurosawa, Hidehisa Nakayama, Nei Kato, Abbas Jamalipour, and Yoshiaki Nemoto, "Detecting Blackhole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method". International Journal of Network Security, Vol.5, No.3, PP.338–346, Nov. 2007
[20] Frank Karg, Stefan Schlott, Andreas Klenk, Alfred Geiss, MichaelWeber, "Securing Ad hoc Routing Protocols", 30th EUROMICRO Conference (EUROMICRO'04), IEEE-2004.
[21] Bhavyesh Divecha, Ajith Abraham, Crina Grosan, Sugata Sanyal "Analysis of Dynamic Source Routing and Destination-Sequenced Distance-Vector Protocols for Different Mobility Models" First Asia International Conference on Modeling & Simulation (AMS'07) : March 2007 pp. 224-229.
[22] D. B. Johnson, D. A. Maltz, Y.-C. Hu and J. G. Jetcheva. "The dynamic source routing protocol for mobile ad-hoc networks". IETF Internet draft, draft-ietf-manet-dsr-04.txt, November 2000.
[23] Performance Comparison of MANET Routing Protocols in Different Network Sizes Computer Science Project David Oliver Jörg, 2003
[24]Rajiv Misra, C.R.Mandal"Performance Comparison of AODV/DSR On-demand Routing Protocols for Ad Hoc Networks in Constrained Situation"0-7803-8964-6/05/$20.00 lEEE 2005
[25] Rama Murti "Wireless Networking" 2008.
[26]D. Djenouri, A. Derhab, and N. Badache. Ad hoc networksrouting protocols and mobility. Int. Arab J. Inf. Technol.3 (2):126–133, 2006
[27]Layuan, Li Chunlin, Yaun Peiyan "Performance evaluation and simulation of routing protocols in ad hoc networks",February 2007, Computer Communication

A Novel Approach of Obtaining Features Using Wavelet Based Image Fusion and Harris Corner Detection

Nilanjan Dey[1], Subhendu Das[2], Pranati Rakshit[3]
[1](Asst. Professor Dept. of IT, JIS College of Engineering, Kalyani, West Bengal) India
[2](M Tech Scholar, Dept. of CSE, JIS College of Engineering, Kalyani, West Bengal) India
[3](HOD Dept. of CSE, JIS College of Engineering, Kalyani, West Bengal) India

ABSTRACT
In this paper we proposed a method of corner detection for obtaining features which is required for tracking and recognizing objects from a fused image. Image fusion is a combination of information gathered from different images which is useful for extraction of more numbers of features from Multibiometric systems, useful for the purpose of biometric recognition and identification. Image fusion is carried out using wavelet based alpha blending technique.

Keywords – **Image fusion, Alpha-Blending, Wavelet, Harris Corner**

I. Introduction
A corner is a point for which there are two dominant and different edge directions in the vicinity of the point. In simpler terms, a corner can be defined as the intersection of two edges, where an edge is a sharp change in image brightness. Generally termed as interest point detection, corner detection is a methodology used within computer vision systems to obtain certain kinds of features from a given image. The initial operator concept of "points of interest" in an image, which could be used to locate matching regions in different images, was developed by Hans P. Moravec in 1977. The Moravec operator is considered to be a corner detector because it defines interest points as points where there are large intensity variations in all directions.

For humans, it is easier to identify a "corner", but a mathematical detection is required in case of algorithms. Chris Harris and Mike Stephens in 1988 improved upon Moravec's corner detector by taking into account the differential of the corner score with respect to direction directly, instead of using shifted patches. Moravec only considered shifts in discrete 45 degree angles whereas Harris considered all directions. Harris detector has proved to be more accurate in distinguishing between edges and corners. He used a circular Gaussian window to reduce noise.

Wavelet Based alpha-blending image fusion technique generates a fused image. Harries Corner detection on fused image [1] gives an effective result for obtaining features, required to track and recognized objects.

A multibiometric system [2] helps to overcome the limitations of the uni-modal biometric systems in the field of biometric recognizing and identifying.

II. Discrete Wavelet Transformation
The wavelet transform describes a multi-resolution decomposition process in terms of expansion of an image onto a set of wavelet basis functions. Discrete Wavelet Transformation has its own excellent space frequency localization property. Applying DWT in 2D images corresponds to 2D filter image processing in each dimension. The input image is divided into 4 non-overlapping multi-resolution sub-bands by the filters, namely LL1 (Approximation coefficients), LH1 (vertical details), HL1 (horizontal details) and HH1 (diagonal details). The sub-band (LL1) is processed further to obtain the next coarser scale of wavelet coefficients, until some final scale "N" is reached. When "N" is reached, we'll have 3N+1 sub-bands consisting of the multi-resolution sub-bands (LLN) and (LHX), (HLX) and (HHX) where "X" ranges from 1 until "N". Generally most of the Image energy is stored in these sub-bands.

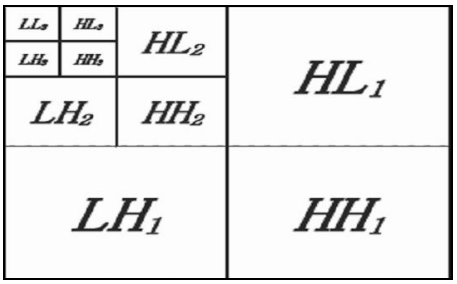


Figure1. Three phase decomposition using DWT.

The Haar wavelet is also the simplest possible wavelet. Haar wavelet is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions.

III. Alpha-Blending Technique
Alpha-Blending [2, 3] is the way of mixing of two images together to form a fused image. Alpha Blending is accomplished in computer graphics by blending each pixel from the first source image with the corresponding pixel in

the second source image. Here's the equation for executing alpha blending

Final pixel = alpha * (First image's source pixel) + (1.0-alpha) * (Second image's source pixel)

The blending factor or percentage of colors from the first source image used in the blended image is called "alpha." The alpha used in algebra is in the range 0.0 to 1.0, instead of 0 to 100%.

According to the formula of the alpha blending the fussed image is given by

FI=alpha*(IM1) + (1.0-alpha)*(IM2) (1)

Where alpha is set as 0.5

RW=Recovered watermark, FI=fussed image, IM1= selected sub-band of the first image, IM2= selected corresponding sub-band of the second Image.

IV. Harris Corner Detection

Harris corner detector [5,6] is based on the local auto-correlation function of a signal which measures the local changes of the signal with patches shifted by a small amount in different directions. Given a shift (x, y) and a point the auto-correlation function is defined as

$$c(x,y) = \sum_W [I(x_i,y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \quad \ldots\ldots (2)$$

Where I (x_i, y_i) represent the image function and (x_i, y_i) are the points in the window W centered on (x, y).

The shifted image is approximated by a Taylor expansion truncated to the first order terms

$$I(x_i + \Delta x, y_i + \Delta y) \approx \left[I(x_i,y_i) + \left[I_x(x_i,y_i)\ \ I_y(x_i,y_i)\right]\right]\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix} \quad \ldots\ldots\ldots\ldots (3)$$

where I_x (x_i, y_i) and I_y (x_i, y_i) indicate the partial derivatives in x and y respectively. With a filter like [-1, 0, 1] and [-1, 0, 1] ^{T}, the partial derivates can be calculated from the image by

Substituting (3) in (2).

$$c(x,y) = [\Delta x \quad \Delta y]\begin{bmatrix}\sum_W (I_x(x_i,y_i))^2 & \sum_W I_x(x_i,y_i)\, I_y(x_i,y_i) \\ \sum_W I_x(x_i,y_i)\, I_y(x_i,y_i) & \sum_W (I_y(x_i,y_i))^2\end{bmatrix}\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix} = [\Delta x \quad \Delta y]C(x,y)\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix}$$

C(x, y) the auto-correlation matrix which captures the intensity structure of the local neighborhood.

Let α_1 and α_2 be the Eigen values of C(x, y), then we have 3 cases to consider:

1. Both Eigen values are small means uniform region (constant intensity).
2. Both eigen values are high means Interest point (corner)
3. One eigen value is high means contour(edge)

To find out the interest points, Characterize corner response H(x, y) by Eigen values of C(x, y).

- C(x, y) is symmetric and positive definite that is α_1 and α_2 are >0
- $\alpha_1 \alpha_2$ = det (C(x, y)) = AC –B^2
  - $\alpha_1 + \alpha_2$ = trace(C(x, y)) = A + C
- Harris suggested: That the $H_{cornerResponse} = \alpha_1 \alpha_2 - 0.04(\alpha_1 + \alpha_2)^2$

Finally, it is needed to find out corner points as local maxima of the corner response.

V. Proposed Method

Step 1. Two images of same size are read and 1-level wavelet decomposition performed for both images.
Step 2. Fused decomposed images using Alpha-Blending technique.
Step 3. Enhanced fused image.
Step 4. Harris corner detection technique applied on the fused image.
Step 5. Extracted corners saved as a feature point for tracking and recognizing objects in the database for matching.

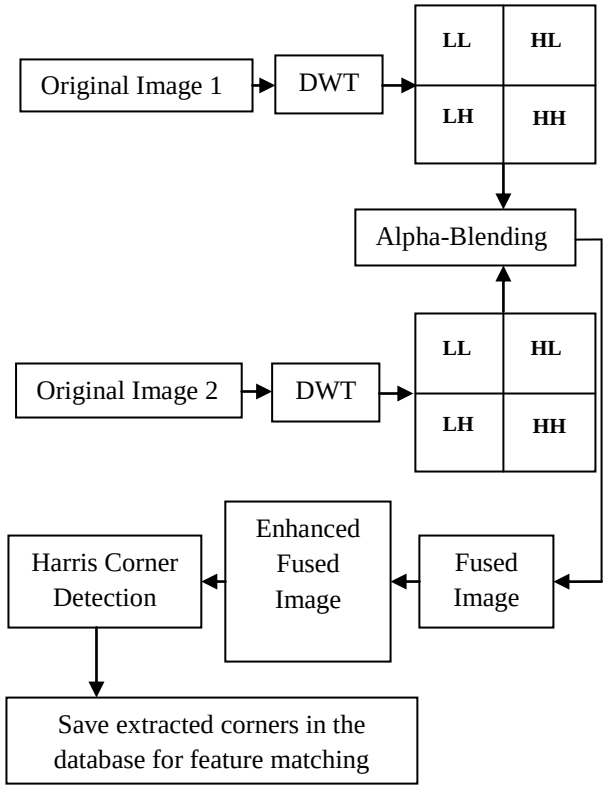
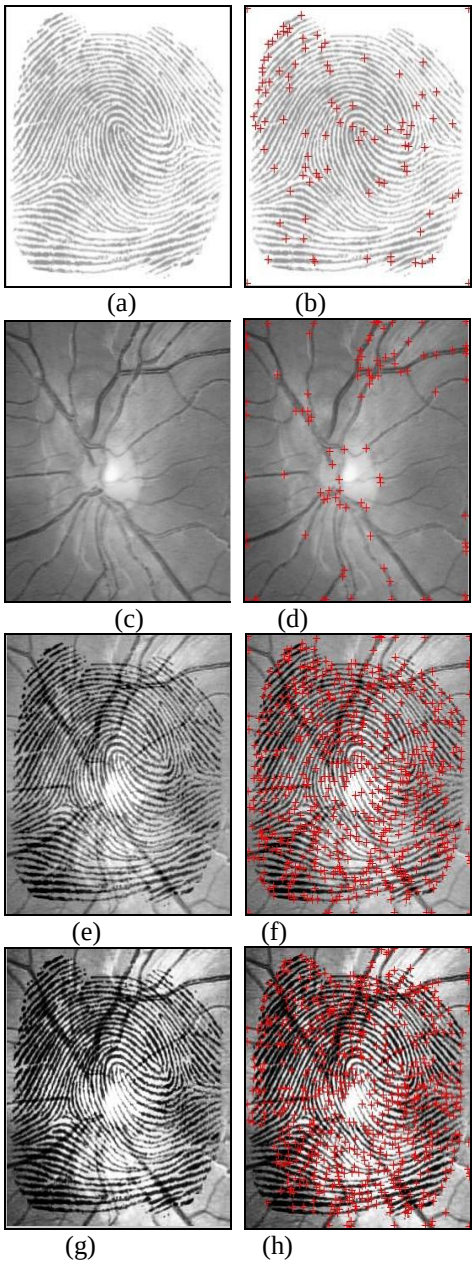


Figure 2.

VI. Result and Discussions



(a)Original Fingerprint (b) Harris Corner Detected Fingerprint image (c) Original Retina Blood Vessel (d) Harris Corner Detected Retina Blood Vessel (e) Limit based Contrast Stretched Fused Image (f) Harris Corner Detected Contrast Stretched Fused Image (g) Histogram Equalized Fused Image (h) Harris Corner Detected Histogram Equalized Fused Image

Figure 3. –Extracted corners using proposed algorithm

Table1

Type	Number of corners found
Harris Corner Detected Fingerprint image	95
Harris Corner Detected Retina Blood Vessel	94
Harris Corner Detected Limit based Contrast Stretch Fused Image	591
Harris Corner Detected Histogram Equalized Fused Image	623

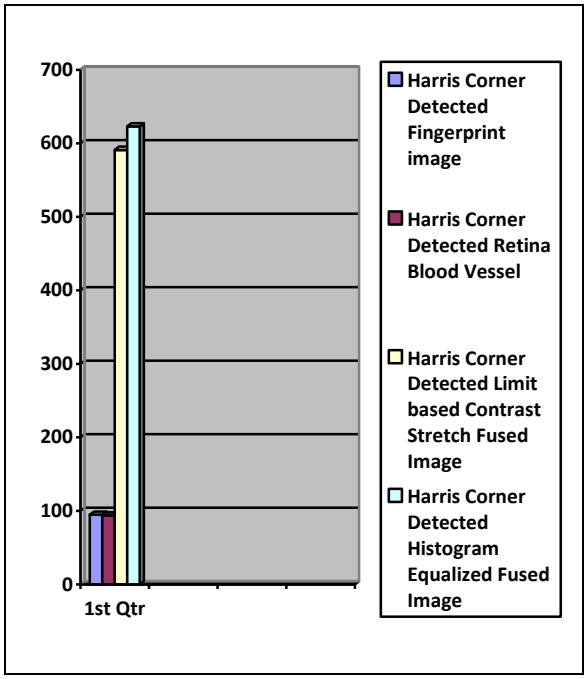


Figure 4 - Graphical representation of extracted corners using proposed algorithm

VII. Conclusion

Harris Corner Detected applied on preprocessed fused image using wavelet decomposition gives a very effective result. The number of corner detected are stored in a database, use for future image processing operations like tracking or recognition of objects.

References

[1] Shekhar Karanwal , Davendra Kumar , Rohit Maurya ,"Fusion of Fingerprint and Face by using DWT and SIFT", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.5, June 2010

[2] A.K.Jain and A. Ross, "Multibiometric systems",Communications of the ACM, vol. 47, no.1, pp. 34 - 40,2004.

[3] Akhil Pratap Shing,Agya Mishra,"Wavelet Based Watermarking on Digital Image",Indian Journal of Computer Science and Engineering,Vol 1 No 2, 86-91

[4] Bo Shen, Iihwar K. Sethi and Vasudev Bhaskaran, (1998) DCT Domain Alpha Blending, IEEE

[5] Harris, C., Stephens, M., 1988, A Combined Corner and Edge Detector, Proceedings of 4th AlveyVision Conference
[6] KonstantinosG. Derpanis, 2004, The Harris Corner Detector

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue.2, pp-400-406 ISSN: 2249-6645

A Theoretical Study on Thermal Conductivity of Zn-Nanofluid

Madan Mohan Ghosh[1], Sudipto Ghosh[2], and Shyamal Kumar Pabi[2]

[1](Department of Metallurgical and Materials Engineering, National Institute of Technology, Durgapur – 713209) India

[2](Department of Metallurgical and Materials Engineering, Indian Institute of Technology, Kharagpur – 721302) India

ABSTRACT

A coupled molecular dynamics (MD)-stochastic model has been developed for predicting the thermal conductivity of a hypothetical water based Zn-nanofluid taking into account the heat transfer during and after collision of the nanoparticles with the heat source. The model has predicted a somewhat lower enhancement in thermal conductivity of water based Zn-nanofluid compared to water based Cu-nanofluid under identical conditions. The theoretical predictions need experimental verification. The model establishes the role of thermal as well as mechanical properties of the nanoparticles in enhancing the thermal conductivity of nanofluids.

Key words: **Molecular dynamics, Nanofluid, Thermal conductivity**

1. Introduction

Nanofluids, i.e. colloidal suspensions of nanoparticles dispersed in liquids, are likely to be the future heat transfer media in advanced heat transfer applications since they have significantly enhanced thermal conductivity compared to that of conventional heat transfer fluids [1-3]. Designing nanofluids for a given practical application requires an accurate theoretical knowledge of the thermal conductivity of nanofluids. However, at present we don't have adequate theoretical understanding on the thermal conductivity of nanofluids [4]. Although some attempts have been made to theorize the thermal conductivity of nanofluids [5-8] it still remains as a subject of hot debate [9]. A generally accepted model which has the capability to explain the thermal conductivity of diverse types of nanofluids is still lacking [10].

The present work aims to theoretically estimate the thermal conductivity of a hypothetical water based Zn-nanofluid. The Zn-nanoparticles suspended in water medium will inevitably undergo Brownian motion, in course of which they will repeatedly collide with the heat source. During the collision rapid conductive heat transfer will take place due to solid-solid contact. Subsequent to the collision, when the nanoparticle undergoes Brownian motion through the bulk fluid it releases the excess heat to the fluid. Therefore, some amount of heat is transferred by conduction mode due to the collision of the suspended nanoparticles with the heat source, in addition to the normal conductive heat transfer through the base fluid itself. The extent of this heat transfer has been estimated with the help of molecular dynamics (MD) simulation in combination with stochastic simulation to compute the enhancement in thermal conductivity of the water based Zn-nanofluid as a function of volume fraction loading of nanoparticles. The scheme of simulation has been presented here. More detailed account of the simulation has been given elsewhere [11]. The thermal conductivity estimated on the basis of the present model for water based Cu-nanofluid has reasonable agreement with the experimental data [11]. However, the predicted thermal conductivity for water based Zn-nanofluid needs to be tested against experimental data. It has been explained with the help of present theoretical model why mechanical properties of nanoparticles are more important than their thermal properties in enhancing the thermal conductivities of nanofluids.

2. Simulation Procedure

In order to theoretically estimate the thermal conductivity of a hypothetical water based Zn-nanofluid first a block shaped Zn-heat source has been generated. A Zn-nanoparticle with diameter 4 nm has also been generated. The Zn-heat source has been equilibrated at 370 K and the Zn-nanoparticle has been equilibrated at temperatures ranging

www.ijmer.com Vol.1, Issue.2, pp-400-406 ISSN: 2249-6645

from 298 K to 358 K using isokinetic thermostat. After equilibration at the predetermined temperatures the Zn-nanoparticle has been directed towards the Zn-heat source with a velocity equal to the collision velocity. The average collision velocity for a Zn-nanoparticle of 4 nm diameter is 4.640 m/s, as estimated from the preliminary stochastic simulation. During collision of the Zn-nanoparticle with the Zn-heat source the temperatures of both the nanoparticle and heat source has been estimated from the average kinetic energy of the atoms containing them. In the MD simulation Lennard-Jones pair potential [12] has been assumed for all interatomic interactions. The time modulation of phase space of atoms of both the nanoparticle and heat source has been generated by velocity-verlet algorithm [12] using a time step size of 10^{-14} s. The period of collision estimated from impact dynamics [13] is 12.91 ps. MD simulations of the collision have been carried out for different initial temperature of the nanoparticle (298-358 K) in order to estimate the temperature rise of nanoparticle due to collision with the heat source.

After the collision the nanoparticle moves through the base fluid by Brownian motion. In the present simulation the Brownian phase space of the particle with progress of time has been evaluated by stochastic simulation [14]. During Brownian motion subsequent to the collision the nanoparticle releases the excess heat to the surrounding fluid as per the rule of flow past a spherical nanoparticle [6]. Based on the characteristic thermal history of the nanoparticles which have been released from different initial distances from the heat source the enhancement in thermal conductivity of the nanofluid for a given volume fraction loading of nanoparticles has been estimated [11].

3. Results and Discussion

Figure 1 shows the temperature variations of the heat source which has been equilibrated at 370 K (Fig. 1(a)) and the nanoparticle which has been equilibrated at 298 K (Fig. 1(b)). It is apparent that at the end of 10000 time steps of 1 fs both the objects get equilibrated at the desired temperatures.

Figure 2 displays the configuration of the nanoparticle and heat source prior to the collision. It is apparent that there is negligible distortion of the objects after equilibration.

The MD simulation of collision of a Zn-nanoparticle (4 nm dia.) with a Zn-heat source has yielded the temperature variations of the nanoparticle and heat source as shown in Fig. 3. Here the initial temperature of the nanoparticle is 298 K and that of the heat source is 370 K, and starting velocity for collision of the nanoparticle with the heat source is 4.640 m/s which is the average collision velocity as estimated from preliminary stochastic simulation for a Zn-nanoparticle of 4 nm diameter undergoing Brownian motion in water. Rapid rise in temperature of the nanoparticle has also been observed here as in the case of Cu-nanoparticle colliding with a Cu-heat source [11]. However, within the collision period of 12.91 ps the Zn-nanoparticle which was initially at a temperature of 298 K did not attain the temperature of the Zn-heat source. Hence, pulse-like heat transfer during the collision is partially effective in this case.

Figure 4 depicts the collision induced temperature rise of the nanoparticle as a function of the initial temperature of the nanoparticle. Here, in all the cases the nanoparticle has been directed towards the heat source with a starting velocity equal to the average collision velocity (4.640 m/s) estimated from preliminary stochastic simulation. It is apparent that the collision induced thermal pick up by the nanoparticle decreases with increasing the pre-collision temperature of the nanoparticle. This is because of the fact that the magnitude of heat flux to the nanoparticle during the collision decreases with decrease in the temperature difference between the heat source and nanoparticle.

The trajectory of a Zn-nanoparticle of 4 nm diameter undergoing Brownian motion in water medium within a time frame of 1 s has been shown in Fig. 5 which is characteristic of Brownian motion of any particle of very small size suspended in a fluid medium. Here the particle has been initially released from a distance of 1 mm from the heat source with zero velocity. It is apparent that during the Brownian motion the nanoparticle repeatedly collides with the heat source. Here the YZ-plane which is passing through origin is the surface of the heat source in contact with the nanofluid.

The characteristic thermal history of a Zn-nanoparticle of 4 nm diameter suspended in a water based nanofluid has been displayed in Fig. 6. This has been evaluated by MD simulation coupled with stochastic simulation. It is apparent that the Zn-nanoparticle (4 nm dia.) which acquires heat from the heat source within ~12 ps during the collision releases the excess heat to the surrounding water medium within 2-3 ms subsequent to the collision.

www.ijmer.com Vol.1, Issue.2, pp-400-406 ISSN: 2249-6645

The enhancement in thermal conductivity as predicted by MD simulation coupled with stochastic simulation for the hypothetical water based Zn-nanofluid has been shown in Fig. 7. The predicted enhancement in thermal conductivity for water based Cu-nanofluid as a function of volume fraction of nanoparticles [11] has also been superimposed in Fig. 7. It shows that at low volume fraction loading of nanoparticles (<0.3 vol.%) the predicted enhancement in thermal conductivity of water based Zn-nanofluid as well as water based-Cu nanofluid varies linearly with volume% loading of nanoparticles. It is also apparent that for a given volume fraction loading the water based Zn- (4 nm dia.) nanofluid shows ~23% lower enhancement in thermal conductivity compared to that of the water based Cu- (4 nm dia.) nanofluid. This is due to much lower thermal conductivity of Zn (113 W m^{-1} K^{-1}) compared to Cu (401 W m^{-1} K^{-1}) [15]. Thus, the present model predicts lower potential of Zn-nanofluid in enhancing the thermal conductivity compared to that of the Cu-nanofluid. These predictions for water based Zn-nanofluids need to be tested against experimental data.

It is to be noted here that the present model considers collision period and thermal conductivity of nanoparticles as important parameters for the enhancement in thermal conductivity of nanofluids. The collision period [13] depends on the density, collision velocity and elastic properties of the nanoparticle. The collision velocity which depends on the Brownian motion parameters has been evaluated based on the preliminary stochastic simulation. The data of thermal conductivity, density, Young's modulus, Poisson's ratio, collision velocity and calculated collision period of Zn, Cu and Ag nanoparticles of 4 nm diameter suspended in respective water based nanofluids have been presented in TABLE 1. It is apparent that Cu and Ag have comparable thermal conductivity (TABLE 1). However, recent studies [16] have revealed 7 times more enhancement in thermal conductivity of water based Ag- (4 nm dia.) nanofluid compared to water based Cu- (4 nm dia.) nanofluid. This is due to ~25% higher collision period [13] of Ag nanoparticle (16.98 ps) compared to the Cu nanoparticle (13.59 ps). On the other hand, the collision period of Zn and Cu are comparable to each other (TABLE 1). However, the thermal conductivity of Zn (113 W m^{-1} K^{-1}) is ~72% less than that of Cu (401 W m^{-1} K^{-1}) [15]. This results in only ~23% lower enhancement in thermal conductivity of water based Zn-nanofluid compared to water based Cu-nanofluid for the same volume fraction loading of nanoparticles. Thus, it appears from the present model that compared to the thermal conductivity of nanoparticles, the collision period which in turn depends on the mechanical properties of the nanoparticles is much more effective in enhancing the thermal conductivity of nanofluids.

4. Conclusions

In conclusion, a model has been developed for predicting the thermal conductivity of water based Zn-nanofluid on the basis of MD simulation coupled with stochastic simulation. In the nanofluid the nanoparticles collide with the heat source repeatedly in course of their Brownian motion through the base fluid. MD simulation has revealed that during the collision rapid heat transfer takes place, although to a less extent compared to Cu-nanofluid. The thermal energy which is acquired by the nanoparticles within ~12 ps during the collision is dissipated to the surrounding water medium within 2-3 ms during subsequent Brownian motion of the nanoparticles through the base fluid. Repeated occurrence of this phenomena results in a significant enhancement in thermal conductivity of the nanofluid. The model predicts a linear variation in the enhancement in thermal conductivity of water based Zn-nanofluid with volume fraction loading of nanoparticles. It is also predicted that for a given volume fraction loading water based Zn-nanofluid would show ~23% lower enhancement in thermal conductivity compared to water based Cu-nanofluid. The theoretical predictions for water based Zn-nanofluids are amenable to experimental verification. The model points out that compared to the thermal properties, the collision period which depends on the mechanical properties of the nanoparticles is a much more effective parameter for the enhancement in thermal conductivity of nanofluids.

www.ijmer.com Vol.1, Issue.2, pp-400-406 ISSN: 2249-6645

REFERENCES

[1] J. A. Eastman, S. U. S. Choi, S. Li, W. Yu, and L. J. Thompson, Anomalously increased effective thermal conductivities of ethylene glycol-based nanofluids containing copper nanoparticles, *Applied Physics Letters, 78(6),* 2001, 718–720.

[2] S. U. S. Choi, Z. G. Zhang, W. Yu, F. E. Lockwood, and E. A. Grulke, Anomalous thermal conductivity enhancement in nanotube suspensions, *Applied Physics Letters, 79(14),* 2001, 2252-2254.

[3] S. Lee, S. U. S. Choi, S. Li, and J. A. Eastman, Measuring thermal conductivity of fluids containing oxide nanoparticles, *Journal of Heat Transfer, 121,* 1999, 280–289.

[4] M. Chandrasekar, S. Suresh, R. Srinivasan, and A. C. Bose, New analytical models to investigate thermal conductivity of nanofluids, *Journal of Nanoscience and Nanotechnology, 9,* 2009, 533-538.

[5] R. K. Shukla, and V. K. Dhir, Study of the effective thermal conductivity of nanofluids, *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*, Orlando, Florida, USA, 2005, 1-5.

[6] S. P. Jang and S. U. S. Choi, Role of Brownian motion in the enhanced thermal conductivity of nanofluids, *Applied Physics Letters, 84(21),* 2004, 4316-4318.

[7] K. C. Leong, C. Yang, and S. M. S. Murshed, A model for the thermal conductivity of nanofluids—the effect of interfacial layer, *Journal of Nanoparticle Research, 8,* 2006, 245-254.

[8] R. Prasher, W. Evans, P. Meakin, J. Fish, P. Phelan, and P. Keblinski, Effect of aggregation on thermal conduction in colloidal nanofluids, *Applied Physics Letters, 89,* 2006, 143119-1-3.

[9] S. U. S. Choi, Nanofluids: from vision to reality through research, *Journal of Heat Transfer, 131,* 2009, 033106-1-9.

[10] X. Q. Wang and A. S. Mujumdar, Heat transfer characteristics of nanofluids: a review, *International Journal of Thermal Sciences, 46,* 2007, 1-19.

[11] M. M. Ghosh, S. Roy, S. K. Pabi, and S. Ghosh, A molecular dynamics-stochastic model for thermal conductivity of nanofluids and its experimental validation, *Journal of Nanoscience and Nanotechnology, 11(3),* 2011, 2196-2207.

[12] J. Li, Basic molecular dynamics, in S. Yip (Ed.), *Handbook of Materials Modeling, Part A* (The Netherlands: Springer, 2005) 565-575.

[13] W. John, G. Reischl, and W. Devor, Charge transfer to metal surfaces from bouncing aerosol particles, *Journal of Aerosol Science, 11,* 1980, 115-138.

[14] E. Nelson, *Dynamical theories of Brownian motion,* 2nd ed. (New Jersey: Princeton University Press, 2001) pp. 46-57.

[15] E. A. Brandes and G. B. Brook (Eds.), *Smithells Metals Reference Book,* 7th edition, (Bodmin, UK: Butterworths, 1992).

[16] M. M. Ghosh, S. Ghosh, and S. K. Pabi, On synthesis of a highly effective and stable silver nanofluid inspired by its multiscale modeling, *Journal of Nanoparticle Research,* under publication.

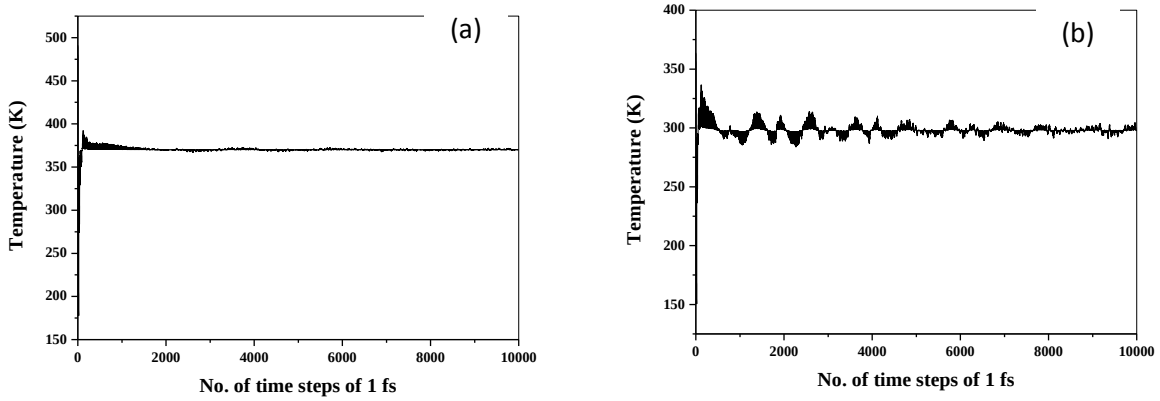


Fig. 1: Temperature variations of a Zn-heat source (a) and a Zn-nanoparticle (b) of 4 nm diameter during equilibration at 370 K and 298 K, respectively.

www.ijmer.com Vol.1, Issue.2, pp-400-406 ISSN: 2249-6645

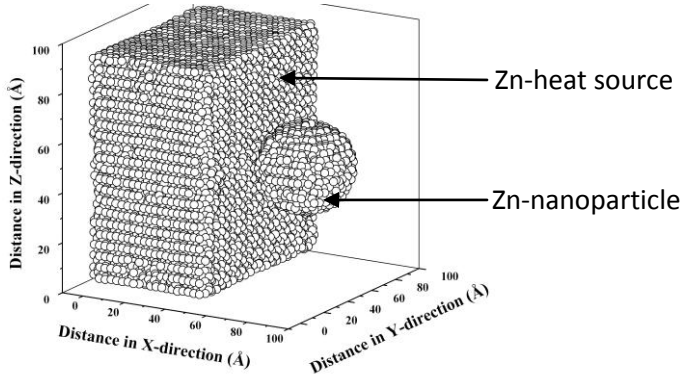


Fig. 2: Configuration of the Zn-nanoparticle (4 nm dia.) and Zn-heat source prior to the collision with each other. The nanoparticle was equilibrated at 298 K and the heat source was equilibrated at 370 K.

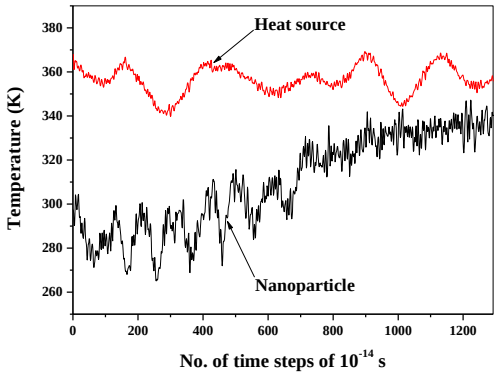


Fig. 3: Temperature variations of a Zn-nanoparticle (4 nm dia.) and a Zn-heat source during collision with each other. The initial temperature of the nanoparticle is 298 K and that of the heat source is 370 K.

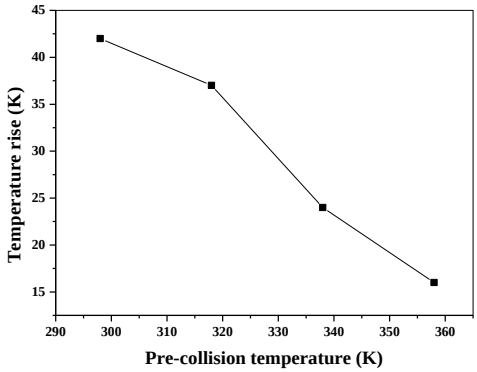


Fig. 4: Temperature rise of a Zn-nanoparticle of 4 nm diameter colliding with a Zn-heat source at a starting velocity of 4.640 m/s, as a function of initial temperature of the nanoparticle.

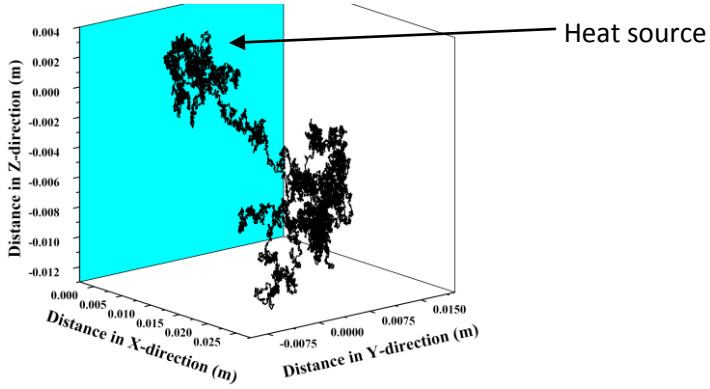


Fig. 5: Trajectory of Brownian motion of a Zn-nanoparticle of 4 nm diameter in water medium within a time frame of 1 s.

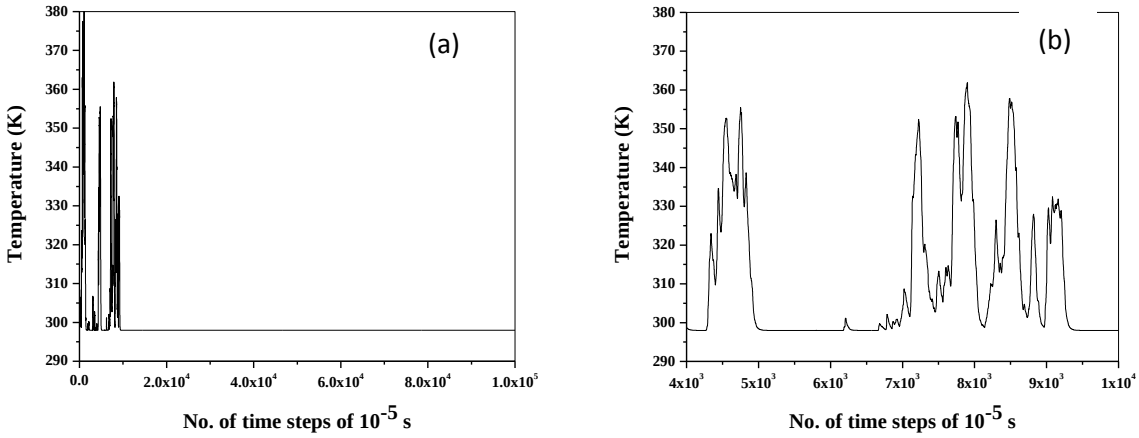


Fig. 6: (a) Simulated thermal profile of a Zn particle of 4 nm diameter moving in water by Brownian motion. Here wider peaks depict multiple collisions with the heat source. (b) A magnified plot of temperature variation of a nanoparticle undergoing multiple collision with heat source in the time interval of 0.04 to 0.10 s in (a).

www.ijmer.com Vol.1, Issue.2, pp-400-406 ISSN: 2249-6645

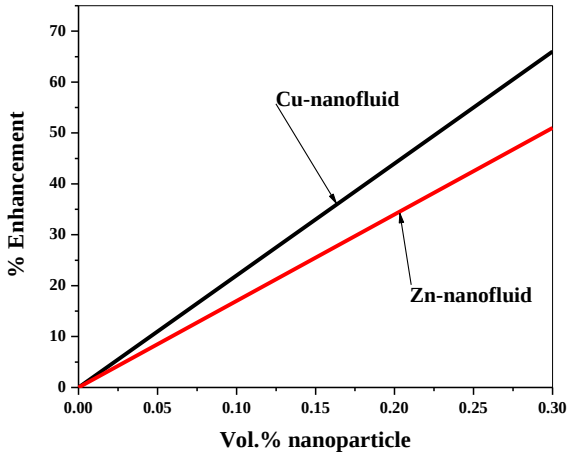


Fig. 7: The enhancement in thermal conductivity of water based Zn- (4 nm dia.) nanofluid, predicted on the basis of MD simulation coupled with stochastic simulation as a function of volume fraction loading of nanoparticles, compared with the predicted enhancement in thermal conductivity of water based Cu- (4 nm dia.) nanofluid.

TABLE 1: Thermal conductivity, density, Young's modulus, Poisson's ratio, collision velocity and collision period of Zn, Cu and Ag nanoparticles of 4 nm diameter.

Nanomaterial	**Thermal conductivity (W/mK)**	**Density (Kg/m^3)**	**Young's modulus (GPa)**	**Poisson's ratio**	**Collision velocity (m/s)**	**Collision period (ps)**
Zn	113	7140	108	0.25	4.640	12.91
Cu	401	8920	120	0.32	4.151	13.59
Ag	429	10500	83	0.37	3.746	16.97

www.ijmer.com Vol.1, Issue.2, pp-407-412 ISSN: 2249-6645

A SOL algorithm and simulation of TCPST for optimal power flow solution using NR method

M. Balasubba Reddy[1], Dr.Y.P. Obulesh[2], Dr.S.Sivanaga Raju[3]
[1](Department of EEE, Prakasam Engineering College, India)
[2](Department of EEE, L.B.R. College of Engineering, India)
[3](Department of EEE, J.N.T.U. College of Engineering, India)

ABSTRACT

This paper introduces the severity of load (SOL) index technique for finding the optimal location of facts devices to achieve the optimal power flow. Objective function in the OPF, that is to be minimised, are overall cost functions which includes the total active power generation cost function. Among various controllers TCPST is considered and optimal location facts device is determined for improved economic dispatch. The OPF constraints are on generators, transmission lines, and TCPST limits. In this paper TCPST for OPF and the achieved improvements are compared with the case where no facts devices are demonstrated.

Keywords **– SOL, TCPST, NR, OPF, FACT**

1. INTRODUCTION

In present days with the deregulation of electricity market, the traditional practices of power system have been completely changed. Better utilization of the existing power system resource to increase capabilities by installing FACTS controllers with economic cost becomes essential [1]. The FACTS devices are capable of changing the system parameters in a fast and effective way. It is known that the benefits brought by FACTS devices include improvement of system stability, enhancement of system reliability, and reduction of operation and transmission investment cost [2].

A few research works were done [3], [4] on the FACTS controllers for improving static performance of the power system. There is also a great need for studying the impact of FACTS controllers and their impact on the power generation cost are also reported [5]. The objective of this paper is to know the real power allocation of generators and to find the best location of FACTS controllers such that overall system cost which includes the minimization of generation cost of power plants and active power loss. Improvements of results with FACTS devices is compared with convention N-R OPF method without FACTS devices.

OPF is a very large, non-linear mathematical programming problem, the main purpose of OPF is to determine the optimal operation state of a power system while meeting some specified constraints. Since the OPF solution was introduced by squires [6], considerable amount of research on different optimization algorithms and solution methods have been done. The main existing techniques for solving the OPF problems are the gradient method, Newton method, linear programming method and decomposition method. Each method has its own advantages and disadvantages, but all of them have their own capabilities for solving the OPF problem [2].

Among the solution methods Newton's method for OPF problem, Newton's method is the most commonly employed. This method requires formulation of Lagrange function combined of objective function with equality and inequality constraints [7]. The flexible AC transmission system is a transmission system which use reliable high speed thyristor based high speed control elements designed based on state of the art developments in power semiconductor devices [8]. The concept of FACTS controllers was first defined by Hingorani in 1988.They are certainly playing an important and major role in the operation and control of modern power system. Facts devices are able to influence and voltages to different degrees depending on the type of device. Typically the devices are divided as shunt connected, series connected and combination of both. The TCPST is series connected device that directly affect the power flows in transmission line to improve power system operation. For OPF control TCPST is used to minimize the total generation fuel cost subject to power balance constraint, real and reactive power generation limits, voltage limits, transmission line limits and FACTS parameter limits. Location of Facts devices in the power system are obtained on the basis of static and dynamic performance [9]. This paper introduces SOL technique for finding the optimal location.

The organization of this paper is as follows. Section 2 introduces OPF without FACTS devices. Modeling of TCPST and problem formulation is described in section 3. The experimental results on the IEEE5 bus and IEEE30 bus systems are presented in section 4. Finally the conclusion and future scope are given.

www.ijmer.com Vol.1, Issue.2, pp-407-412 ISSN: 2249-6645

2. OPF WITHOUT FACTS DEVICES

The objective of active power optimization is to minimize production cost while observing the transmission line and generation active and reactive power limits. The problem can be stated as follows.

$$\text{Minimize} \quad F_T = \sum_{i=1}^{m} C_i(P_{Gi}) \qquad \text{... (1)}$$

$$\text{Subjected to} \quad \sum_{i=1}^{m} P_{Gi} - \sum_{k=1}^{n} P_{Dk} - P_L = 0 \qquad \text{... (2)}$$

$$P_L \le P_L^{\max} \qquad \text{... (3)}$$

$$P_{Gi}^{\min} \le P_{Gi} \le P_{Gi}^{\max} \qquad \text{... (4)}$$

Where n is the number of system buses and m is the number of generating units respectively. $C_i(P_{Gi})$ is production cost of the unit at i^{th} bus, F_T is the total production cost of m generators, P_{Gi} min &P_{Gi} max are minimum and maximum active power limits of the unit at i^{th} bus. P_{Dk} is the active power load at bus k, P_L is the network active power loss, P_l, $P_l^{\max}$ are the active power flow and its limit on line l.

The augmented lagrangian is,

$$L(P_{Gi}) = F_T(P_{Gi}) + \lambda\left(\sum_{k=1}^{n} P_{Dk} + P_L - \sum_{i=1}^{m} P_{Gi}\right) + \sum_{l=1}^{N_l} \mu_l\left(P_l - P_l^{\max}\right) + \sum_{i=1}^{m}\left[\mu_i^{\max}\left(P_{Gi}^{\max} - P_{Gi}\right) + \mu_i^{\max}\left(P_{Gi} - P_{Gi}^{\max}\right)\right] \qquad \text{... (5)}$$

λ is for power balance equation.

$\mu_i^{\min}$ and $\mu_i^{\max}$ are lower and upper active power limits of unit at i^{th} bus.

μ_l is for active power flow limit on line l.

N_l is the number of transmission line flow violations.

3. MODELING OF TCPST

The structure of a TCPST is given in Fig.1. The shunt connected transformer draws power from the network and provides it to the series connected transformer in order to introduce a voltage VT at the series branch. Compared to conventional phase shifting transformers, the mechanical tap changer is replaced by a thyristor controlled equivalent. The purpose of the TCPST is to control the power flow by shifting the transmission angle.

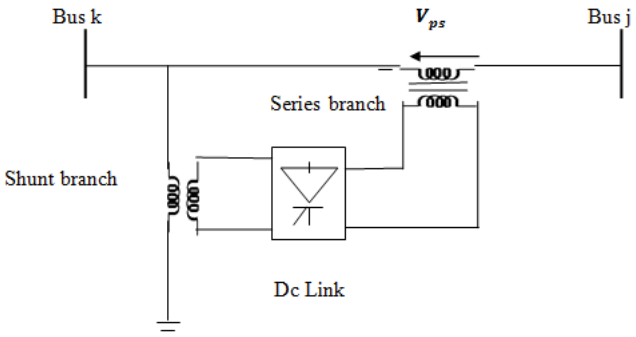


Fig.1 Structure of TCPST

A TCPST model used is given in Fig. 2 where the TCPST corresponds to a variable voltage source with a fixed angle of 90° with respect to the primary voltage. The manipulated variable is the phase shift δ which is determined by the magnitude of the inserted voltage V_{ps}.

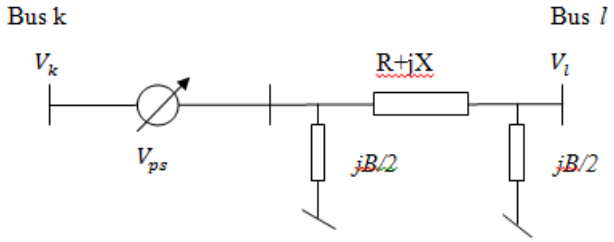


Fig.2 Basic model of TCPST

It is assumed that the device is lossless. Thus, the relationship between the primary and the secondary voltage i.e, where the magnitude of the inserted voltage is determined from the phase shift by,

$$V_k = V_k + V_T \qquad \text{... (6)}$$

$$V_k e^{j\theta_k} = V_k e^{j\theta_k} + V_T e^{j(\theta_k - 90^0)} \qquad \text{... (7)}$$

$$V_T = V_K \tan\delta \qquad \text{... (8)}$$

The OPF uses Newton's method as its optimization engine, enabling an OPF phase-shifter model that is both flexible and robust towards convergence. It can be set to simulate a wide range of operating modes with ease. The power flow equations as provide the starting point for the derivation of the phase-shifter OPF formulation.

$$P_k = V_k{}^2 G - V_k V_m \left[G\,Cos(\theta_k - \theta_m - \phi) + B\,Sin(\theta_k - \theta_m - \phi) \right] \quad \text{... (9)}$$

$$Q_k = -V_k{}^2 B - V_k V_m \left[G\,Sin(\theta_k - \theta_m - \phi) - B\,Cos(\theta_k - \theta_m - \phi) \right] \quad \text{... (10)}$$

$$P_m = V_m{}^2 G - V_m V_k \left[G\,Cos(\theta_m - \theta_k - \phi) + B\,Sin(\theta_m - \theta_k - \phi) \right] \quad \text{... (11)}$$

$$Q_m = -V_m{}^2 B - V_m V_k \left[G\,Sin(\theta_k - \theta_m - \phi) + B\,Cos(\theta_k - \theta_m - \phi) \right] \quad \text{... (12)}$$

Based on the circuit theory, the injection equivalent model of the phase shifter can be obtained. Then by considering the phase shifter into the transmission line the injected powers can be written as,

$$P_{ks} = -V_i{}^2 G_{ij} \tan^2 \phi - V_k V_l \tan\phi \left[G_{kl}\,Sin\delta_{kl} - B_{kl}\,Cos\,\delta_{kl} \right] \quad \text{... (13)}$$

$$P_{ls} = -V_k V_l \tan\phi \left[G_{kl} Sin\delta_{kl} + B_{kl} Cos\delta_{kl} \right] \quad \text{... (14)}$$

Hence to calculate the distribution factors, dc load flow is used. Therefore the above equations can be simplified as,

$$P_{ks} = \tan\phi\, B_{kl}\, Cos\delta_{kl} \quad \text{... (15)}$$

$$P_{ks} = -\tan\phi\, B_{kl}\, Cos\,\delta_{kl} \quad \text{... (16)}$$

4. OPF formulation with TCPST Lagrangian Function

The main aim of the optimization algorithm described in this chapter is to minimize the active power generation cost in the power system by adjusting suitable controllable parameters. For a phase-shifter model with phase-shifting facilities in the primary winding, the Lagrangian function may be expressed by,

$$L(x,\lambda) = f(P_g) + \lambda^t h(P_g, V, \theta, \phi_t) \quad \text{... (17)}$$

In this expression, $f(P_g)$ is the objective function which is to be optimize, term $h(P_g, V, \theta, \phi_t)$ represents the power flow equations; *x* is the vector of state variables, *k* is the vector of Lagrange multipliers for equality constraints; and P_g, V, θ, and ϕ_t are the active power generation, voltage magnitude, voltage phase angle, and phase-shifter angle for tapping position *t*, respectively. The inequality constraints, $h(P_g, V, \theta, \phi_t) < 0$, are not shown because they are included only when variables are outside limits. The Lagrangian function of the power flow mismatch equations at buses *k* and *m* is incorporated into the OPF formulation as an equality constraint, given by the following equation

$$L_{km}(x,\lambda) = \lambda_{pk}(P_k + P_{dk} - P_{gk}) + \lambda_{qk}(Q_k + Q_{dk} - Q_{gk}) + \lambda_{pm}(P_m + P_{dm} - P_{gm}) + \lambda_{qm}(Q_m + Q_{dm} - Q_{gm}) \quad \text{...(18)}$$

In this expression, P_{dk}, P_{dm}, Q_{dk}, *and* Q_{dm} are the active and reactive power loads at buses k and m; P_{gk}, P_{gm}, *and* Q_{gk}, Q_{gm} are the scheduled active and reactive power generations at buses *k* and *m*; and $\lambda_{pk}, \lambda_{pm}, \lambda_{qk}$ *and* λ_{qm} are Lagrange multipliers for active and reactive powers at buses *k* and *m*. A key function of the phase-shifting transformer is to regulate the amount of active power that flows through it, say P_{km}. In the OPF formulation this operating condition is expressed as an equality constraint, represented by the following Lagrangian function,

$$L_{flow}(x,\lambda) = L_{km}(P_{km} - P_{specified}) \quad \text{… (19)}$$

In this expression, $\lambda_{flow-km}$ is the Lagrange multiplier associated with the active power flowing from bus *k* to bus *m*; $P_{specified}$ is the required amount of active power flow through the phase-shifter transformer. The overall Lagrangian function of the phase shifter, encompassing the individual contributions is,

$$L_{ps}(x,\lambda) = \lambda_{flow-km}(P_{km} - P_{specified}) \quad \text{… (20)}$$

4.1 Linearized System of Equations

Representation of the phase-shifting transformer in the OPF algorithm requires that matrix W be augmented by one row and one column, with ϕ_t becoming the state variable. Furthermore, if the phase shifter is set to control active power flow then the dimension of matrix W is increased further by one row and one column. Hence, for each phase shifter involved in the OPF solution the dimension of W is increased by up to two rows and columns, depending on operational requirements. If the two-winding transformer has phase-shifting facilities in the primary winding, the linearized system of equations for minimizing the Lagrangian function using Newton's method is

$$\begin{bmatrix} W_{kk} & W_{km} & W_{k\phi} \\ W_{mk} & W_{mm} & W_{m\phi} \\ W_{\phi k} & W_{\phi m} & W_{\phi\phi} \end{bmatrix} \begin{bmatrix} \Delta z_k \\ \Delta z_m \\ \Delta z_\phi \end{bmatrix} = - \begin{bmatrix} g_k \\ g_m \\ g_\phi \end{bmatrix} \quad \text{... (21)}$$

In this expression, the structure of matrix and vector terms $W_{kk}, W_{km}, W_{mk}, W_{mm}, W_{\phi k}, W_{\phi m}, W_{k\phi}, W_{m\phi}$ *and* $W_{\phi\phi}$ is given by Eqns. (22)–(28), respectively.

www.ijmer.com Vol.1, Issue.2, pp-407-412 ISSN: 2249-6645

$$W_{kk} = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_k^2} & \frac{\partial^2 L}{\partial \theta_k \partial V_k} & \frac{\partial P_k}{\partial \theta_k} & \frac{\partial Q_k}{\partial \theta_k} \\ \frac{\partial^2 L}{\partial V_k \partial \theta_k} & \frac{\partial^2 L}{\partial \theta_k \partial V_k} & \frac{\partial P_k}{\partial V_k} & \frac{\partial Q_k}{\partial V_k} \\ \frac{\partial P_k}{\partial \theta_k} & \frac{\partial P_k}{\partial V_k} & 0 & 0 \\ \frac{\partial Q_k}{\partial \theta_k} & \frac{\partial Q_k}{\partial V_k} & 0 & 0 \end{bmatrix} \quad \dots (22)$$

$$W_{mk} = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_m \partial \theta_k} & \frac{\partial^2 L}{\partial \theta_m \partial V_k} & \frac{\partial P_k}{\partial \theta_m} & \frac{\partial Q_k}{\partial \theta_m} \\ \frac{\partial^2 L}{\partial V_m \partial \theta_k} & \frac{\partial^2 L}{\partial V_m \partial V_k} & \frac{\partial P_k}{\partial V_m} & \frac{\partial Q_k}{\partial V_{mk}} \\ \frac{\partial P_k}{\partial \theta_k} & \frac{\partial P_k}{\partial V_k} & 0 & 0 \\ \frac{\partial Q_k}{\partial \theta_k} & \frac{\partial Q_k}{\partial V_k} & 0 & 0 \end{bmatrix} \quad \dots (23)$$

$$W_{mk} = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_m{}^2} & \frac{\partial^2 L}{\partial \theta_m \partial V_m} & \frac{\partial P_m}{\partial \theta_m} & \frac{\partial Q_m}{\partial \theta_m} \\ \frac{\partial^2 L}{\partial V_m \partial \theta_m} & \frac{\partial^2 L}{\partial V_m{}^2} & \frac{\partial P_m}{\partial V_m} & \frac{\partial Q_m}{\partial V_m} \\ \frac{\partial P_m}{\partial \theta_m} & \frac{\partial P_m}{\partial V_m} & 0 & 0 \\ \frac{\partial Q_m}{\partial \theta_m} & \frac{\partial Q_m}{\partial V_m} & 0 & 0 \end{bmatrix} \quad \dots (24)$$

$$\Delta z_k = \begin{bmatrix} \Delta\theta_k & \Delta V_k & \Delta\lambda_{pk} & \Delta\lambda_{qk} \end{bmatrix}$$

$$\Delta z_m = \begin{bmatrix} \Delta\theta_m & \Delta V_m & \Delta\lambda_{pm} & \Delta\lambda_{qm} \end{bmatrix}$$

$$g_k = \begin{bmatrix} \Delta\theta_m & \Delta V_m & \Delta\lambda_{pk} & \Delta\lambda_{qk} \end{bmatrix}^t \quad \dots (25)$$

$$g_m = \begin{bmatrix} \Delta\theta_m & \Delta V_m & \Delta\lambda_{pm,} & \Delta\lambda_{qm} \end{bmatrix}^t$$

The additional matrix terms in Eqn. (17) reflect the contribution of $\emptyset_t$, the phase shifter state variable. These terms are given explicitly by,

$$W_{k\phi}{}^t = W_{\phi k} = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_k \partial \phi_t} & \frac{\partial^2 L}{\partial V_k \partial \phi_t} & \frac{\partial P_k}{\partial \phi_t} & \frac{\partial Q_k}{\partial \phi_t} \\ \frac{\partial^2 L}{\partial \theta_k \partial \lambda_\phi} & \frac{\partial V_k}{\partial \lambda_\phi} & 0 & 0 \end{bmatrix} \quad \dots (26)$$

$$W_{m\phi}{}^t = W_{\phi m} = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_m \partial \phi_t} & \frac{\partial^2 L}{\partial V_m \partial \phi_t} & \frac{\partial P_m}{\partial \phi_t} & \frac{\partial Q_m}{\partial \phi_t} \\ \frac{\partial^2 L}{\partial \theta_m \partial \lambda_\phi} & \frac{\partial^2 L}{\partial V_m \partial \lambda_\phi} & 0 & 0 \end{bmatrix} \quad \dots (27)$$

$$W_{\phi\phi} = \begin{bmatrix} \frac{\partial^2 L}{\partial_{\phi t}{}^2} & \frac{\partial^2 L}{\partial_{\phi t} \partial_{\lambda\phi}} \\ \frac{\partial^2 L}{\partial_{\lambda\phi} \partial_{\phi t}} & 0 \end{bmatrix} \quad \dots (28)$$

$$\Delta Z_\phi = \begin{bmatrix} \Delta\phi_t & \Delta\lambda_\phi \end{bmatrix}^t \quad \dots (29)$$

$$g_\phi = \begin{bmatrix} \Delta\phi_t & \Delta\lambda_\phi \end{bmatrix}^t \quad \dots (30)$$

If the phase-shifting mechanism is on the secondary winding rather than the primary winding, the state variable ϕ_u replaces ϕ_t in Eqns. (26) – (30). It is noted that the first and second partial derivatives for the various entries in Eqn. (21) are derived from the Lagrangian function of Eqn. (17), The derivative terms corresponding to inequality constraints are entered into matrix only if limits are enforced as a result of one or more state variables having violated limits.

The procedure described by Eqns. (9) – (25) corresponds to a situation where the phase shifter is set to control active power flowing from buses k to m, which is the phase shifter standard control mode. However, in OPF solutions the phase shifter variables are normally adjusted automatically during the solution process in order to reach the best operating point of the electrical power system. In such a situation, the phase shifter is not set to control a fixed amount of active power flowing from buses k to m, and matrix W is suitably modified to reflect this operating condition.

The initial conditions given to all variables involved in the study impact significantly the convergence pattern. Experience has shown that the phase-shifter model is very robust towards convergence when the phase-shifting angle is initialized at 0^0. State variables are initialized similarly to the power flow problem (i.e.1 p.u. voltage magnitude and 0^0 voltage angle for all buses). The Lagrange multiplier for the power flow constraint, $\lambda_{flow-km}$, is set to zero. These values enable very robust iterative solutions.

4.2 Optimal setting of TCPST Parameters

The voltage angle between the sending and receiving end of the transmission line can be regulated by TCPST. It is modelled as a series compensation voltage $U_{FACTS} = \Delta U_{TCPST}$ which is perpendicular to the bus voltage i.e. $V_i \angle 90^0$. According to the model of the FACTS devices, the rated values (RV) of each FACTS device is converted into the real compensation as follows:

The working range of the TCPST is between the −5 degrees to +5 degrees.

$$\phi_{TCPST} = \text{RV} \times 5(\text{degree}) \quad \dots (31)$$

The cost of a TCPST is more related to the operating voltage and the current rating of the circuit concerned. Thus, once the TCPST is installed, the cost is fixed and the cost function can be expressed as follows,

$$C_{TCPST} = d * P_{\max} + IC\ (RS) \quad \dots (32)$$

where,

d is a positive constant representing the capital cost

IC is the installation costs of the TCPST.

P_{max} is the thermal limit of the transmission line where TCPST is to be installed.

The unit for generation cost is *Rs/Hour* and for the investment costs of FACTS devices are *Rs*. They must be

unified into *Rs/Hour*. Normally, the FACTS devices will be in-service for many years. However, only a part of its lifetime is employed to regulate the power flow. In this proposed work, 5 years is applied to evaluate the cost function. Therefore the average value of the investment costs is calculated using the following equation

$$C_1(f) = \frac{C(f)}{8760 \times 5} \; Rs/hr \qquad \ldots (33)$$

where, C (f) is the total investment costs of FACTS devices

5. SEVERITY OF OVER LOADABILITY INDEX (SOL) COMPUTATION

The location of the FACTS devices in this work is decided based on the severity of the overloading of that particular branch in which the device is incorporated. The process of ranking the branches based on their load ability in the order of their severity involves the following steps.

Step1: Establish the criterion to be considered in formulating the ranking

Step2: For the criterion established in (Step 1), define a scalar mathematical function which has a large value of branch load that which stress the system relative to that criterion, and a small value for those which do not; this function is called a "SOL index."

The SOL index is such that contingencies resulting in system conditions yielding large valued over load indices are considered more severe than system conditions with smaller over load indices. In the overload ranker, the SOL index is defined as,

$$SOL = \sum_{i=1}^{n} \left(\frac{Pi}{Pi,\max} \right)^2 \qquad \ldots (34)$$

where,

P_i is the real power flow in line "i",

$P_{i,\max}$ is the maximum of active power transfer over the i^{th} line and

'n' is the set of monitored lines contributing to SOL.

5.1 Calculation of SOL for IEEE 5 Bus system

Table 1: SOL index of all buses by running the general OPF for IEEE 5 bus system

Bus No./Node No.	SOL index of each bus	Ranking
[3]	**0.5812**	**1**
[4]	0.5310	2
[5]	0.3285	3

As compared the above SOL-indices for the IEEE 5 bus system among the 3 load buses (3, 4, 5) the bus 3 is having the maximum SOL index, it is considered to be the critical bus. Hence line indices will provide accurate information with regard to the stability condition of the lines.

5.2 Calculation of SOL for IEEE 30 Bus system

Table 2: SOL-indices by running the general OPF of maximum loaded buses in IEEE 30 bus system

Branch Number	SOL indices of different branches	Ranking
[30]	0.7776	2
[24]	0.5672	4
[29]	**0.8873**	**1**
[28]	0.7486	3
[26]	0.5491	5

As we considered the SOL-index table of the IEEE 30 bus system there will be the 5 load buses (24, 26, 28, 29, 30) with the bus (29) is having the maximum load ability, it is considered to be the critical bus. The branch connected to that particular weakest or critical bus will be the optimal location for the FACTS device to be placed. Hence the branch [29]-[30] is chosen to be the optimal location in the IEEE 30 bus case.

6. RESULT ANALYSIS

The IEEE 5-bus test system is taken illustrate the use of the optimal power flow Newton– Raphson method and is also used to illustrate the use of the OPF with TCPST and associated data. Comparison of line flows in NR OPF without facts device and OPF with TCPST are given in Table 3 to Table 6.

6.1 IEEE 5- Bus Systems

Table 3: Nodal parameters for the IEEE 5 -bus system without FACTS devices

Parameters	Bus Number				
	1	2	3	4	5
Voltage magnitudes	1.1096	1.1000	1.0784	1.0779	1.0726
Phase angles	0.00	-1.33	-3.64	-3.83	-4.46
λ_p (Rs/MW/hr)	179.08	177.83	177.37.9	173.33	169.73

Table 4: Nodal voltages in the five-bus network withTCPST

(With active power flow regulation)

Parameter	Bus Number					
	1	2	3	4	5	6
Voltage magnitude (p.u.)	1.109	1.100	1.076	1.079	1.073	1.079
Phase angle (deg)	0.000	-1.193	-4.098	-3.102	-4.097	-2.705
λ_p (Rs/hr)	175.64	178.54	176.44	178.54	175.64	169.84

Table 5: Phase-shifter angles in the five-bus test system

No. of Iterations	Phase angle settings [$\emptyset_t$ (deg)]	
	(No active power control)	(Active power control)
0	0.000	0.000
1	-0.325	-1.874
2	-0.363	-2.122
3	-0.346	-2.009
4	-0.346	-2.010

Table 6: Active power generation cost without and with Facts device

Quantity	NR based OPF method Without Facts device	NR based OPF method with TCPST
Active power generation cost (Rs/hr)	34,046.325	33,674.85
Active power loss (MW)	3.55	3.03
Active power generation (MW)	168.04	168.05

6.2 IEEE 30-BUS SYSTEM

By comparing the SOL-index under normal situation the optimal location of the FACTS device is decided. Hence for the IEEE 30 Bus system (28-29) is the optimally decided branches for the FACTS devices to be incorporated in the electrical power system.

Table 7: The active power and reactive power for IEEE 30 bus system

Case type	Active power loss (MW)	Reactive power loss (MVAR)
Without FACTS Device	18.58	52.73
With TCPST	18.49	46.41

Table 8: The initial and final costs of active power generation in IEEE 30 bus system

Case type	IEEE 5 bus system		IEEE 30 bus system	
	$P_{Gen,\ initial\ cost}$ (Rs/hr)	$P_{Gen,\ final\ cost}$ (Rs/hr)	$P_{Gen,\ initial\ cost}$ (Rs/hr)	$P_{Gen,\ final\ cost}$ (Rs/hr)
Without FACTS Device	35,000	34,046	36,900	36,765
With TCPST	35,000	33,874	36,900	35,325

From the above sections 6.1 and 6.2 it is observed that the generation cost is reduced to *680 Rs /hr* in IEEE5 bus system and *3825 Rs/hr* in IEEE30 bus system respectively when compared to TCPST, and with the base case i.e. 4.66% reduction in the active power generation cost compared to (2.76%, and 3.25%) without FACTS and with TCPST.

7. CONCLUSION

In this paper SOL technique is effectively and successfully implemented to minimize the operating cost in OPF control with TCPST. The SOL approach achieves better solutions. The thyristor firing angle, a newly introduced state variable in OPF formulations, is combined with nodal voltage magnitudes and angles of the power network in a single frame of reference unified iterative solutions via newton's method. In this firing angle, the thyristor firing angle is regulated in order to achieve an optimal level of compensation under either condition, constrained or unconstrained power flow across the compensated branch. From the above results it is evident that there will be an active power loss reduction by OPF NR method with TCPST will be 14.08 % more compared to NR OPF method and also the active power generation cost decreased by 371.475 Rs/hr by the use of TCPST. The work carried out in this paper can be extended to reduce active power loss and to improve system stability by using various FACTS devices further.

References:

[1]. S.Gerbex, R.Cherkaouiand A.J.Germond," Optimal location of multiple type Facts devices in a power system by means of genetic algorithm," IEEE Trans. Powersystem, Vol.16, pp.537-544, August 2001.

[2] Tjing Tlie and Wanhong Deng, "Optimal flexible Ac transmission systems (FACTS) devices allocation," Electric power and energy systems,Vol .19.No.2.pp.125-134.1997.

[3]. X.Duan, J.Chen F.peng, Y.Luo,Y.Huang, "Power flow control with facts devices," IEEE Trans.Power systems, pp.1585-1589,2000.

[4]. L.Gyugyi, C.Dsehauder, S.L.Williams, Etai., "The unified power flow controller: A new approach to power transmission control," IEEE Trans Power delivery, Vol.10,No.2, pp.1085-1097,1995.

[5]. S.Gerbex, R.Cherkaouiand A.J.Germond," Optimal placement of facts controller in power system by genetic based algorithem," IEEE international conference on power electronics and drive and systems, Hongkong, 1999.

[6]. Squires R.B,1961. Economic dispatch of generation directly from power system voltage and adnittances, IEEE trans on Pas-79 (3):1235-1244.

[7]. Seyed Abbas Tehar and seyed mohammad Hadi Tabei," A multi objective HPSO algorithm approach for optimally location of UPFC in deregulated power systems," American journal of applied sciences 5 (7):835-843,2008.

[8]. Abdel-moanen M.A Narayana Prasad Padhy," optimal power flow incorporating FACTS devices-Bibilography and survey," IEEE 2003.

[9]. S,N,Singh, A.K.David," Optimal location of facts devices for congestion management," Electric power system research 58 (2001) 71-79.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-413-417 ISSN: 2249-6645

Removal of Artifacts in Multi-channel Visual Evoked Potentials

V. Adinarayana Reddy[1], P. Chandra Sekhar Reddy [2], G. Hemalatha [3], T. Jaya Chandra Prasad[4]

[1]RVPECW, Cuddapah, [2]JNTUH, Hyderabad, [3]KSRMCE, Cuddapah, [4]RGMCET, Nandyal,

***Abstract*— The primary goal of this work is to introduce temporal artifact detection strategy to remove artifacts in multichannel evoked potentials. An artifact is defined as any signal that may lead to inaccurate classifier parameter estimation. Temporal domain artifact detection tests include: a standard deviation (STD) test that can detect signals with little or abnormal variations in each channel, a clipping (CL) test detect amplitude clipped EPs in each channel and a kurtosis (KU) test to detect unusual signals that are not identified by STD and CL tests. An attempt has been made to apply these techniques to 14-channel visual evoked potentials (VEPs) obtained from four different subjects.**

Keywords **– evoked potentials, standard deviation , clip, kurtosis.**

I. INTRODUCTION

Evoked potentials (EPs) are event related potentials (ERPs) superimposed in electro-encephalogram (EEG). Evoked potentials are usually considered as the time locked and synchronized activity of a group of neurons that add to the background EEG. Evoked Potentials indicate how well the brain is processing stimuli from the sense organs (eg. eyes, ears or skin) and can help diagnose illnesses.

An evoked potential (EP) is a signal that is generated as a result of the transmission of information induced by the application of a sensory stimulus to a sensory pathway. Examples of such stimuli are electric stimuli, visual stimuli, and auditory stimuli [1]. The application of a stimulus invokes a sequence of action potentials that is transmitted via a nervous pathway to the central nervous system (CNS).

The activation of different parts in the nervous pathway leads to variations in the electromagnetic field that can be recorded on the scalp. Using surface electrodes a sequence of positive and negative peaks can be recorded; such a sequence is called a sensory evoked potential. These peaks are characterized by their amplitude and time after the stimulus, at which they occur the (post stimulus) latency. Evoked potentials are simultaneously recorded on the scalp with the spontaneous EEG.

The EEG signal has much larger amplitude than the evoked potential. Averaging techniques are used to extract the signal related to the stimulus and reduce the amplitude of the ongoing EEG signal.

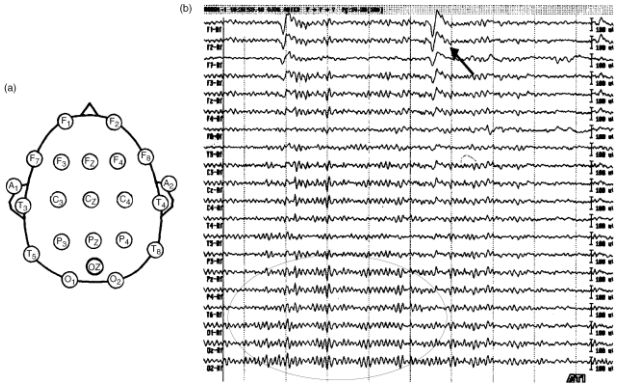


Fig. 1. The M single channel EPs in response to stimulus c.

Evoked potentials are used extensively in the study of human brain functions and in clinical investigations to study normal and abnormal brain functions. They are used to test conduction in the visual, auditory, and somatosensory systems. During surgery they can be used to monitor the condition of structures at the operative site. Fig.1. shows M single channel evoked potentials in response to stimulus c.

Sensory evoked potentials can also be used for monitoring effects of anesthetics on the central nervous system (CNS). The choice of stimulus type to be used depends on the part of the nervous system to be investigated and the circumstances under which measurements are to be made.

We define artifacts as patterns in the training set that lead to inaccurate estimation of classifier parameters and patterns in the test set that yield misleading performance evaluations. In real time classification, such artifacts can give inaccurate test results which can have serious consequences, such as inaccurate diagnosis in clinical evaluations [2].

Visual evoked potentials are very useful in detecting blindness in patients those cannot communicate, such as babies or animals. If repeated stimulation of the visual field causes no changes in EEG potentials then the subject's brain is probably not receiving any signals from his/her eyes. Other applications include the diagnosis of optic neuritis, which causes the signal to be delayed. Fig.2 (a) shows visual evoked potential recording setup where pattern reversal method is used as stimulus, and Fig.2 (b) shows a typical visual evoked potential.

Artifacts in EP waveform recordings typically result from voltage changes due to eye blinks, eye movements, muscle activities, and power line noise. Artifact detection in EPs is essential because artifacts are known to frequently occur in evoked potential data acquisition [3]-[7].

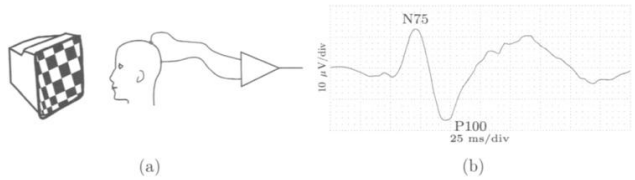


Fig.2. Visual evoked potentials. (a) Recording setup where pattern reversal method is used as stimulation and (b) typical VEP morphology.

II. ARTIFACT DETECTION STRATEGY

Artifacts are rejected by first removing signals with excessively large amplitude variations or signals with little or no amplitude variations using a standard deviation test. Signals with samples that have been clipped are removed using a clipping test [8]-[9]. Kurtosis test is used to detect and reject artifacts that are not detected by standard deviation test. It enhances the peaks of the average evoked potentials. These tests can be used to identify faulty stuck-at recording channels that always give the same readings.

If a channel has stuck at fault, the EPs of that channel are discarded from further analysis. We assure that, if an artifact occurs in one channel then the responses of all the channels are also artifacts. This assumption is valid as the EPs of neighboring channels are highly correlated. Therefore for a given trial, if an artifact is detected in any one or more channels, single trial data of all the channels for that trial are removed.

The three tests are described using $z_{m/c;n}$ to represent single trial EP n, $n = 1,2,...,N$, in the ensemble of class c, c = 1,2,…,C, recorded at channel m, m = 1,2,…,M. Where N is the number of single trial EPs in each ensemble, C is the number of brain activity categories, and M is the number of channels. The c-class ensemble of EPs collected at channel m will be referred to as m/c ensemble [10]-[13].

A. The clipping (CL) test

This test is designed to exclude single trials whose amplitude have been clipped. An evoked potential will be detected as a clipped signal if more than λ samples have the same maximum or minimum values .

To determine if $z_{m/c;n}$ is clipped,

let $\lambda_1 = \max\left[z_{m/c;n}(k) \right]$

and $\lambda_2 = \min\left[z_{m/c;n}(k) \right]$,

where $z_{m/c;n}(k)$ is sample *k, k=1,2,…,K,* of $z_{m/c;n}$

Let

$$v_{1k} = \begin{cases} 1, & \text{if } z_{m/c;n}(k) = \lambda_1, k = 1,2,...,K \\ 0. & \text{otherwise} \end{cases}$$

Similarly let

$$v_{2k} = \begin{cases} 1, & \text{if } z_{m/c;n}(k) = \lambda_2, k = 1,2,...,K \\ 0, & \text{otherwise} \end{cases}.$$

The single trial EP $z_{m/c;n}$ is clipped if

$$\sum_{k=1}^{K} v_{1k} \geq \lambda \quad \text{or} \quad \sum_{k=1}^{K} v_{2k} \geq \lambda .$$

If $z_{m/c;n}$ is clipped for one or more values of m, then the MCEP $z_{c;n}$ is regarded as clipped and removed from the ensemble of class c. The parameter λ is not a function of c. Fig.3 shows an example of a clipped evoked potential.

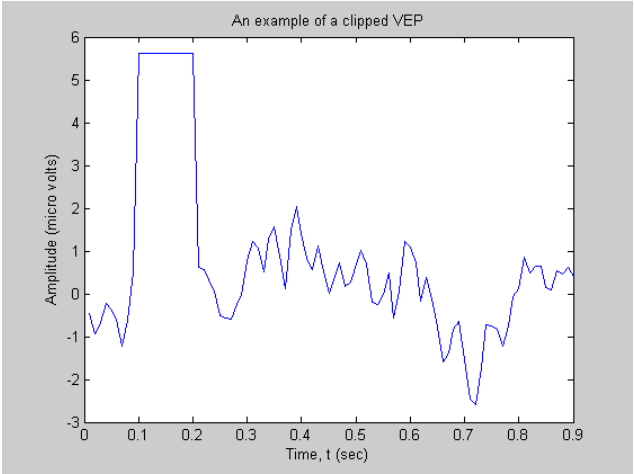


Fig. 3. A signal with clipped peaks

B. The Standard Deviation test

Standard deviation of a single trial response $z_{m/c;n}$ in the m/c ensemble is defined as

$$\sigma_{m/c;n} = \left(\frac{1}{K} \sum_{k=1}^{K} \left(z_{m/c;n}(k) - \hat{z}_{m/c;n} \right)^2 \right)^{1/2}$$

If the standard deviation $\sigma_{m/c;n}$ of the samples of a single trial response $z_{m/c;n}$ in the m/c ensemble is outside a threshold window $[\tau_{\sigma 1}, \tau_{\sigma 2}]$, then *n*th single trials of all M channels are regarded as artifacts and are discarded from the m/c ensemble [14]. That is, multi channel EP $z_{c;n}$ is an artifact,

if $\delta_n \geq 1$.

Where $\delta_n = \sum_{m=1}^{M} \rho_{m/c,n}$

and

$\rho_{m/c,n} = 1$, if $\sigma_{m/c;n} < \tau_{\sigma 1}$ or $\sigma_{m/c;n} > \tau_{\sigma 2}$, $m = 1,2,\ldots,M.$

The threshold $\tau_{\sigma 1}$ is selected to be close to zero, in order to detect responses that are relatively constant over the entire duration of the event related potential (ERP), whereas the threshold $\tau_{\sigma 2}$ is determined empirically. If the standard deviation is less than the threshold $\tau_{\sigma 2}$, or greater than the threshold $\tau_{\sigma 2}$ for all n at any c, the channel is regarded as faulty and the EPs of the faulty channel are removed from further processing. Fig.4 shows an example of artifact detected by standard deviation test.

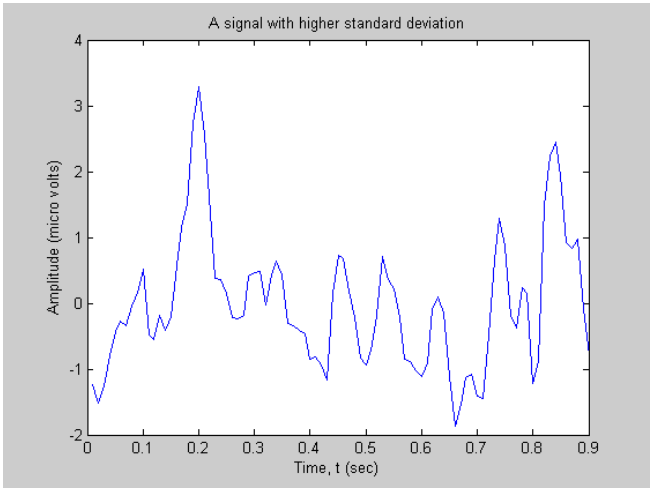


Fig. 4. One of the artifact signals detected by standard deviation test.

C. The Kurtosis test

Kurtosis is the fourth order moment, which is useful in the detection of transients due to external noise such as switching on/off of electrical or electronic equipment.

If the kurtosis

$$\kappa_{m/c;n} = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{z_{m/c;n}(k) - \hat{z}_{m/c;n}}{\sigma_{m/c;n}} \right)^4$$

of the samples of a single trial response $z_{m/c;n}$ in the *m/c* ensemble is outside a threshold window $[\lambda_{\kappa_1}, \lambda_{\kappa_2}]$, then the *n* th single trials for all M channels are regarded as artifacts and are discarded from m/c ensemble.

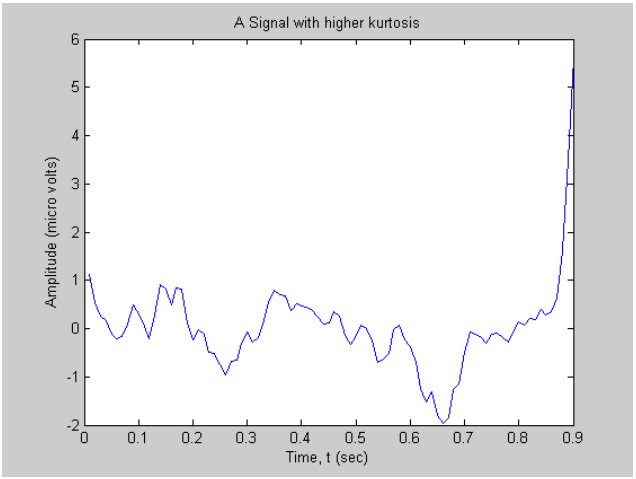


Fig. 5. One of the artifact signals detected by kurtosis test.

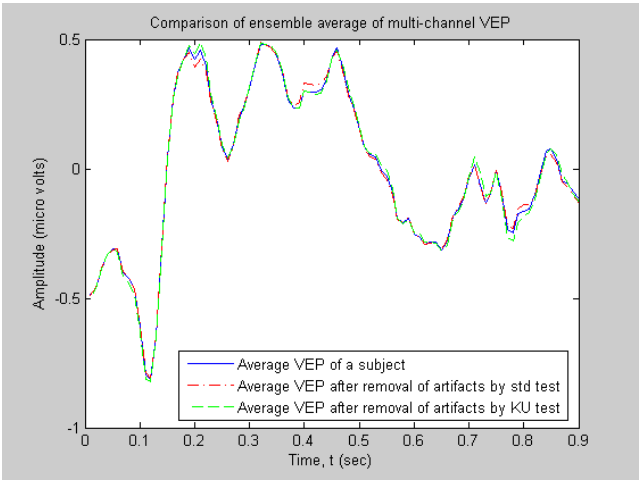


Fig.6. Comparison of average of actual VEP with average VEP after removal of artifacts using standard deviation test and kurtosis test.

This test detects and excludes signals with higher peaks so that average evoked potential will be smoothened. Fig.5 shows an artifact detected by kurtosis test. Fig.6 shows a comparison of averages of actual evoked potential with average VEP after removal of artifacts using standard deviation and kurtosis tests.

Quality Factor

Quality factor, $\theta = 1 - \hat{\theta}$

Where $\hat{\theta} = \frac{a}{N}$

a = No. of artifacts detected
N = No. of trials of data tested

III. SIMULATION AND RESULTS

The artifact detection strategies using standard deviation test, clip test and kurtosis test were applied to 14-channel VEP ensembles acquired from four different subjects. Single trial EPs having clipped peaks, lower (close to zero) or higher standard deviation or kurtosis or both, are detected as artifacts and removed while classifying the EPs. Examples of artifacts detected by standard deviation and kurtosis are shown in Fig. 3 to Fig. 5.

Following table shows details of artifacts detected in 14-channel 71-trial evoked potentials of a typical subject

No. of atrtifacts detected using standard deviation test alone	3
No. of artifacts detected using kurtosis test alone	3
No. of artifacts detected using KU test after removal of artifacts using STD test	2
Total no. of artifacts detected using STD and KU tests	5
Quality factor before removal of artifacts	91.55%
Quality factor after removal of artifacts using STD test but before removal of artifacts using KU test	92.65%
Quality factor after removal of artifacts using STD and KU tests	100%

IV. CONCLUSIONS

The primary objective of this work is to identify and reject artifacts. The artifacts were first detected using a sequence of within channel standard deviation and clipping tests. Some more artifacts which could not be detected by these two tests are identified by using kurtosis test. It is observed that removal of artifacts using kurtosis test improves peaks of the average VEP and also it improves the performance of evoked potential classifiers, much more effectively in addition to that provided by standard deviation test.

REFERENCES

[1] Rodrigo Quan Quroga, *Evoked potentials Encyclopedia of Medical Devices and Instru-mentation, Second Edition, John Wiley & Sons, Inc.*(2006).

[2] R.J. Croft, R.J. Barry, "Removal of ocular artifact from EEG: a review," *Clin. Neurophysiol.* 30 (1) (2000) 5–19.

[3] G.L. Wallstrom, R.E. Kass, A. Miller, J.F. Cohn, N.A. Fox, "Automatic correction of ocular artifacts in the EEG: a comparison of regression based and component-based methods," *Int. J. Psychophysiol.* 53 (2) (2004) 105–119.

[4] T.D. Lagerlund, F.W. Sharbrough, N.E. Busacker, "Spatial filtering of multichannel electroencephalo-graphic recordings through principal component analysis by singular value decomposition," *J. of Clin. Neurophysiol*. 14 (1997) 73–82.

[5] S. Casarotto, A.M. Bianchi, S. Cerutti, G.A. Chiarenza, "Principal component analysis for reduction of ocular artifacts in event-related potentials of normal and dyslexic children," *Clin. Neurophysiol.* 115 (3) (2004) 609–619.

[6] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. Mckeown, V. Iragui, T.J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology* 37 (2000) 163–178.

[7] C.A. Joyce, I.F. Gorodnitsky, M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," *Psychophysiology* 41 (2) (2004) 313–325.

[8] P.J. Rousseeuw, A.M. Leroy, "Robust regression and outlier detection," *Wiley Series in Probability and Mathematical Statistics, Wiley, New York*, 1987.

[9] R. Barandela, E. Gasca, "Decontamination of training samples for supervised pattern recognition methods," in: *Proceedings of Joint IAPR International Workshops SSPR and SPR 2000, Springer, NewYork*, 2000, pp. 621–630.

[10] F. Vazquez, J.S. Sanchez, F. Pla, "A stochastic approach to Wilson's Editing Algorithm," *IbPRIA* 2005, pp. 34–42.

[11] L. Gupta, D.L. Molfese, R. Tammana, P.G. Simos, "Non-linear alignment and averaging for estimating the evoked potential," *IEEE Tran. Biomed. Eng.* 43 (4) (1996) 348–356.

[12] L. Gupta, J. Phegley, D.L. Molfese, "Parametric classification of multichannel evoked potentials," *IEEE Trans. Biomed. Eng.* 49 (8) (2002) 905–911 49(9) (2002) 1070.

[13] L. Gupta, B. Chung, M.D. Srinath, D.L. Molfese, H. Kook, "Multichannel fusion models for the parametric classification of differential brain activity," *IEEE Trans. Biomed. Eng.* 52 (11) (2005) 1869–1881.

[14] Hyunseok Kook, Lalit Gupta, Srinivas Kota, Dennis Molfese, H. Lyytinen, "An offline/real-time artifact rejection strategy to improve the classification of multi-channel evoked potentials," *Pattern Recognition* (2008), pp. 1985-1996.

 **About the Author- V. Adinarayana Reddy** received his graduate degree in Electronics and Telecommunication Engineering from The Institution of Electronics and Telecommunication Engineers, New Delhi in1996 and M. Tech in Electronic Instrumentation and Communication Systems from Sri Venkateswara University, Tirupati in 1999. He joined as faculty in the Department of Electronics and Communication Engineering at KSRM College of Engineering, Cuddapah. Currently he is working as Professor and Head of the Department, Electronics and Communication Engineering at Rajoli Veera Reddy Padmaja Engineering College for Women, Cuddapah. His research area of interest includes signal processing and communication systems.

 **About the Author- Dr Putta Chandra Sekhar Reddy** received the B.Tech. degree in the electronics and commu-nications engg from JNTUH, Hyderabad, India and M.E from Bharatiyar University, Coimbatore. He received M.Tech and Ph.D from JNT University, Hyderabad, India. He joined as faculty in JNTU, Anantapur. Currently he is working as Professor Co-ordinator in JNTUH Hyderabad, India. He is an author of numerous technical papers in the Fields of High speed networking and wireless networks. His research interests include mobile and wireless communications and networks, personal communications service and high speed communications and protocols.

 **About the Author- G. Hema Latha** received her B.Tech. Degree in Electronics and Communication Engineering from Sri Venkateswara University, Tirupati in 1997, and M.Tech in Instrumentation and Control Systems from Sri Venkateswara Unversity, Tirupati in 2003. Smt. Hemalatha joined faculty in Electronics and Communication Engineering at G. Pulla Reddy Engineering College, Kurnool. At present, she is working as Associate Professor in Electronics and Communication Engineering, KSRM College of Engineering, Cuddapah. Her research interests include Signal Processing and Communication Systems.

 **About the Author-. Dr.T.Jayachandra Prasad** obtained his B.Tech in Electronics and Communication Engg., from JNTU College of Engineering, Anantapur 515002, and Master of Engineering degree in Applied Electronics from Coimbatore Institute of Technology,Coimbatore. He earned his Ph.D. Degree (Complex Signal Processing) in ECE from JNTUCE, Anantapur. He joined as faculty at KSRM College of Engineering, Cuddapah. Later he worked as Head of the department, ECE, at KSRMCE. Presently he is working as the Principal of RGMCET, Nandyal-518502. Dr. Jayachandra Prasad is having more than 22 years of experience and has more than 32 technical publications in National /International journals and conferences.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-418-424 ISSN: 2249-6645

Water tanks design in urban spaces designed for optimal use of flowing water from precipitation to climate

Hassan Bayadi[1], Mahdi Koohi[2]
[1](Department of Economic Managing and Vice chairman of the Tehran City Council)
[2](Department of Electronic Engineering Islamshahr Branch, Islamic Azad University, Tehran. Iran)

Abstract
Increasing global demand for the use of water in effect of uncontrolled population growth, industry growth, increasing the level of planting agricultural products as well as contaminating water resources especially surface runoffs due to mixing with wastewater and chemical pollutants caused abundant problems for communities and unpleasant periods for next generation. In this research with choosing and studying environmental condition and rain in Mashhad city which faced with extreme lack of water in order to save runoffs in raining seasons using them to meet the need of parks in the city, constructing suitable and adjacent reservoir was proposed. The rainfalls with different occurrence possibilities the amount of runoff was calculated by HEC-1 software. And surveying two parameters, i.e. peak and volume of runoff the best volume of reservoir is obtained which this prohibit destructive floods and use it in drought condition for irrigating green areas and plant covers. In this research in designing the amount of volume and building sample reservoir which followed by the ease of designing with less expenditure which showed efficiency of the above systems in optimal use of resources and saving in consumption.

KEYWORDS-**HEC-1,Software,Watertanks design, urban spaces.**

I. INTRODUCTION
Societies in which house, food and environment could provide human with well-being and comfort favorably, there made appropriate motivations for effort, movement, construction and cultural and industrial improvement were resulted. Study of ancient civilization show that if in some historical periods climate and environmental difficulties could make put barriers on the way of effort of construction, human with his innovation and invention minimized these constraints. Huge reservoir, ices, wells, wind breakers, and magnificent and tall buildings designed in the heart of the desert are the indication of industrious human's effort and generating well-being which occurred in the last century. Today, although civilized and developed human could, with the help of technology, make appropriate easy conditions with constructing towers and skyscrapers and developed cooling and heating systems. Nonetheless, such heavy expenses payment of mechanic systems as well as using fossil fuels for all people in all places is not possible. Thus, referring to the simple techniques which our antecedents applied to using environmental potentials can provide optimal well-being for majority of people in city or village. Out antecedents who had never access to cheap oil and gas only with the use of rain, wind, shining, sun and shadow factors and utilizing temperature differences in 24 hours provided optimal conditions for work and life. Use of these experiences with new knowledge can be helpful for today human in optimal water, fuel and energy consuming. As we know 34 percent of the earth is covered by water, 60 percent of animal body configuration and about 85 percent of plant active contests are formed by water [1] and the necessity of consistent exchanges of these elements between outer and inner environment of alive creature make it clear that without water life is impossible. With a look to the past we will find out that civilizations and populations were formed next to rivers or areas with ample water and fertilized grounds and when due to population increase or other reasons human have been forced to immigrate, they have chosen areas for residing to have access to water as easy as possible. Global population increase which entails more consumption in all fields especially increasing need for water, how will determine the human's future? This is a question which worries the most optimistic people. The history of saving and optimal use of water and water providing is begin from those days which human being chosen group life and the first residents were built next to rivers like the Nile, Tigris, Euphrates and Indus due to the easiness of supplying water. He learned transition from river flow and saving from lakes and where there was no access to the rivers to meet his needs he began to dig wells. Following the nature with digging groves and making suitable steeps on the ground transferred water to the consumption place and after thousands years with thought growth could make channels under the ground and by well transfer the underground water to the surface. Also using soil, stone and logs constructed dams in the water and little by little enhanced their use from saving into securing their

resident against floods, irrigating uphill grounds and getting energy (wind mill)[2,3].

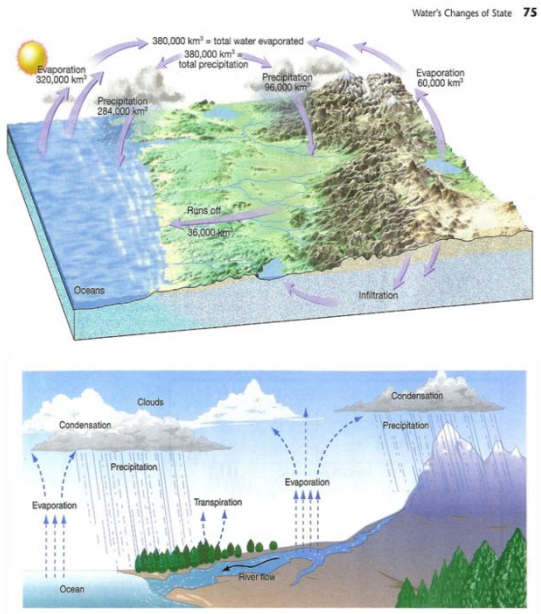
II. Importance And Necessity Of Implicit Subject

No doubt, among all the disasters which are threatening the earth now lack of water (as a crisis) will be turned to a great issue. United Nations organization has predicted that in the year of 2050 more than 5/4 people of the world will beexposed to injuries and losses resulted from lack and contamination of water. Now also every year 2 million children lose their life due to lack of access to healthy beverage water, daily 10 thousand children died due to diarrhea resulted from using polluted water and one third of people who are hospitalized is resulted from diseases which their primary reason is somehow related to water. The specialists believe that in the next century water will be rise as a political issue among countries and its economic value will be more than the most today expensive materials. Supplying healthy beverage water is often costly and about half of the world people don't have access to that. Unfortunately, in most of our cities this costly water is not used properly and the big part of it is wasted or is used in unnecessary cases [6]. The amount of water that exists on the earth and contributed in the water cycle is nearly fixed. The volume of existing water-rate of the earth is about 1454 million km sq. which according to table (1) is concentrated in different parts of the earth [6]. The huge bulk of this water exists in the oceans and is useless for drinking and agriculture.

Table(1): estimate of water quantity on earth

Gathering place	Exchange time(year)	Percent of total	Water volume (km1063)
oceans	3000	94.2	1370
Hypogeous source	5000	4.13	60
Ice layers	8000	1.65	24
Surface waters	7	0.019	28
earth wet	1	0.0055	80
rivers	0.031	0.00008	1200
aerosphere	0.027	0.00096	140

Regarding to this estimate fresh water-rate is estimated 40 million km sq. the existing water in the atmosphere which is vapor is only 35% of the whole fresh water, but because of short of time of exchange it has an important role in supplying and distributing water resources because evaporated water from oceans level and humid surfaces is transferred by horizontal and vertical flows of air and in all areas of the earth is purred as rain and snow. This was repeated continuously for about 40 years and thereby provided rivers water, undergroundresources, lakes and other resources of fresh water. As was showed in figure (1) [6].Thus it is necessary to supply fresh water in the raining area and runoff resulted from it be studied and utilized in advance. The world population formany reasons has a quick growth so that today the world population is over 5 billion people. Natural resources limit and quick growth of populationcaused that different societies are inneed ofplanning in different field. The goal of planning is optimal use of existing equipments favorably. This doesn't mean the reduction of consumption, but is thank to economic justifications and continuous use of existing resources. Dry and mild climate in a vast part of Iran and high growth of population caused planning in the field of fresh water resources to have high importance. Therefore, we can label in this field saving water behind dams, change in the way of irrigating, water consumption and so on. By huge investments in order to use and save water some small plans can be performed to prohibit wasting water area by area.



Figure(1). Diagram of water traveling in nature

III. Case Study In This Plan

In surveying every goal we should analyze basic information in a proper way. In the current plan two main goals, fighting against flood and water supplying and some lateral goals were proposed which here it will suffice to deal with the main goals and regarding to that the suggested structure is the same and just they are different in details and also various aspects of the plan have many shared parts, after determining informational needs of each of these goals, the data are rank according to the priority and explained in the next section in title of "recognizing studying area". Now, here we survey informational aspects and needs primarily on hydraulic structure that can reduce the amount of flood by collecting runoff and then the same way is applied to a water supply resource according to the

use. According to the UNICCO surveys on water studies, using hydrologic and planning of collecting systems and prohibiting urban floods four data groups should be studied which are:
1-Physiographic features of basin: topography, steep, geographical position, type of soil, …
2-The data related to surface waters: rivers, streams, dams, and resources of surface water.
3-Underground waters and soil features,
4-climate forecasting data: rain, temperature, humid…

IV. OPPOSITION AGAINST FLOOD

1-Generating water equipment out of the cities to protect them from entering up floods or floods resulted from overflow of basins and rivers adjacent to the cities.
2-Building barriers out of cities to trap the floods or slow them down in order to increase the time of concentration in the lower parts.
3-Generating water equipment inside the cities and preserving existing floods to control and entered floods.
4-Using urban facilities and their physiographic features in urban development design in order to guide and natural discharge of floods in city.
5-Enacting and performing constructing rules to reinforce urban buildings resistance against flood destructive effects. Among the above cases, three cases are related to the inside of the cities and two cases are related to the out of the cities which it indicating the importance of city constraint in controlling floods. Thus in controlling floods projects we can't ignore the role of reinforcing urban constraint by infeasibility of their covered surface.

V. WATER PROVIDING

In recent years some big and subjective plans occasionally for making changes in area climate were represented which are not acceptable, like: transferring …rivers by channels, transferring the water from lakes to deserts in building lakes there, making channels and cutting mountains to transfer humid air and building lakes by fossilized water. Now it is determined that with the above plans not only can change the climate of the area, but there will arise some new and unknown problems. Thus, we should seek for solutions while is accordance with the nature, satisfy our present and future needs. The increase of productivity and the optimal use of rain is accepted by all the responsible people and specialists and finding a practical way to this goal "supplying water of each area from that area" needs further investigating and research.

VI. AN OVERVIEW TO WATER ISSUE IN DRY AREAS

We will talk in detail on distributing criteria of areas and applying the word humid, semi-humid, semi-desert, desert and… in physical and climate features of Khorasan. In this part we deal with the importance of supplying water for dry areas. Dry and semi-dry areas are covered most of the lands on the earth that there exists relatively much population and due to bad climate conditions always faced with many problems. About 90 percent of Iran has dry and semi-dry weather [4]. In dry areas according to required conditions to form soil and other effective factors in its change and evolution, both fertilized soil and usable water are low, but the issue of water is more important than other limiting factors because if there is appropriate water sufficiently we can somehow modify inappropriate soil. And if climate conditions are incompatible we can abate them in part, e.g.: with mo0re irrigation we can null vapor unfavorable effect and use the warmth of these areas for better performance of agriculture corpses and/or according to the length of daylight and also being prolonged the time of growing season in desert and dry areas using optimal irrigating we can produce different agriculture products. And also when the tempest is regarded as the restricting factor in the area we can make alive wind breakers with planting trees and irrigating them. But in dry areas lack of rain causes that the use of surface waters and underground waters not to meet the residents of this area. Thus, it is necessary for a better and more use of that little rain in these areas help to solve this problem in part.

VII. GENERATING REQUIRED WATER RESOURCES

A way of more productivity is also utilizing rainfalls which are the intention of the current plan. Saving runoff resulted from rains is in resources which are generated especially for this purpose and are considered to be exploited when is needed. A reservoir can be built for many reasons and whatever application of it is justifiable for the amount of investment in this way. Because building a hydraulic building also should have economic justification like other structures. However, the serious lack of water and specialists' worrying prediction and international communities on this lack for future decade and also lack of substitute and the dependence of life on water justifies every kind of investment. But the restriction of equipments makes us to put some measure on the top which need less fiscal resources and have high efficiency. For this purpose, in this section we will mention goals and applications of suggestedwater reservoir and required data for such structures.Data analysis is presented in the next section after proposing information jointly with flood section.

VIII. SELECTING THE STUDY AREA AND PERFORMING METHOD

It has been many years that we were witnessing occurrence of destructive floods in some provinces such as the vast province of Khorasan. And a phenomenon called drought is seriously threatening most of the areas. No doubt, lack ofwater was one of the most important issues which obsessed the mind of most people in charge especially planners of agriculture section and urban affairs so that the citizens of these areas hardly can supply their beverage water and use it in form of rationing. one of the reasonsof choosing Mashhad city in respect with optimal use of urban runoffs first is thank to its population and also 12 million pilgrims and tourists which were affected by the destructive drought effects that in the year of 1378, paid 10 billion Rials for irrigation and in most areas of the city also to water the parks they use beverage water that at least with studying this plan we can use season runoffs which are resulted from rainfalls in order to supply the water of parks of this city which should be the pattern of religious cities in the world.

IX. IN SELECTING THE AREA THE FOLLOWING FACTORS SHOULD BE TAKEN INTO ACCOUNT:

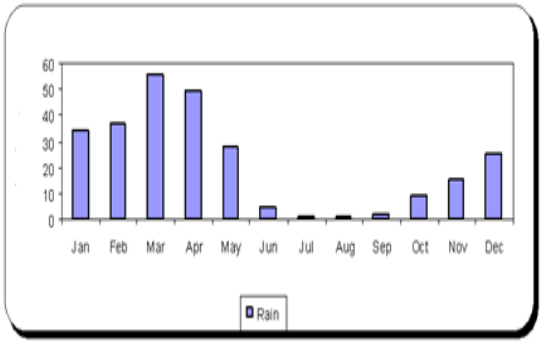
1-Hydrologic balance: the earth hydrologic system regarding to the amount of existing water being restricted we can consider a close system, but we survey small system of basin of Mashhad plain that is placed in a part of hydrologic cycle flow path.

2-Geographical situation: Mashhad town with a population of 2500000 are resided and annually 1500000 of passengers and pilgrims is a religious-tourism city. The average altitude of this city is 110 and longitude of 59 degree and 38 east ….of 36 degree has a width of 200 km.

3-climate situation: due to that in many cases climate parameters in synoptic scale (those maps which showed about one fourth of the use of the earth) become meaningful and also masses of air usually cover a big area of the country. It is necessary that local climate be studied as a part of whole.

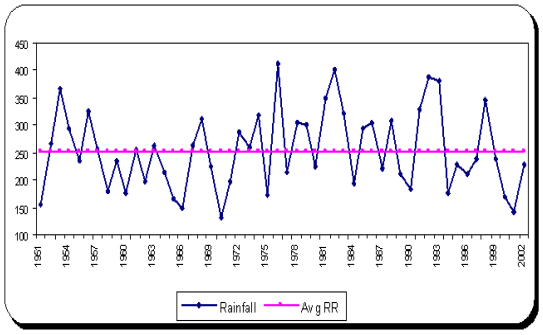
X. COLLECTING INFORMATION AND FINDINGS OF RESEARCH

Mashhad has a winter raining regime and both forms of solid and liquid is seen. Basically raining in areas can be resulted from a vast air system in large scale or locally be in an air mass. The typical type of local raining is thunder. According to this that two main goals of the plan are preventing damages of flood and saving water for use, thus the issue of raining like any other hydrologic plan both generally and in detailed be surveyed. On one hand, entering water in the area is almost allocated to rain, thus rainfalls are the base of water budget. In this paper the statistical of rain of Mashhad weather in 40 recent years was examined and also notable results were obtained, as shown in figure (2).



Figure(2)-Mash had Monthly Rain Average

The average of raining in Mashhad during a statistical period of 1960-2000 is of 259 mm. rainfalls mostly occurred in winter and the maximum average of monthly rain is of 5/55 mm in March and the minimum average of monthly rain was o 6/0 mm in July. Graphs (2-4) show the average raining of weather station of Mashhad during the statistical period. The maximum 24 hours rainfall in 30th of December 1970 was reported of 47 mm. rainfall collection annually during statistical period with mobile averages of 3 and 5 years which are showed in figures (3) and (4) annually rain graph and specialists' opinion [7] verified the absent of this in annual rainfalls but in mobile average graphs less rain and much rain periods are determined.

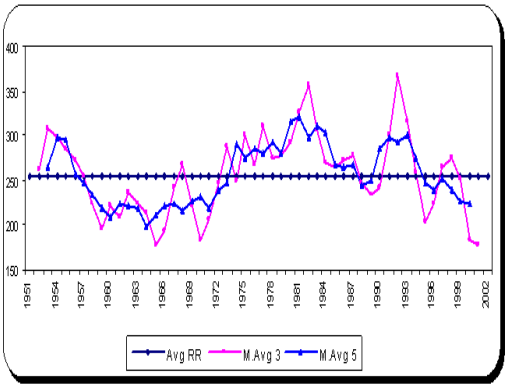


Figure(3)-Mashhad average long-term and rain yearly

Of course the amount of rainfalls in this city is low, but this low level is notable because regarding to the definition of rain mm (one mm is equal to one liter water in one square meter or ten cubic meter in one acre) and regarding to the amount of mean rainfall with a simple calculation we find that the average of rainfalls is annually in each acre 2590=10*295 cubic meter. The vast of Mashhad city is about 200 km and the extension of

Mashhad area is 16500 km with allocating the above calculation we can conclude that:
The average of annual rainfalls in Mashhad city and Mashhad province is:

The amount of water in cubic meters=the mean of rainfall*area in acre*10
In Mashhad city→259*100*200*10=51800000
InMashhadprovince→259*100*16500*10=4273500000



Figure(4)-Mashhad average changing rain

Although for providing reservoirs or constructing dams and also hydraulic designs information on the whole annual rainfall is necessary, in the most studies more details such as …of rain time and for calculating…and flood even the elements of one rainfall like its intense, duration and the abundance of occurrence should be examined necessarily because especial features of flood such as volume of flood, discharge or peak flow and the time of getting this discharge is depended to rain features as the most important factors of making floods and erosion and so on. The above elements are as follow:

1-Duration or continuity (from the beginning to the end of rainfalls).
2-Intensity (the amount of rain in time unit, mm per hour).
3- Abundance of occurrence (the average of the number of years which there is between two similar raining).

While the most extreme flood is resulted from a rain which its insistence is equal to basin concentration time and the time of a physical parameter which its amount is different for each basin, thus if we want to calculate the amount of flood in an area for different basins we should have the maximum intensity of rain in continuities which its amount is equal to the time of concentration of basin. So, to have relationship and a graph on which we can estimate the rainfall for different continuities is that of main requirement which is used in water structure design. On the other hand water structurelike: bridge, flood block, dam, drainage, collecting streams for sewage and whatever related according to the importance and sensitivity to destruction or the amount of money that is spent for building it or dangers which are resulted from ruining is determinedwith return period. Thus, intensity or rainfall that is used for certain continuity in its design should be related to considered return period.Rain intensity changes into return period ofa statisticalfunction are possibilities that have different amount for every climate area.Having this relationship is that of design requirements in waterworks. From integration of the above relationship we can achieve uniform functions or graphs on which we can estimate the amount intensity in different continuities and return periods. The aim of this report is representing such relationships and graphs for the targeted station. These functions and graphs which have been extracted from existing data is represented in a way that the amount of rain intensity in 5 minutes to 12 hours continuities in return periods of 20, 25, 50, 100 years are obtained. Choosing rain continuality changes range and standard return periods is for their high application which we deal with them in water structure design or whatever similar.

XI.PLANSIMULATIONAND IMPLEMENTATION

In simulating our plan HEC-1 software was used. One of the advantages of using HEC-1 in comparison to other models is the simple way of organizing information. HEC-1 determines format and information of basin and formats of discharge using some input cards or codes. HEC-1 is one 6 soft wares in the name of HEC which is provided by American engineers. In fact this software is simulation software for runoff rainfall; and in a general definition is able to estimate the resulted runoff by using an observed rain.
The most important information which can be obtained if a certain rain entered is the amount of discharge hydrograph from urban area. The advantage of using this method is determining the discharge volume and amount of discharge peak. The obtained information from this method is much more completed than the two other, while it needs more comprehensive information than them. Another important difference between them is that is has more complex calculations than those two previous usually for this method due to high calculations the most common software is used. One of the most known and used program in this field is HEC-1. In order to work with this software the following information are necessary.
1-Rain information: This information can be given to the program as rain hydrograph, certain information of rain gauge, likely rain and soon.
2-Damages information: This information is parameters related to different methods of

calculating damages in hydrology science. HEC-1 is able to consider different methods for calculating damages of rain and of course introducing every certain method, its related information must be given to the program.
3-Unit hydrograph information: unit hydrograph is in fact information which is obtained from previous memory of system during different rains. In the case of lack of this information we can use approximate artificial unit hydrograph in calculation.

XII. DETERMINING APPROPRIATE VOLUME OF RESERVOIR USING HEC-1

HEC is a software used in hydrology engineering and is used in order to simulating rain in a basin. In this project in order to determine required volume for saving runoffs resulted from rain this software has been used. To perform calculations related to simulation of runoff rain we need to determine 3 main factors for the study area, which are:
1-Rainfall 2- unit hydrograph 3. The amount of damage

XIII.CALCULATING SUGGESTED RESERVOIR VOLUME

These days, there are thunders with different intensity in order to determine required volume for gathering these runoffs during a year it is assumed that we have in average thunder for each mentioned day and the number of occurrence is considered to match with return period of that rain so that an occurrence possibility of two years rain is 50 percent rain and for 100 years is 01/0. So the required volume for gathering the number of thunders annually is calculated as follow:
1440*1/2+3600*0/2+3960*0/1+4320*0/05+4680*0/02+5040*0/01*7/7=16909/2170003 m
Thus, a reservoir with volume of 170003 is suggested for specified area. Following that we can determine other reservoirs for other areas which are to gather runoffs.

THE OBTAINED RESULTS FROM PERFORMING HEC-1 SOFTWARE AS RUNOFF VOLUME:

resulted from rain with different return periods are showed in table(2). As it is said in determining rain intensity for different states we considered fixed continuity which regarding to likewise ratio between rain and continuity of this premise doesn't have that much effect on determining runoff volume of whole rain.
In order to design a reservoir for gathering surface waters and analysis should be conducted on the number and the possibility of occurrence of these thunders during a year. Mashhad city has the mean of annual rain of 3/259 mm. the main distribution of this rain is as scattered thunders with a shallow depth (less than 10mm). Such rains do not make runoffs or flood in urban areas and many of them are not retainable because has a high percent of damages. The mean of the days which Mashhad faced is 7/7 days per year.

Table(2): resulted from rain with different return periods

Result of run off volume(yr)	4320	5/02	20	3960	3600	1440
Result of run offvolume(mm/hr)	4680	5/72	50	3960	3600	1440
Result of runoff volume(m3)	5040	6/23	100	3960	3600	1440

XIV. CONCLUSION

Supplying water for different consumers is one of the most vital activities for dry and semi-dry areas. Whatever activity to increase the percent of…. existing water resources includes high profit for the area. In this paper by using a systematic approach to the issue of optimal exploitation from existing water resources, we tried to propose a method to prevent damages resulted from rain in the best way possible. Using water resources often was considered in form of big and costly projects. Such projects in respect with high volume of and high investment are in need of much time and cost to be implemented. It is the case, when we can take advantage of many facilities by a proper and systematic approach to the environment and at the same time we can save much costs and damages. One of the existing and related challenges to this issue is urban runoffs. These runoffs are direct consequences of urbanity and urban life which often cause many problems because water is flown in pavements, streets, and settlements. In this research a solution has been proposed in which we can use existing unwanted water resources resulted from rain by using this certain position in urban areas. Gathering and removing runoffs resulted from rain in urban lands is implied as kind of security, Sanitary and well-being services which should be presented to urbanite community this services in addition to tangible profits and valuable has some intangible profits which affect procedure of development. According to the all studies gathering runoffs, directing them to the water reservoirs and using them in dry season is suggested to municipalities and owners of big complexes because in this method:

Supplying water is done with the less cost.

1-uses facilities and potential of the area and reduces the danger of floods.
2-can use existing natural positions like holes and artificial complications.
3-replacing small and cheap with high efficiency projects with big and costly projects
4-generating and reinforcing contribution culture and saving in water is possible.
In this respect we can present suggestions which of course are also presented in previous cases that are as follow:

1- Surveying and finding existing different areas in the city which has the appropriate and similar potential to the study place of this project in order to implement the project the best way possible
2- designing certain place of reservoir in a way that in addition to gathering water it reduces peak power and resulted floods volume
3- using natural positions like holes or natural and artificial complications for building such reservoirs
4- using many small and cheap projects as alternative for a big and costly project.
5- 5- We hope that the method and presented results in this thesis will be a step toward developing water supply position of the urban areas especially those which have dry climate.

REFRENCES

1-Plant Biology.Hassan's religion.Publication of Education Page 76 1985.
2-Water and Irrigation Technology in Ancient Iran - considering enayatollah Reza, Gholamcorse, MohammadAli Imam Shoushtari and Ali Akbar entezami - The Ministry of Water and Power 1980.
3-Finding water, water supply, irrigation and Bsnjy in ancient Iran - Ahmad sponsor arrow 20 October 1991.
4-Arid regions. Vol. Dr. ParvizKrdvany. Third Edition.2004
5-water resources issues in Khorasan Province - Dr. Sadollahvelayati 1996.
6-Principles of Applied Hydrology, Ninth Edition, Dr. Amin Alizadeh , 1997, University of Imam Reza, Page408.
8-Bulletin of the National Center for Scientific climatology of Mashhad - Twentieth Session of drought Research Council - Dr. Houshang vertical - July 2000.
8 - HydroliqueUrbaine. Tome 1 Dupont. Ander - Eyrolles 1981.
9-Ministry of Planning and Budget, and create. Design principles and criteria for urban water supply - Water Standards Publication No. 117-3 Office of the Ministry of Energy in 2006.
10- Monthly Sepehr - Volume VII - Number Twenty-Eight.2007.

Hassan Bayadi received B.S degree From the university of Karaj in I.E Engineering and the M.S degree in I.E engineering from the university Of science &technology of Mazandaran Currently, he is an Urban planning researcher with the Department of I.E engineering Elmikarbordi university.

Hot Spot Temperature Analysis in 3 Phase Transformers Using FEM Method

Diako Azizi[1], Ahmad Gholami[2], Diar Azizi[3]

[1](Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran)
[2](Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran)
[3](Department of Mechanical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran)

ABSTRACT

Transformers represent the largest portion of capital investment in transmission and distribution substations. In addition, transformer outages have a considerable economic impact on the operation of an electrical network. Hot-spot temperature (HST) value is a significant factor that directly impacts on the operation of transformers. For this purpose, energy (thermal) and Maxwell equation are solved with together until temperature is obtained. For numerical solution of above equation, finite element method (FEM) is used.The selected model for simulation is 3 phases, 10 MVA transformers.

Keywords-**Energy, hot spot, magnetic, temperature, transformer**

I. INTRODUCTION

Transformers are static pieces of apparatus with two or more windings, which by electromagnetic induction; transform a system of alternating voltage and current into another system of voltage and current usually of different values and at same frequency for the purpose of transmitting electrical power. Since transformers belong to the most valuable assets in electrical networks, it is suitable to pay higher attention to these operating resources. An outage impacts the stability of the network and the associated financial penalties for the power utilities can be increased. In a transformer operation, a part of the electrical energy is converted into the heat. Although this part is quite small comparing to total electric power transferred through a transformer, it causes significant temperature rise of transformer constructive parts, which represents the limiting criteria for possible power transfer through a transformer. That is why the precise calculation of temperatures in critical points is of practical interest. Thermal impact leads not only to long-term oil/paper insulation degradation; it is also a limiting factor for the transformer operation [1]. Therefore the knowledge of the temperature, especially the hot-spot temperature, is of high interest.

A hot-spot temperature calculation procedure is given in the International Standards [2]-[4]. In [5], [6] the algorithm for calculating the hot-spot temperature of a directly loaded transformer, using data obtained in a short circuit heating test, is given. These papers propose improvements in the modeling of thermal processes inside the transformer tank.

Calculation methods have to be based on an energy balance equation. Some attempts of heat transfer theory results application to the heat transfer from winding to oil are exposed in [7]. The usage of average heat transfer coefficient is typical in a transformer designing process to calculate needed number (area) of cooling surfaces [8]-[13].
In this paper, a procedure for obtaining the temperature distribution in the transformer is proposed. For numerical simulation of mentioned equation, it has been used finite element method.

II. MATHEMATICAL FORMULATION FOR HEAT CONDUCTION EQUATION [14]

The structure of a transformer winding is complex and does not conform to any known geometry in the strict sense. Under fairly general conditions, the transformer windings can be assumed cylindrical in formation; hence, a layer or a disc winding is a finite annular cylinder [10]. The thermal and physical properties of the system would be equivalent to a composite system of insulation and conductor. It has been assumed that heat is generated throughout the body at a constant rate, and oil in the vertical and horizontal ducts take away the heat through the process of convection. However, in an actual transformer winding, the conductor is the only heat source. Later in this section, formulations are given for calculating different thermal and physical properties of the system. It has been assumed that temperature is independent of space variable, due to the fact that winding structure is symmetrical. The temperature at any point on the periphery of circle for a specific value of r and z is deemed a constant (i.e., presence of spacers has been ignored, thus reducing three dimensional problem to a two-dimensional) one with r and z as space variables. Dielectric loss in insulation is assumed to be small compared to copper losses in the conductor. The surface of disc or layer has been assumed smooth.

The generalized system of non-homogeneous heat conduction equation with non-homogeneous boundary condition, in Cartesian coordinate system is written thus [10,15]:

$$k.\left(\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}\right)+Q=0 \qquad (1)$$

In the region a<x<b and 0<y<l.
At the inner cylindrical surface:

$$-k_1 \frac{\partial T}{\partial x} + h_1 T = f_1(y) \quad (2)$$

At the outer cylindrical surface:

$$k_2 \frac{\partial T}{\partial x} + h_2 T = f_2(y) \quad (3)$$

At the bottom flat surface:

$$-k_3 \frac{\partial T}{\partial y} + h_3 T = f_3(x) \quad (4)$$

At the top flat surface:

$$k_4 \frac{\partial T}{\partial y} + h_4 T = f_4(x) \quad (5)$$

Equations (1) through (5) represent the general heat conduction equation with convection at all four boundary surfaces. In the above equations, temperature T is a function of space variables x and y. The term Q is the heat source function and has been modified here to take care of variation of resistivity of copper with temperature. The heat source term Q can be of the form:

$$Q = Q_0.[1 + \alpha_c.(T - T_{amb})] \quad (6)$$

Where, α_c is the temperature coefficient of electrical resistance of copper wire. With this representation, the function Q becomes temperature dependent, distributed heat source. Boundary functions $f_1(y)$ and $f_2(y)$, derived from Newton's law of cooling are of the form (7) and (8):

$$f_1(y) = h_1.(T_b + m_1 y) \quad (7)$$

$$f_2(y) = h_2(T_b + m_2 y) \quad (8)$$

The term Tb is the temperature at the bottom of the disc or layer, as applicable. Terms m_1 and m_2 are the temperature gradient along the winding height (for layer) or along disc thickness for a disc. Similarly, functions f_3 and $f_4$representing temperature profiles across bottom and top surfaces, having the form same as shown in (7) and (8), where temperature gradient term has been taken as zero:

$$f_3(x) = h_3 T_b \quad (9)$$

$$f_4(x) = h_4 T_{top} \quad (10)$$

In general, thermal conductivities are shown different, and are indicated accordingly. But in an actual case, the thermal conductivities in radial directions (k_1 and k_2) are equal and conductivities in axial direction will also be the same (k_3 andk_4). Thermal conductivity has been treated as a vector quantity, having components in both radial and axial direction.

Resultant thermal conductivity of the system can be estimated as:

$$K = \sqrt{k_r^2 + k_z^2} \quad (11)$$

Where:

$$k_r = \frac{log\frac{r_n}{r_1}}{\left(\frac{log\frac{r_2}{r_1}}{k_1} + \frac{log\frac{r_3}{r_2}}{k_2} + \cdots + \frac{log\frac{r_n}{r_{n-1}}}{k_n}\right)} \quad (12)$$

$$k_z = \frac{k_{cu}\, k_{in}\,(t_{cu} + t_{in})}{(t_{in}\, k_{cu} + t_{cu}\, k_{in})} \quad (13)$$

Term K represents resultant thermal conductivity of insulation and conductor system. Heat transfer coefficients h_1 to h_4 (h_{tc}), are different across all the four surfaces. To determine boundary functions f_1 to f_4, it is necessary to calculate heat transfer coefficients across the four surfaces. Difficulty has been encountered in calculation heat transfer coefficient. It is reported elsewhere [10], that it depends on as many as 13 factors (e.g., winding size, type, duct dimensions, oil velocity, type of oil circulation, heat flux distribution, oil thermal properties, etc.). In this work, corrections have been given for temperature dependence of the thermal and physical properties of oil, such as viscosity, specific heat, volumetric expansion and thermal conductivity. It was found that there is negligible effect of specific heat, coefficient of volumetric expansion and conductivity in the present working range of loading. Following are some of the heat transfer relations and relevant formulae in natural cooling (ON) mode. These formulae have been used to calculate the h_{tc} [16]. Local Nusselt number for laminar flow over vertical plates has been shown below:

$$Nu = 0.6 Ra_{hf}{}^{0.2} \quad (14)$$

$$Ra_{hf} = Gr_{hf} P_r \quad (15)$$

Where Ra_{hf} and Gr_{hf} are the local Rayleigh and Grashof number based on heat flux (q_w) at characteristic dimension (δ).Pr is the Prandtl number of transformer oil. Expression of Rayleigh number based on constant heat flux is:

$$Ra_{hf} = \frac{g\beta\, C_p \rho^2 q_w \delta^4}{k_{oil}^2 \mu} \quad (16)$$

Mean Nusselt number in this case can be computed as:

$$Nu_m = 1.25[Nu]_{\delta=1}$$

However, correction to formula (14) has to be given for cylindrical curvature. The correction factor in this case is of the following form ($30<P_r<50$):

$$f(\xi) = 1 + 0.12\xi \quad (17)$$

Where:

$$\xi = \frac{2\sqrt{2}}{Gr^{0.25}} \times \frac{\delta}{r} \quad (18)$$

Here G_r is Grashof number based on temperature difference. Heat transfer coefficient can be computed as:

$$h = \frac{Nu\, k_{oil}}{\delta} \quad (19)$$

Where, h represents local coefficient. Mean coefficient (h_m) can be calculated from mean Nusselt number, as in (15). After knowing $h_m$for a particular surface, temperature difference between the winding surface and oil can be found out, dividing the $h_m$by the heat flux through the surface. Mean Nusselt number of top surface of annular cylindrical winding for laminar and turbulent regime, will normally be of the form:

$$Nu_m = 0.54Ra^{0.25} = 0.61Ra_{hf}^{0.2} \quad (20)$$

$$Nu_m = 0.15Ra^{\frac{1}{3}} = 0.24Ra_{hf}^{\frac{1}{4}} \quad (21)$$

$$\delta = \frac{b-a}{2} \quad (22)$$

Where a and b are the inner and outer radius of annular disc or layer. Nusselt number of bottom surface of annular cylindrical winding for laminar and turbulent regime is in (23):

$$Nu_m = 0.27Ra^{0.25} = 0.35Ra_{hf}^{0.2} \quad (23)$$

The axial oil temperature gradient in presence of cooling by a constant heat flux can be found out by using (24), due to[17] thus:

$$\frac{\partial T}{\partial y} = 4.25 \times 10^{-2} \frac{q_w}{k_{oil}} \left(\frac{1}{\pi D_m}\right)^{\frac{4}{9}} Ra_{hf}^{-\frac{1}{9}} \quad (24)$$

Where l is the winding height, $D_m$is the mean diameter of annular disc or layer of windings. Determination of boundary conditions in forced convection (OF mode) too requires calculating htc. The expression when the oil velocity is lower values, the mean Nusselt number based on temperature difference is of the form of (25), corresponding mean Nusselt number based on constant heat flux is of the form of (26):

$$Nu_m = 1.75\left[G_z + 0.012\left(G_z Gr^{\frac{1}{3}}\right)^{\frac{4}{3}}\right]^{\frac{1}{3}} \times \left(\frac{\mu_b}{\mu_w}\right)^{0.125} \quad (25)$$

$$Nu_m = 1.63\left[1.8{G_z}^{0.9} + 0.02\left(G_z Gr^{\frac{1}{3}}{}_{hf}\right)^{1.16}\right]^{\frac{1}{3}} \times \left(\frac{\mu_b}{\mu_w}\right)^{0.125} \quad (26)$$

$$G_z = RePr\frac{D}{l} \quad (27)$$

Where $G_z$calledGraetz number. The terms μ_b and μ_w are viscosity of oil computed at oil bulk mean temperature of oil and at winding wall temperature, respectively. R_e is the Reynolds number and Pr is the prandtl number. The relative importance of natural and forced cooling is indicated by the factor f_r=Gr/Re^2. If $f_r \cong 1$ then both of the cooling modes have to be considered. At a lower value of this factor, natural cooling can be ignored. The oil viscosity is an important property, which depends on temperature. The formula to take the oil temperature variation into consideration is given in (28) below from [9, 10]:

$$\mu = \alpha.exp\left(\frac{\gamma}{T_m}\right) \quad (28)$$

Where:

$$\alpha = 0.0000013573\ (kg.m^{-1}.s^{-1}) \quad (29)$$

$$\gamma = 2797.3\ (K) \quad (30)$$

The viscosity was calculated at the mean oil and wall temperature. Initially, the winding surface temperature is not known, so a starting guess for winding surface temperature has to be made, after calculating the value of h, the temperature difference (T_w-T_{oil}) is to agree with the assumed value. To calculate winding wall temperature at different surfaces, only bottom oil temperature is necessary. In this paper, the bottom oil rise over ambient temperature has been calculated as:

$$\theta_u = \theta_{fl}.\left(\frac{I_r^2 R+1}{R+1}\right)^n \quad (31)$$

Where μ_u is the bottom oil temperature rise over ambient Temperature $\mu_{fl}$is the full load bottom oil temperature rise over ambient temperature obtained from an off-line test;$R$is the ratio of load loss at rated load to no-load loss. The variable $I_r$is the ratio of the specified load to rated load:

$$I_r = \frac{I}{I_{rated}} \quad (32)$$

The exponent *n* depends upon the cooling state. The loading guide recommends the use of *n*=0.8 for natural convection and *n*=0.9-1.0 for forced cooling.
Assumptions have been made in the calculation of h_{tc} in the ducts provided in the disc-type windings under oil-forced (OF) modes of heat transfer (DOF and NDOF). The cooling in OF mode is due to mixed mode (natural and forced) convection. While the oil-flow velocity in both vertical and horizontal ducts has been assumed to be equal, the mode is DOF.

If the flow velocity in horizontal ducts is assumed negligible compared to velocity in the vertical ducts, the mode isNDOF. In case of DOF mode, the h_{tc} was assumed to be a function of both heat flux through the surface and the oil-flow velocity. The mechanism of heat transfer in this mode of cooling is same in both axial and radial direction of the disc. In case of NDOF mode, the h_{tc} in the vertical duct has been estimated by the same formula as for DOF. But the convection of heat in the horizontal ducts has been assumed to be a purely natural type.

III. SIMULATION AND RESULTS

The FEM of the transformer is performed in order to verify the effectiveness of the theoretical equations used in the design process and validate the design parameters.
The FEM in the case study involves these stages: magnetic and thermal field distribution within the transformer.

III.A. Design parameters of the transformer

To investigate the proposed approach, a three-phase transformer was designed. Solenoid-type windings are used for the conceptual design. The primary and secondary windings have 16 layers and each layer 45 turns. The main parameters of the transformer are shown in table 1.

TABLE I
DESIGN PARAMETERS OF TYPICAL TRANSFORMER WITH AUXILIARY WINDINGS.

Rating:	Voltage:	Current:	Frequency:	Phase:
10MVA	63/20 kV	90/290A	50 Hz	three-phase

III.B. Simulations of magnetic field

In this section, the simulation results obtained using COMSOL Multiphysics software is discussed. The color plotting of the magnetic field distribution in the case study obtained as a result of the simulation process is explained. The color plots of the magnetic field distribution obtained as a result of 2D simulations is shown in figure 1.
Referring to the numerical flux density values shown in the figure 1, the maximum value of the leakage magnetic field is 0.015 T.

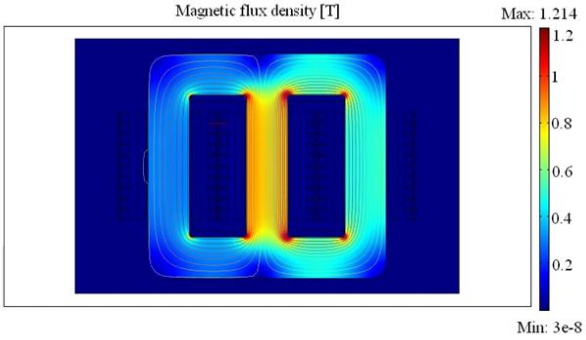


Fig. 1.Magnetic field distribution in typical transformer

III.C. Simulation of thermal field

The color plots of the thermal field distribution obtained as a result of 2D simulations is shown in figure 2.
It can be seen that, in figure 2, the maximum value of the temperature is 136ºC..

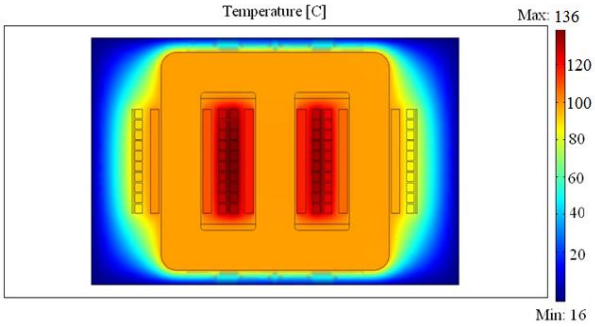


Fig. 2.Thermal field distribution in typical transformer

IV. CONCLUSION

In this paper, an attempt has been made to suggest a method to improve the accuracy of prediction of the temperature of the hottest spot in power transformer by solving the heat transfer partial differential equation (PDE) numerically.
The purely numerical approach for evaluating hot spot and its location followed in this paper seems to correspond reasonably well with the results of calculations and actual tests and on site measurements [9, 10]. The authors wish to point out that the IEEE loading guide and other similar documents offer relations for the calculation of the HST based on p.u. load. The formulations tend to ignore the possibilities of two transformers that are rating identical but have a different winding structure and varying heat loss/ unit volume. The method suggested by the authors gives due representation for this omission and, hence, is believed to give more accurate estimates. The thermal model presented here can predict the hot-spot location, with a reasonable degree of accuracy.

REFERENCES

[1] L. W. Pierce, 'Predicting liquid filled transformer loading capability', IEEE Trans. On Industry Applications, Vol. 30, No. 1, pp. 170–178,Jan./Feb. 1994.

[2] IEC Standard, Publication 354, Loading guide for oil immersedtransformers, second ed. (1991).

[3] IEEE Loading Guide for Mineral Oil Immersed Transformer, C57.91, pp.18–19, 46–53, 1995.

[4] IEEE Guide for Determination of Maximum Winding Temperature Risein Liquid-Filled Transformers, IEEE Std. 1538-2000, Aug. 2000.

[5] Z. Radakovic, 'Numerical determination of characteristic temperatures indirectly loaded power oil transformer', Eur Trans Electr Power 13:47–54, 2003.

[6] Z. Radakovic and Dj. Kalic, 'Results of a novel algorithm for thecalculation of the characteristic temperatures in power oil transformers', Electrical Engineering, Vol. 80, No. 3, pp. 205-214, Jun 1997.

[7] G. Swift, T. Molinski, W. Lehn, and R. Bray, 'A fundamental approachto transformer thermal modeling—Part I: Theory and equivalent circuit', IEEE Trans. on Power Delivery, Vol. 16, No. 13, Apr. 2001.

[8] W. H. Tang , O. H. Wu and Z. J. Richardson, 'Equivalent heat circuitbased power transformer thermal model', IEE Proc. Elect. Power Appl. Vol.149, No.2, March 2002.

[9] L. W. Pierce, 'An investigation of the thermal performance of an oilfilled transformer winding', IEEE Transaction on Power Delivery, Vol.7, No. 3, pp. 1347-1358, July 1992.

[10] M. K. Pradhan and T. S. Ramu, 'Prediction of hottest spot temperature (HST) in power and station transformers', IEEE Trans. Power Delivery, Vol. 18, No. 4, pp. 1275–1283, Oct. 2003.

[11] S. A. Ryder, 'A simple method for calculating winding temperaturegradient in power transformers', IEEE Trans. Power Delivery, Vol. 17, pp. 977–982, Oct. 2002.

[12] G. Swift and Z. Zhang, 'A Different Approach to Transformer ThermalModeling', IEEE Transmission and Distribution Conference, NewOrleans, April 12-16, 1999.

[13] Z. Radakovic and K. Feser, 'A new method for the calculation of thehot-spot temperature in power transformers with ONAN cooling' IEEE Trans. Power Delivery, Vol. 18, No. 4, pp. 1–9, Oct, 2003.

[14] M. A. Taghikhani, A. Gholami, 'Temperature distribution in power transformer windingswithNDOF and DOFcooling', Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE.

[15] Ma. Tsao-Tsung, 'A Novel Algorithm Based on Wavelet and Parallel Neural Networks for Diagnosing Power Transformer Status', International Review of Electrical Engineering (I.R.E.E.), Vol. 5, No. 3, 2010.

[16] [14] F. P. Incropera and D. P. DeWitt, "Fundamentals of Heat and MassTransfer." 4th ed., New York/USA: J. Wiley & Sons, 1996.

[17] Sykulski J K, Goddard K F and Stoll R L 1990 Hightemperature superconducting demonstration transformer:design considerations and first test results IEEE Trans.Magn. 35 3559–61.

Authors' information

1-DiakoAzizi was born in 1985. He has received B.Sc. degree in Electrical Engineering from Tabriz University, Tabriz, Iran in 2007. And he received the Master degree in Electrical Power Engineering from the University of Science and Technology, Tehran, Iran in 2009. He is presently pursuing the Ph.D. degree in Electrical Power Engineering, Iran University of Science and Technology. His research interests are aging of insulations in electrical machines.

2-Ahmad Gholami has received his B.Sc. Degree in electrical engineering from IUST, Tehran, Iran, in 1975, the M.Sc. and PhD. Degrees in electrical engineering from UMIST, Manchester, England, in 1986 and 1989 respectively. He is currently an associate professor in the Electrical Engineering Department of Iran University of Science and Technology. His main research activities are high voltage engineering, electrical insulation, insulation coordination, transmission lines and substations planning.

3-Diar Azizi was born in 1982 in Iran. He has received B.Sc. degree in Mechanical Engineering from the Iran University of Science and Technology, Tehran, Iran in 2005. He is presently pursuing the M.sc degree in Mechanical Engineering in University Science and Research branch, Islamic Azad University, Tehran, Iran. His research interests are Heat Transfer, Fluid Mechanic, and Thermodynamic.

www.ijmer.com Vol.1, Issue.2, pp-430-442 ISSN: 2249-6645

Study on the influence of spray parameters of Ar-N_2 Plasma spray process using CFD analysis

Mr. Mohammed Yunus[1], Dr. J. Fazlur Rahman[2]
[1]Research scholar, Anna University of Technology, Coimbatore,
Assistant Professor, Department of Mechanical Engineering
H.K.B.K.C.E. Bangalore
[2]Professor Emeritus, Department of Mechanical Engineering
H.K.B.K.C.E. Bangalore,

ABSTRACT

Thermal sprayed surface coatings are extensively used for a wide range of industrial applications. In the plasma spray process used in thermal spraying, the temperature of substrate and conduction of heat along the thickness of ceramic oxide coatings (TCOC) play an important role in the bond strength and microstructure of the coatings, which decides its performance and quality. In this study, using Ansys-CFX, a 3D numerical model is developed to study heat exchange between plasma jet and substrate and along the TCOC to predict life of the coating. The plasma jet temperature, velocity distribution and heat exchange to the substrate surface and coatings have been thoroughly analyzed for the effects of various spray parameters (SP) such as gas composition, standoff distances (sod), velocity and temperature of a jet. It is found that 3D modeling has shown promising results on substrate heating. The effect of spray parameters could also be assessed and validated by comparing with experimental results.

Keywords **- Computational Fluid Dynamics, heat flux, spray process, Numerical modeling, Partially Stabilized Zirconia, Temperature and velocity distribution Zirconia Toughened Alumina.**

1. INTRODUCTION

In ceramic oxide coatings, an atmospheric plasma spraying process (APS) is widely used. Plasma spraying has been extensively used in various

industrial components for producing different kinds of coatings, such as wear, corrosion, pitting and thermal resistance coatings [1], [3-4]. In this technique, plasma gas which is a high temperature ionized gas. When a strong electric arc is struck between tungsten electrode (cathode) and a nozzle (anode) in the presence of Argon and nitrogen/ hydrogen mixture in the chamber, the gas gets ionized which called plasma is reaching the temperature of the order of 14,000^{0}C to 20,000^{0}C. Injected particles of coating materials are heated inside the plasma jet and molten droplets are projected on the substrate with high velocities to form the coating.

The properties of the resulting layers are strongly depend on how large particle's velocity, temperature and melting degree are at the moment of impinging on the substrate[2],[5],[7]. Therefore, the manufacturing process requires the adjusting of a large number of parameters to get a good quality of coating (to suit our functional requirement) and for life and quality prediction of coating. Some of the parameters are

1. Flow rate, Gas composition, velocity and temperature of the plasma, substrate heating, type of coatings, standoff distance (distance between nozzle and substrate), size of the powder particles.
2. Conduction of heat in different types of coatings.

It is impossible in practice to determine the respective influence of all these single factors, on the resulting thermal barrier coating. Thus an accurate modeling of the whole plasma spraying process provides us with a powerful tool to understand better, the process and to determine the optimal conditions for the TBC production.

Much work has been done to study the effect of above parameters in the plasma jet without considering the presence of coating surface with substrate condition, using two and three-dimensional analysis [2],[7], [9]& [15-18]. It is found that the presence of coated substrate has major effect on these parameters. In order to ascertain its effect, the present study, based on numerical modeling analysis, has been done and the effect of various parameters have been discussed from the point of view of good quality coating.

Three different commercially available ceramic coatings powders namely, Partially Stabilized Zirconia (PSZ), Zirconia toughened alumina (ZTA consist of 80% alumina and 20% PSZ) and Super-Z alloy (20% alumina and 80% PSZ) were used for the coatings [1], [3] and [19].

1.1 Modeling approach

To simulate the plasma jet, it is assumed that plasma is in steady state, in local thermo-dynamical equilibrium, optically thin, Incompressible, turbulent and mass diffusivity is equal to thermal diffusivity [15-17]. The plasma jet is impinged on pre heated substrate (between 27^0C to 200^0C) in an open atmosphere and plasma gas used is a mixture of argon and nitrogen. To increase the enthalpy and thermal conductivity of the plasma jet, usually a small amount of nitrogen is added to argon [6-7] &[12]. Twelve different computational geometries are created for this study. Figure 1.(a), (b), (c), (d),(e), (f) respectively, shows computational domains for torch to substrate distances, 0.08, 0.1, 0.11, 0.14 and 0.15 m. These geometries are created using Ansys ICEM CFD10.0. The mesh is refined at core region of the jet to treat the large temperature and velocity gradients both in axial and in radial directions [7] & [9]. The nozzle exit diameter is 0.007 m. The Ansys-CFX 11.0 has been used for preprocessing, solving and post processing of results.

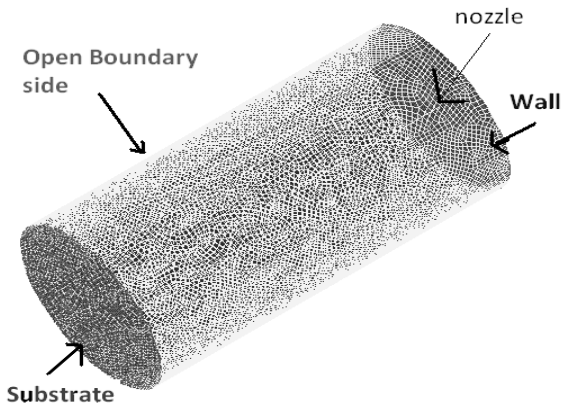
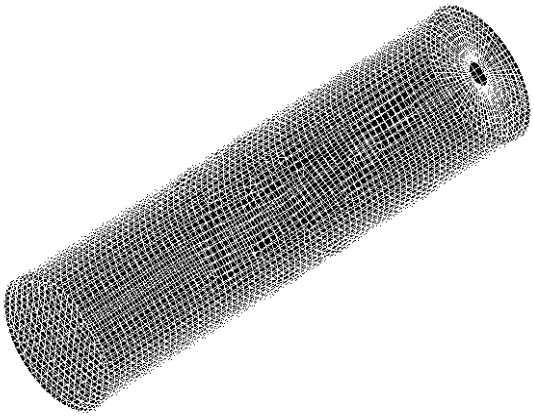


Figure 1. Computational geometry simulates the plasma jet with substrate at a standoff distance of 0.08m.

Figure 2. Computational geometry simulates the plasma jet with substrate at a standoff distance of 0.1m.

Figure 3. Computational geometry simulates the plasma jet with substrate at a standoff distance of 0.11m.

Figure 4. Computational geometry simulates the plasma jet with substrate at a standoff distance of 0.14m.



www.ijmer.com Vol.1, Issue.2, pp-430-442 ISSN: 2249-6645

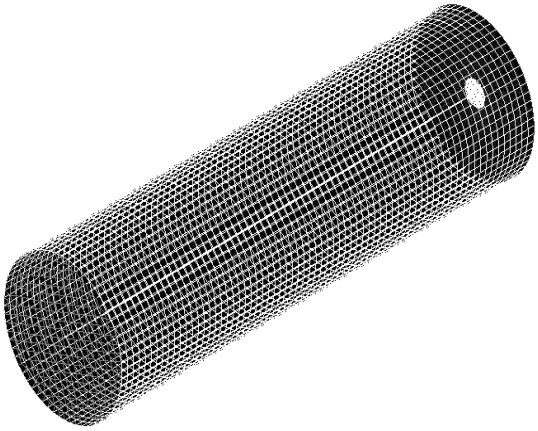


Figure 5. Computational geometry simulates the plasma jet with substrate at a standoff distance of 0.15m.

2. Results and Discussion

Analysis carried out for 16 kW plasma spray torch chosen to simulate the nozzle exit temperature and velocity profiles have been used to prepare different kinds of coatings such as Partial stabilized zirconia (PSZ), Super-Z and Zirconia Toughened Alumina (ZTA).

It is clear from this result that effect of nitrogen content in the plasma gas on torch power and efficiency is stronger than that of the argon gas flow rate. Hence, temperature and velocity of the plasma jet decrease with increasing argon gas flow rate. The similar effect has been seen for other stand-off distances.

The gas flow rates of both argon and nitrogen are fixed at three different values percentages of Ar90%, $N_2$10%, Ar 75%, N_2 25% and Ar50%, $N_2$50% respectively. The total heat flux to the substrate increases with increasing flow rate. The similar effect has been observed at stand-off distances of 0.1 and 0.125 m.

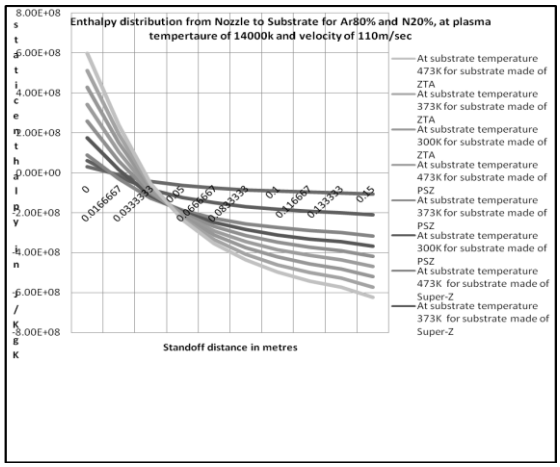


Fig.6.a. Enthalpy distribution for sod =0.15m at 110m/sec.

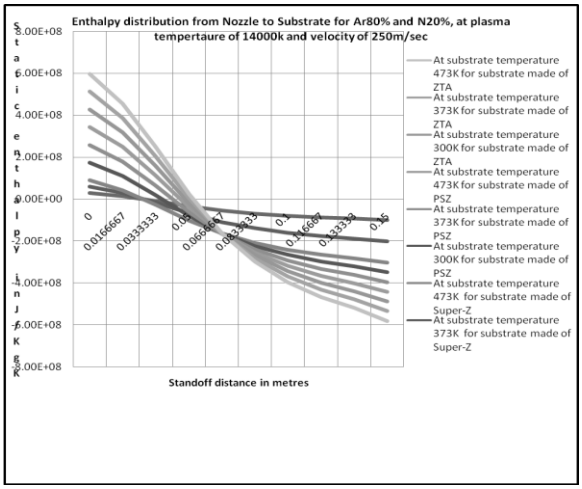


Fig.6.b. Enthalpy distribution for sod =0.15m at 250m/sec.

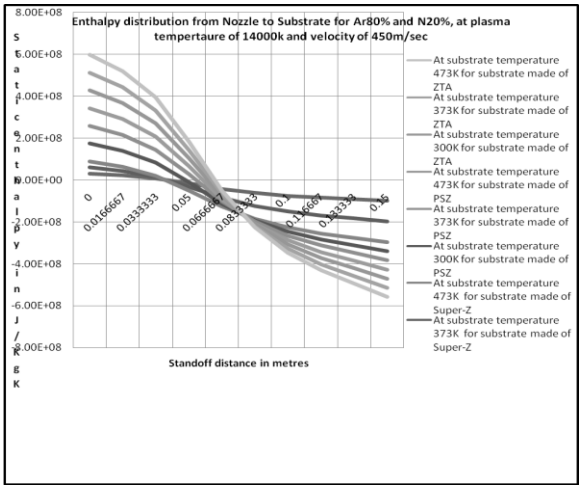


Fig.6.c. Enthalpy distribution for sod =0.15m at 450m/sec.

www.ijmer.com Vol.1, Issue.2, pp-430-442 ISSN: 2249-6645

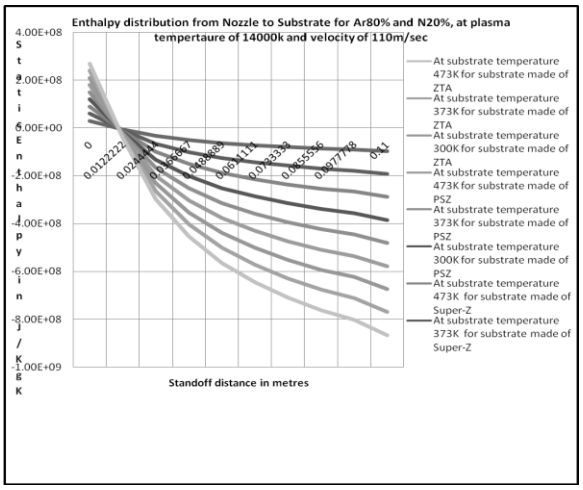


Fig.6.d. Enthalpy distribution for sod =0.11m at 110m/sec.

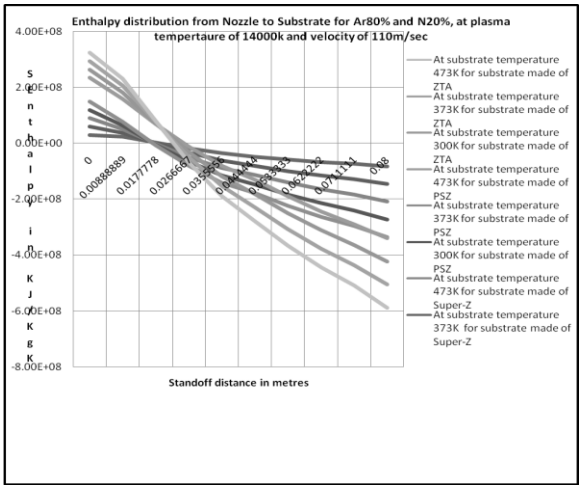


Fig.6.e. Enthalpy distribution for sod =0.08m at 110m/sec.

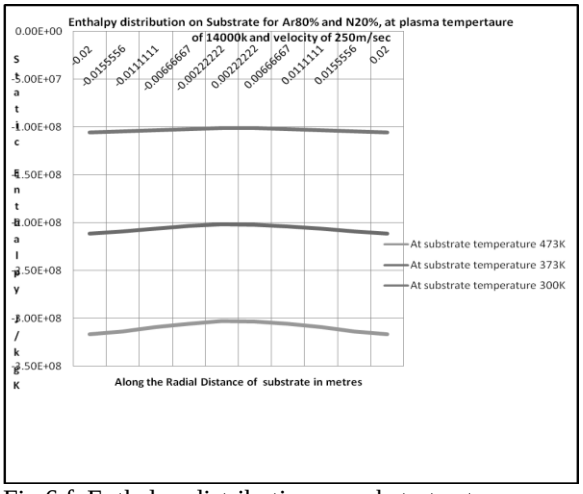


Fig.6.f. Enthalpy distribution on substrate at 250m/sec.

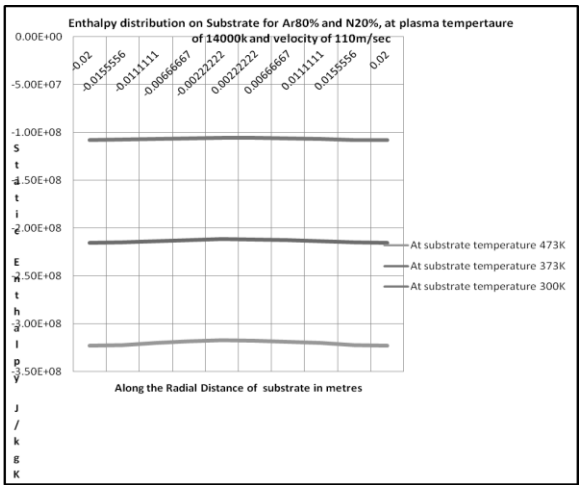


Fig.7.a. Enthalpy on substrate in radial direction at 110m/sec and sod = 80mm.

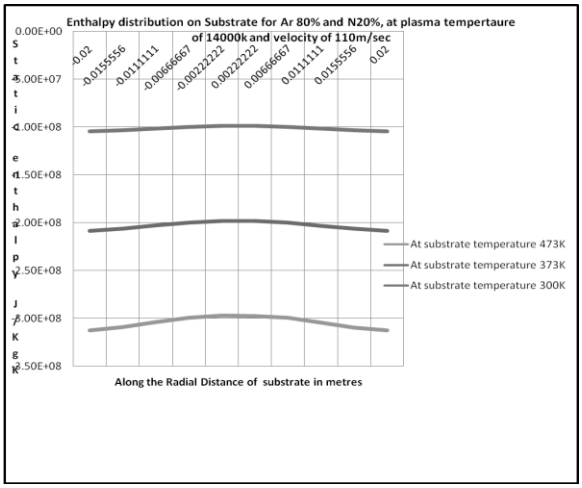


Fig.7.b. Enthalpy on substrate in radial direction at 110m/sec and sod = 100mm.

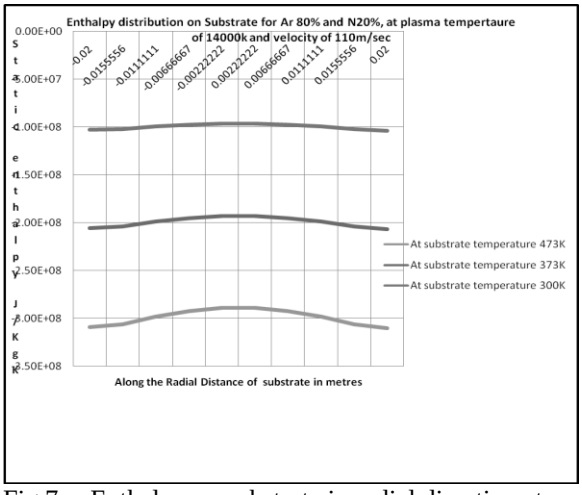


Fig.7.c. Enthalpy on substrate in radial direction at 110m/sec and sod = 110mm.

www.ijmer.com Vol.1, Issue.2, pp-430-442 ISSN: 2249-6645

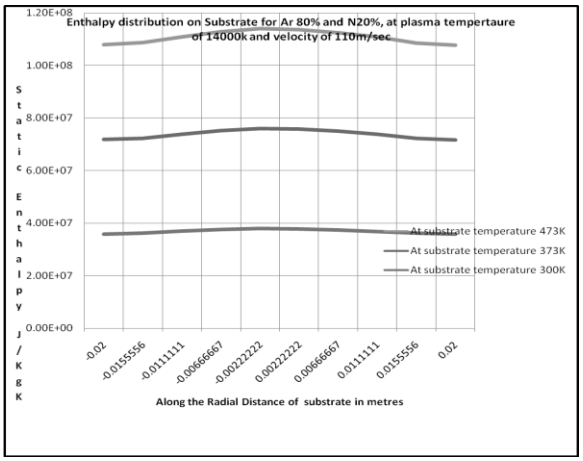


Fig.7.d. Enthalpy on substrate in radial direction at 110m/sec and sod = 140mm.

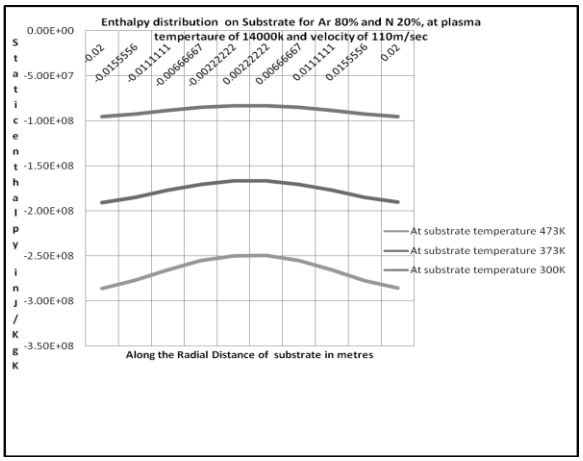


Fig.7.e. Enthalpy on substrate in radial direction at 110m/sec and sod = 150mm.

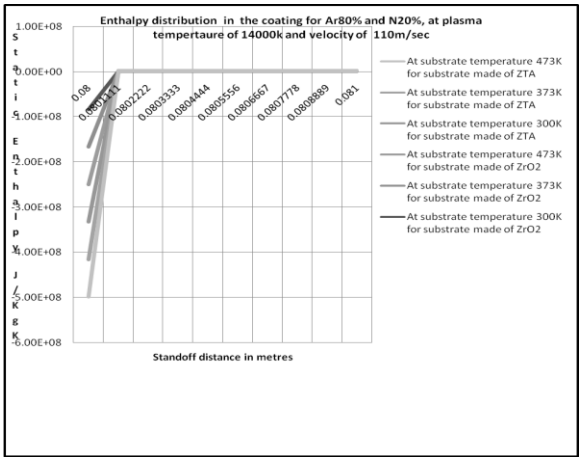


Fig.8.a. Enthalpy along coating thickness at 110m/sec and sod = 100mm.

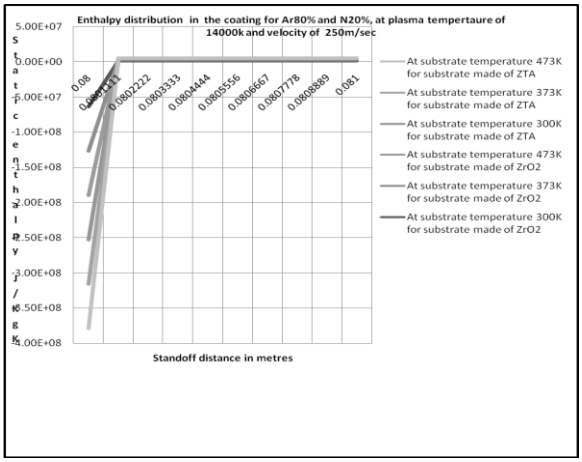


Fig.8.b. Enthalpy along coating thickness at 250m/sec and sod = 80mm.

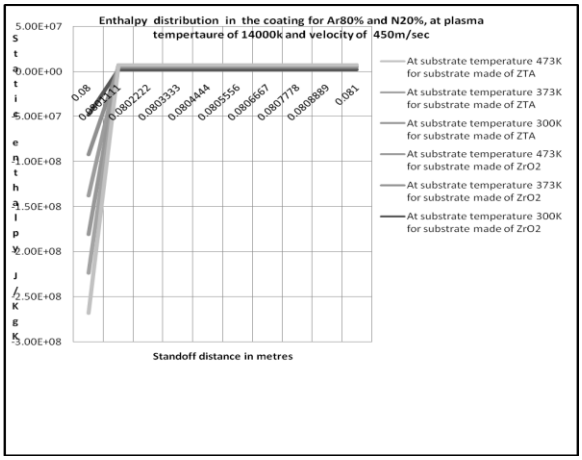


Fig.8.c. Enthalpy along coating thickness at 450m/sec and sod = 80mm.

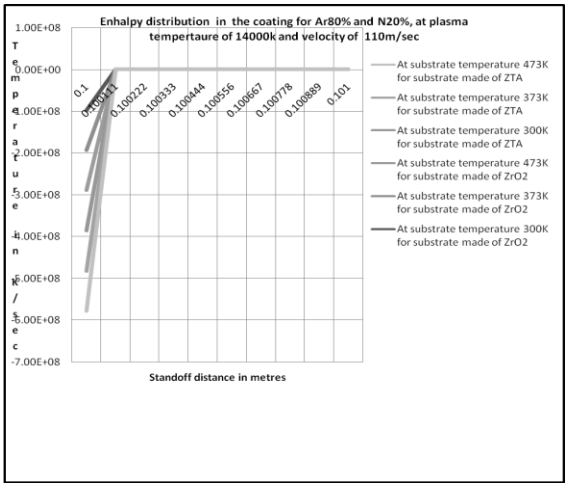


Fig.8.d. Enthalpy along coating thickness at 110m/secandsod=100mm.

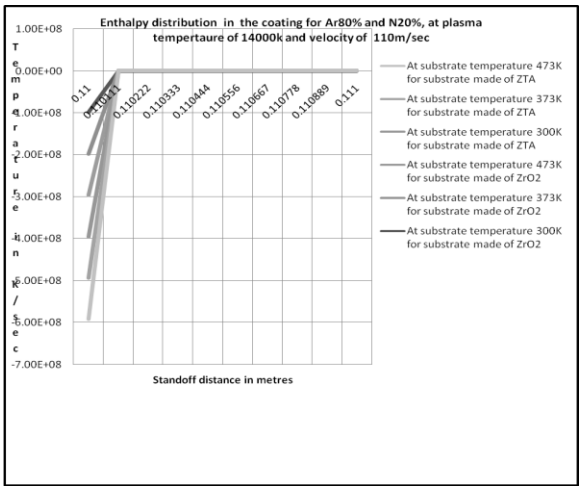


Fig.8.e. Enthalpy along coating thickness at 110m/sec and sod = 110mm

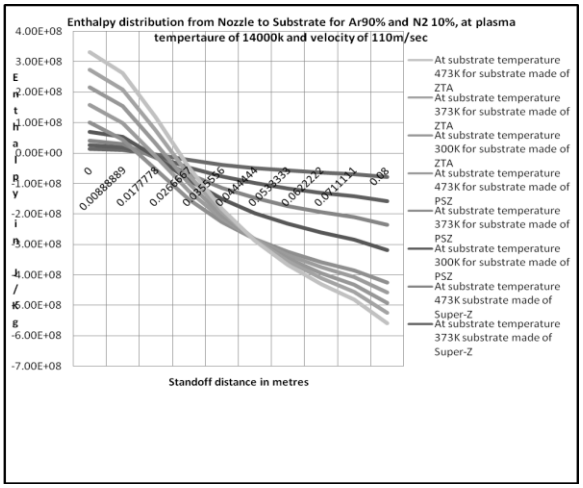


Fig.9.a. Enthalpy distribution for sod =0.08 m at 110 m/sec and Ar90%+$N_2$10%.

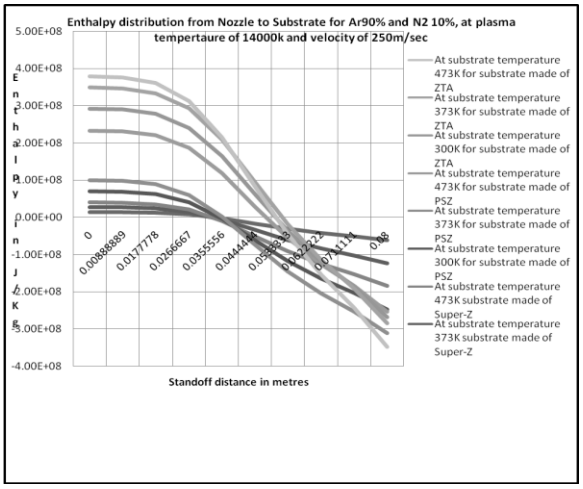


Fig.9.b. Enthalpy distribution for sod =0.08m at 250 m/sec and Ar90%+$N_2$10%.

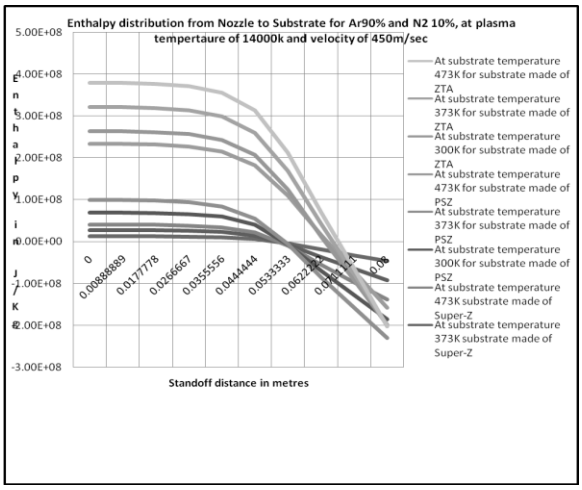


Fig.9.c. Enthalpy distribution for sod =0.08m at 450 m/sec and Ar90%+$N_2$10%.

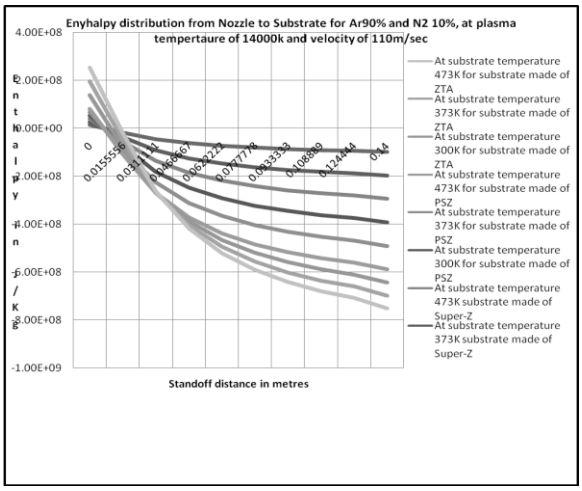


Fig.9.d. Enthalpy distribution for sod =0.14m at 110 m/sec and Ar90%+$N_2$10%.

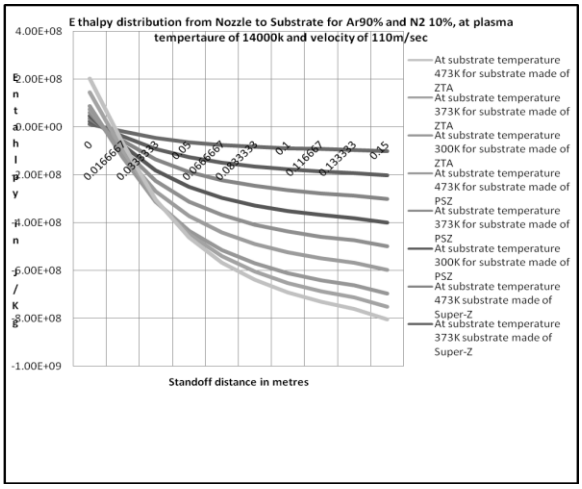


Fig.9.e. Enthalpy distribution for sod =0.15 m at 110m/sec and Ar90%+$N_2$10%.

www.ijmer.com Vol.1, Issue.2, pp-430-442 ISSN: 2249-6645

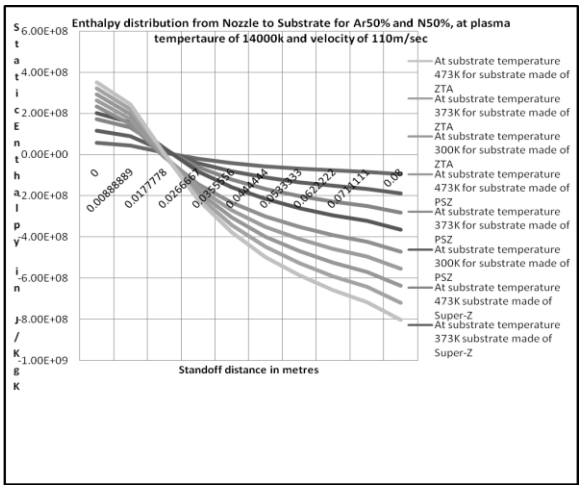


Fig.10.a. Enthalpy distribution for sod =0.08 m at 110 m/sec and Ar50%+$N_2$50%.

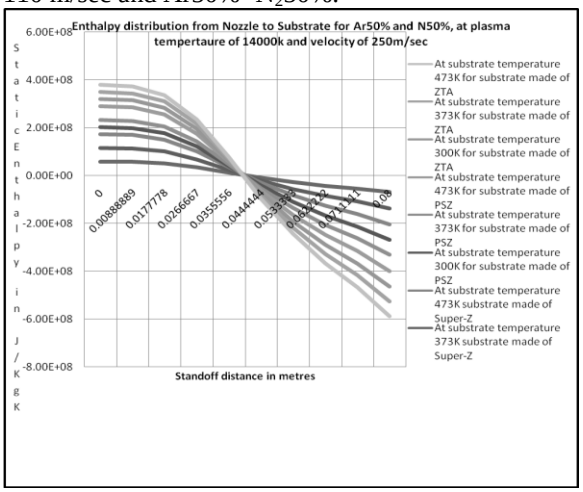


Fig.10.b. Enthalpy distribution for sod =0.08 m at 250 m/sec and Ar50%+$N_2$50%.

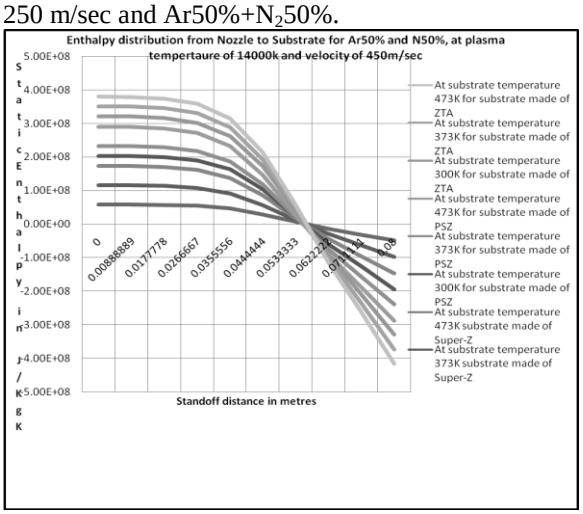


Fig.10.c. Enthalpy distribution for sod =0.08 m at 450m/sec and Ar50%+$N_2$50%.

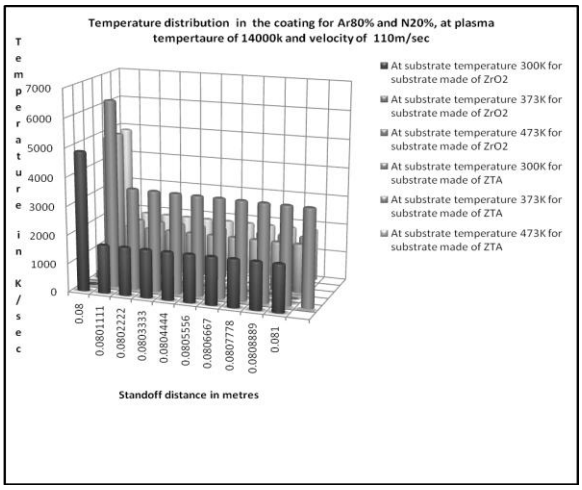


Fig.11.a. Temperature along the thickness of coating for sod =0.08 m at 110 m/sec and Ar50%+$N_2$50%.

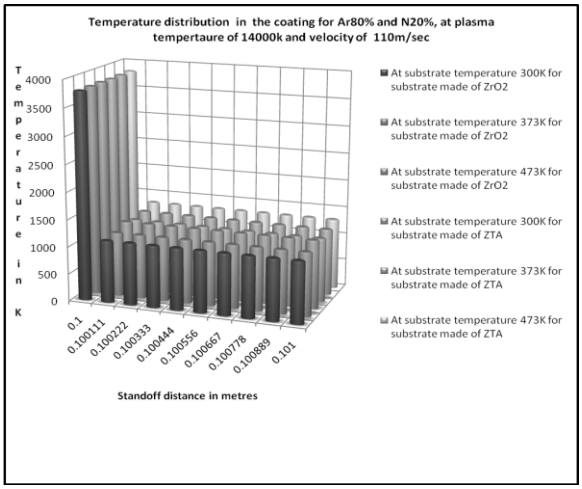


Fig.11.b.Temperature along the thickness of coating for sod =0.10 m at 110 m/sec and Ar50%+$N_2$50%.

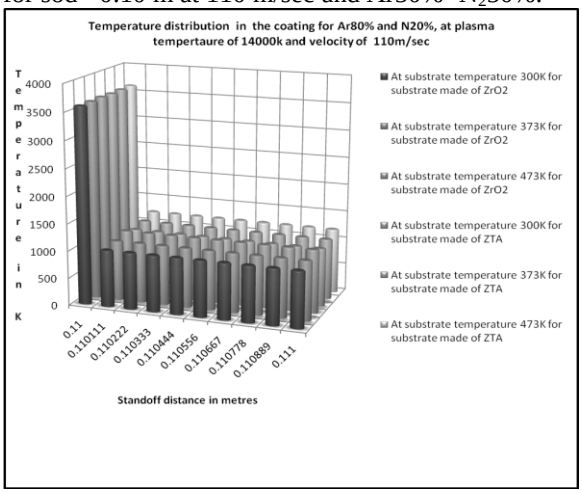


Fig.11.c.Temperature along the thickness of coating for sod =0.11 m at 110 m/sec and Ar 50%+N_2 50%.

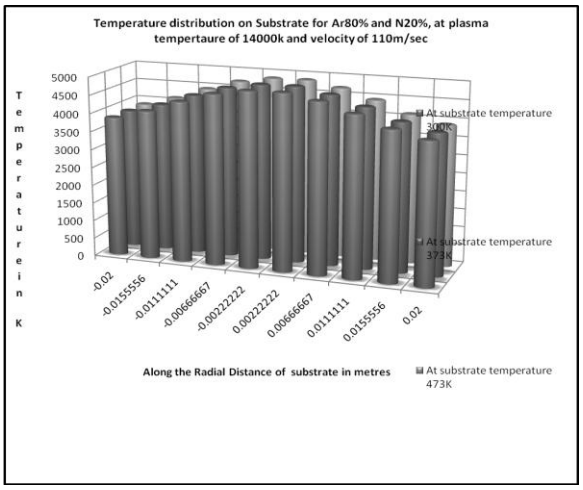


Fig.12.a.Temperature along the radial direction of substrate for sod =0.08 m. at 110 m/sec.

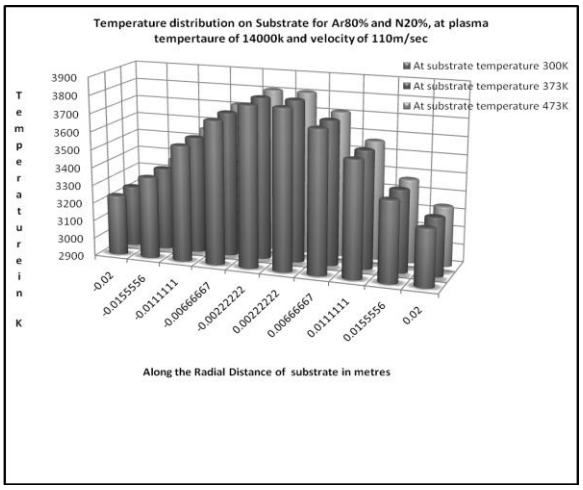


Fig.12.d.Temperature along radial direction of substrate for sod =0.1 m at 110 m/sec.

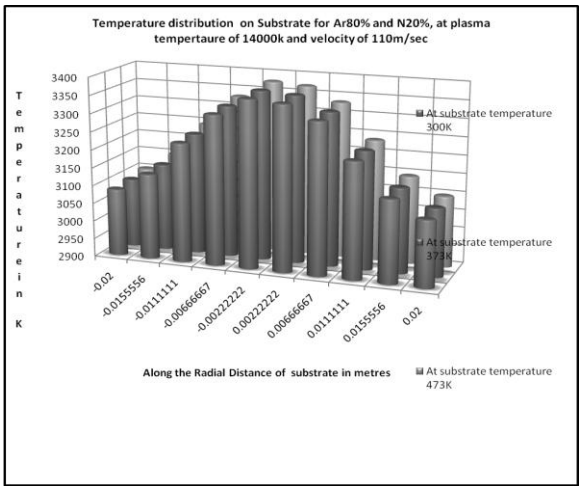


Fig.12.b.Temperature along radial direction of substrate for sod =0.11 m at 110 m/sec.

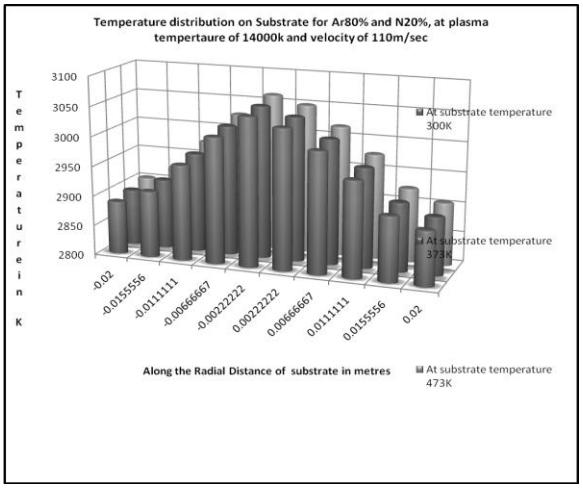


Fig.12.e.Temperature along radial direction of substrate for sod =0.15 m at 110 m/sec.

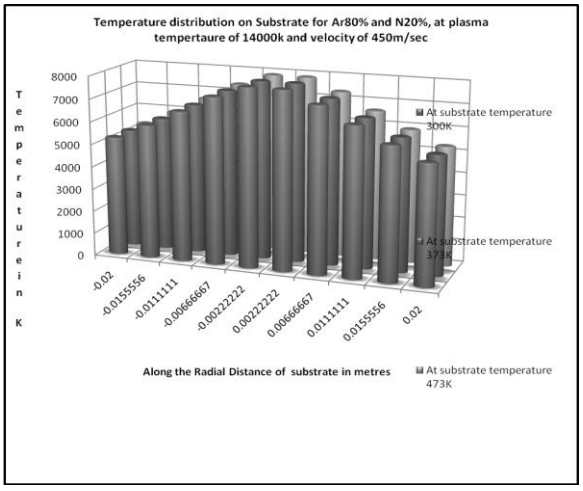


Fig.12.c.Temperature along radial direction of substrate for sod =0.08 m at 450 m/sec.

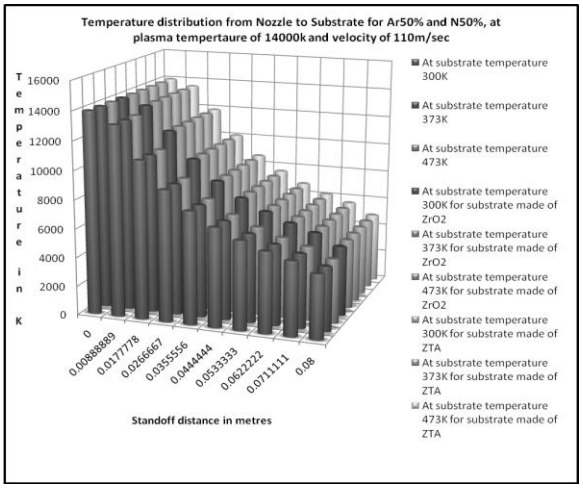


Fig.13.a.Temperature distribution for sod =0.08 m at 110 m/sec and Ar50%+$N_2$50%.

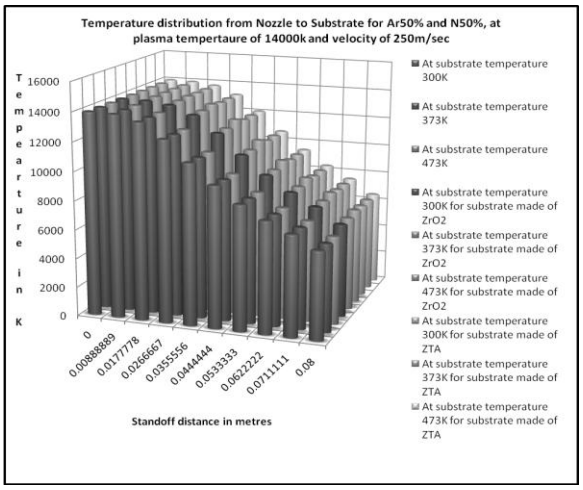


Fig.13.b.Temperature distribution for sod =0.08m at 250 m/sec and Ar50%+$N_2$50%.

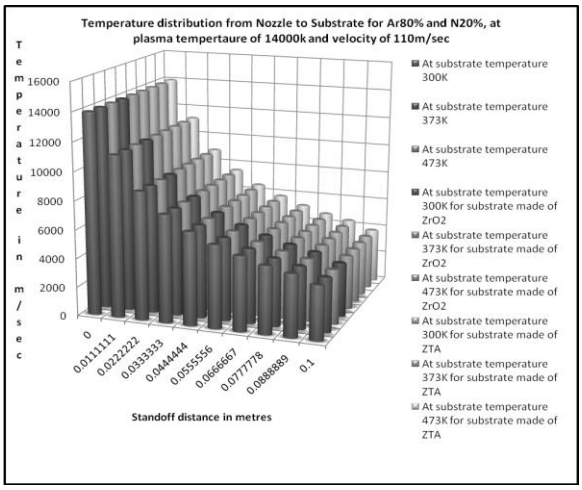


Fig.13.e.Temperature distribution for sod =0.1m at 110 m/sec and Ar80%+$N_2$20%.

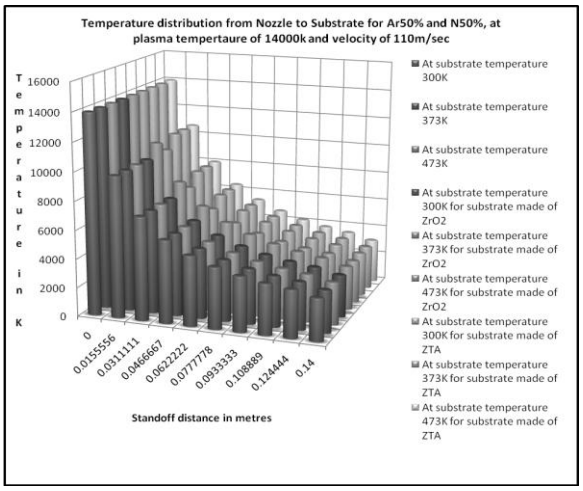


Fig.13.c.Temperature distribution for sod =0.14 m at 110 m/sec and Ar50%+$N_2$50%

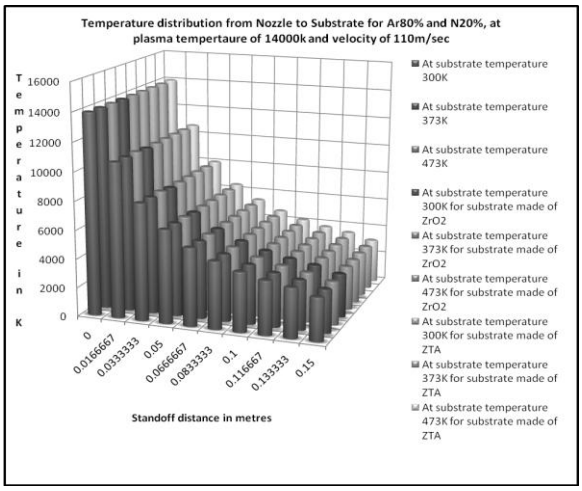


Fig.13.f.Temperature distribution for sod =0.15 m at 110 m/sec and Ar80%+$N_2$20%.

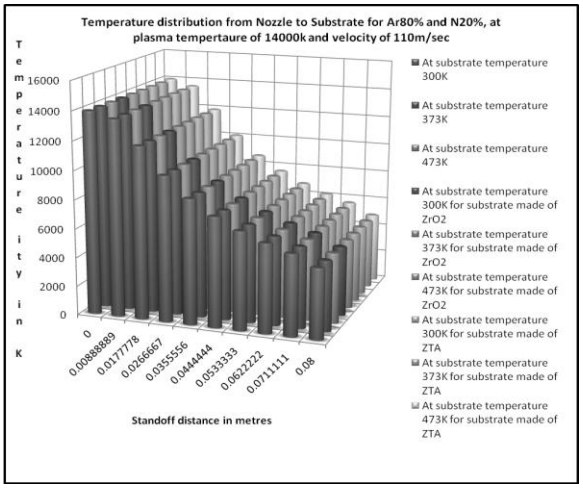


Fig.13.d.Temperature distribution for sod =0.08m at 110 m/sec and Ar80%+N220%.

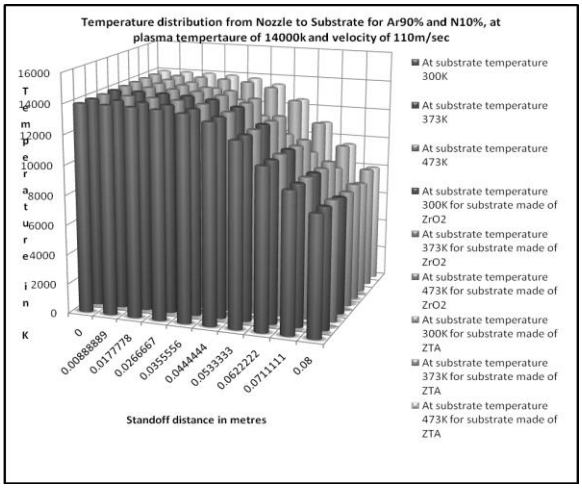


Fig.13.g.Temperature distribution for sod =0.08 m at 110 m/sec and Ar90%+$N_2$10%.

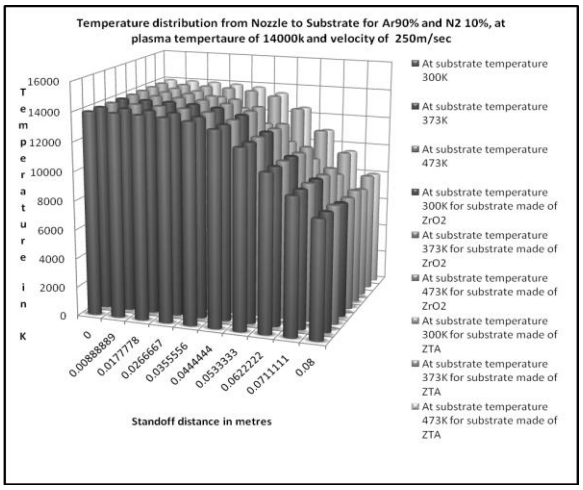


Fig.13.h.Temperature distribution for sod =0.08 m at 250 m/sec and Ar 90%+$N_2$10%.

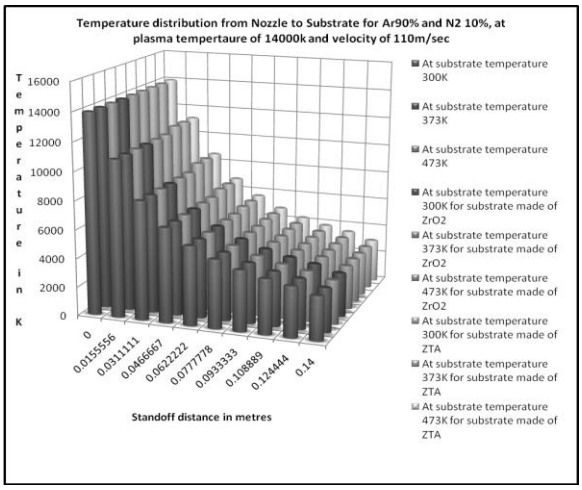


Fig.13.k.Temperature distribution for sod =0.14 m at 110 m/sec and Ar 90%+$N_2$10%.

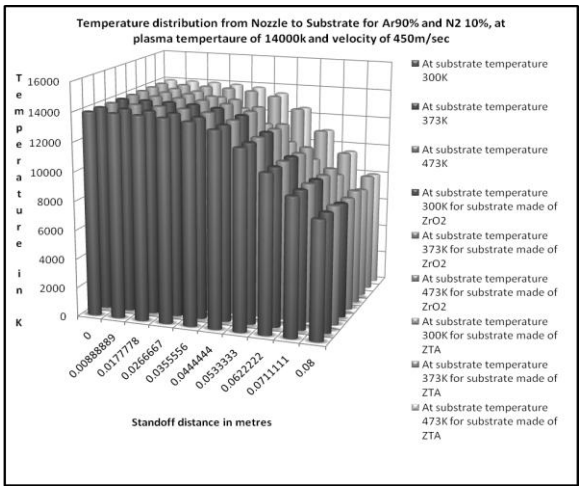


Fig.13.i.Temperature distribution for sod =0.08 m at 450 m/sec and Ar 9 0%+N210%.

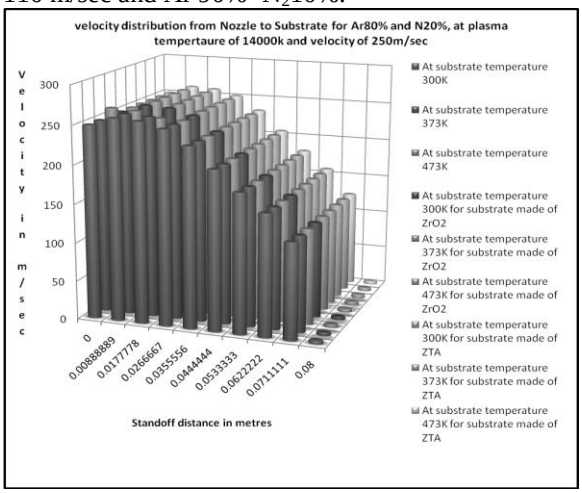


Fig.14.a.Velocity distribution for sod =0.08 m at 250 m/sec and Ar 80%+N220%.

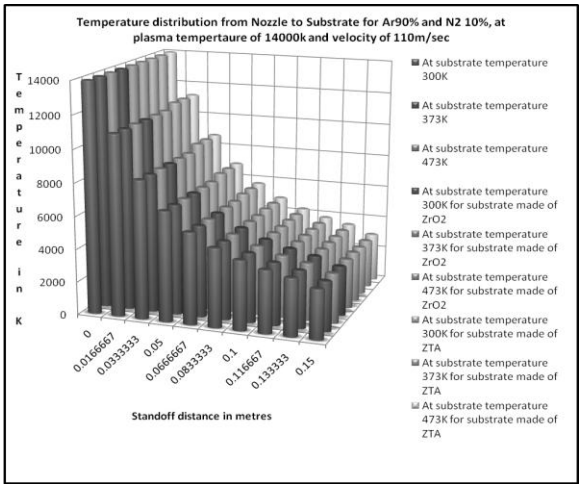


Fig.13.j.Temperature distribution for sod =0.15 m at 110 m/sec and Ar 90%+$N_2$10%.

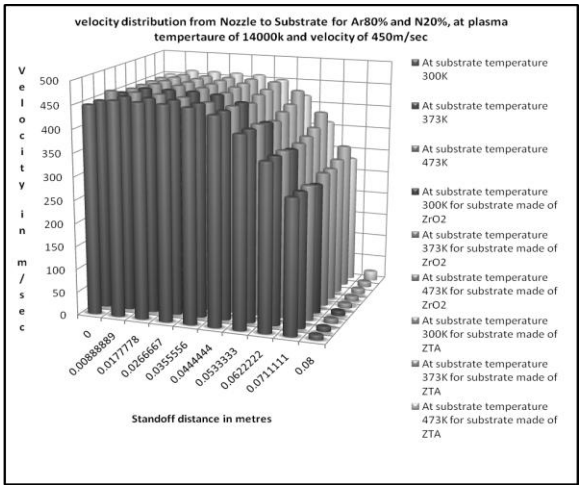


Fig.14.b. Velocity distribution for sod =0.08 m at 450 m/sec and Ar 90%+$N_2$10%.

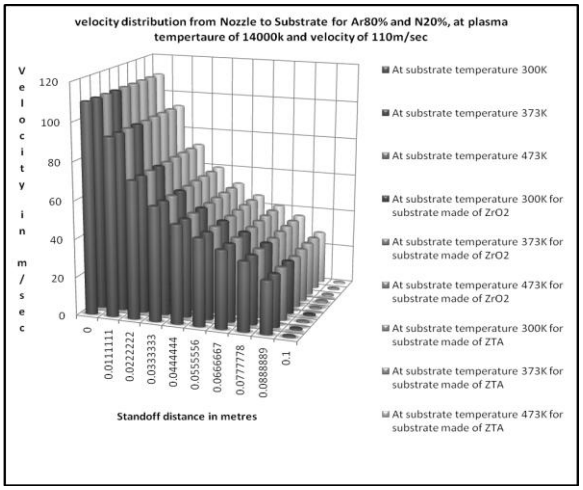


Fig.14.c. Velocity distribution for sod =0.1 m at 110 m/sec and Ar 80%+$N_2$20%.

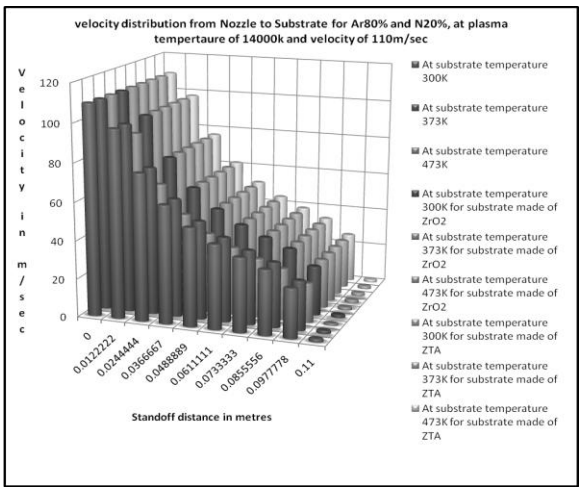


Fig.14.d. Velocity distribution for sod =0.11 m at 110 m/sec and Ar 80%+N220%.

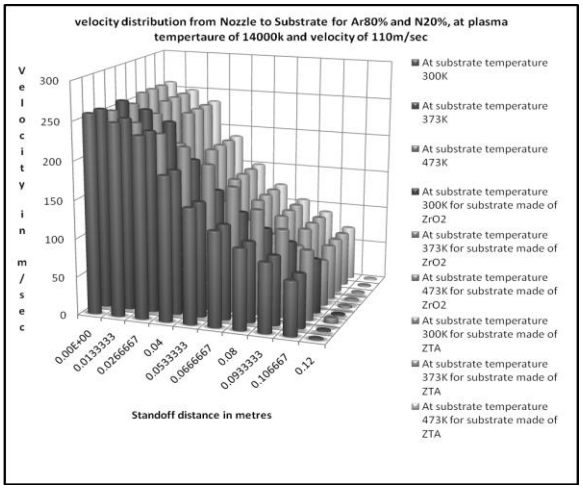


Fig.14.e. Velocity distribution for sod =0.12 m at 110 m/sec and Ar 80%+$N_2$20%.

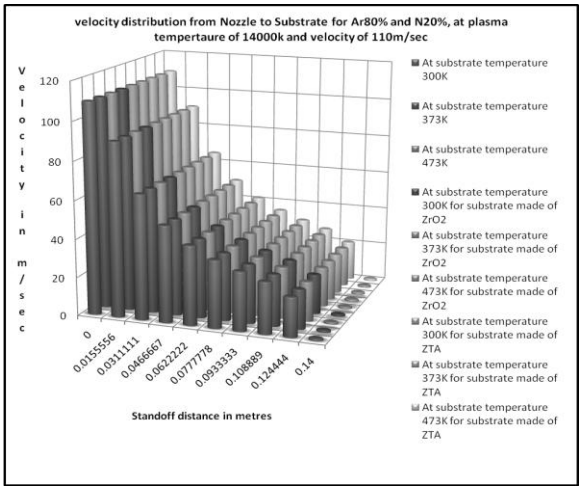


Fig.14.e. Velocity distribution for sod =0.14 m at 110 m/sec and Ar 80%+$N_2$20%.

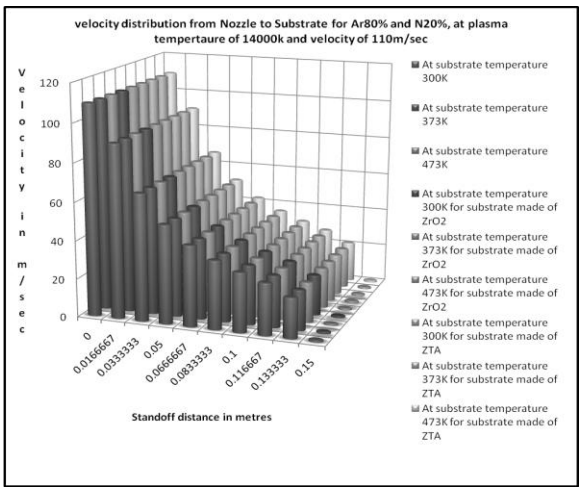


Fig.14.f. Velocity distribution for sod =0.15 m at 110 m/sec and Ar 80%+$N_2$20%.

The effect of gas flow rate on heat flux to the substrate at the stand-off distance of 0.08 m is shown in Fig. 8(a) to 8(e), 9(a) to 9 (e), 10 (a) to 10(c) and, 11(a) to 11 (c). Increasing gas flow rate decreases the heat flux to the substrate. The similar effect has been observed at stand-off distances of 0.1m and 0.15 m. As expected, the temperature observed at the center of the substrate is higher and the temperature is gradually decreasing along both radial and axial directions as shown in fig. 12(a) to 12(e), 13(a) to 13(k). The effect of stand-off distance on the temperature distribution in the substrate is significant. Similar results are obtained for all other cases. With increase in velocity, heat flux (enthalpy) more or less remain constant in case of super-Z, whereas in case

of ZTA and PSZ gradually reduces. Hence at higher velocities, super-Z can be preferred.

Thermo-fluid fields are strongly altered close to the substrate whereas the negligible effect of substrate is seen near the inlet. It is noted that the thermo-fluid fields are not symmetric due to application of the three-dimensional profiles at the inlet. Three-dimensional effect of nozzle exit profiles on the temperature and velocity fields of impinging jets shrinks along the axial directions due to the turbulent mixing of Ar-N_2 plasma with cold air. The similar effects have also been observed for other stand-off distance (0.1 and 0.13 m) as shown in fig. 14(a) to 14(f). For a given velocity with increase in substrate temperature, reduction in enthalpy imparted to substrate is marginal in case of super-Z and gradually increases in PSZ and ZTA as shown in fig. 6(a) to 6(e),. Hence Super-Z is preferred. Hence the order of preference is Super-Z, PSZ and ZTA.

The effect of stand-off distance on heat flux to the substrate at different radial distances for figures shown in 7(a) to 7(e). The heat flux to the substrate decreases with increasing stand-off distance and at the center is stronger and falls along the radial distance. There is no significant effect of stand-off distance on the heat flux to the substrate at the radial distance of 0.3 m. At longer radial distance, the difference between heat fluxes to the substrate reduces. The similar effect has been observed for other cases. Between 90mm to100mm standoff distances, the enthalpy imparted to the substrate gradually reduces upto 90mm and thereafter that is maintained constant upto 30mm. Hence it is inferred the standoff distance for getting good quality of coating between 100mm to 130mm.

The temperature and enthalpy variation along the radial direction of substrate remains more or less constant at standoff distances (less than 90mm) and decreases at increasing standoff distance in general. Reduction is more in case of enthalpy and temperature along radial direction decreases with decrease in substrate temperature. Enthalpy and temperature distribution along coating thickness remains constant at all standoff distances and substrate temperature and velocity of jets (which is because of micro level coating thickness).Marginal change in composition of gas does not have any effect on enthalpy distribution over coating surface remains unaltered.

With increase in percent of Argon in the gas mixture enthalpy imparts to the substrate is comparatively more than at lower compositions of Argon in gas mixture. The temperature obtained at surface coating in respect of super-Z is more compared to ZTA below Standoff distance 90mm and thereafter there is increase in ZTA compared to Super-Z.

3. CONLUSIONS

A three-dimensional numerical model is developed to simulate the Ar-N_2 plasma jet impinging on a flat surface coated with TBC coatings, namely Zirconia Toughened Alumina (ZTA), Partially Stabilized Zirconia (PSZ) and Super-Z. Since the arc root attachment at the anode creates three-dimensional effects both on temperature and velocity of plasma inside the torch, the three-dimensional feature extends to the plasma jet. However, this effect diminishes towards the substrate with different substrate temperatures. Plasma jet velocity is more sensitive to the gas flow rate than temperature especially near the inlet and the effect of gas flow rate on both temperature and velocity of the plasma jet diminishes along the axial direction. The atmospheric condition has strong influence on the heat transfer (rate of heat energy transfer through a given surface) between the plasma jet and the substrate. The effect of gas flow rate on the heat flux to the substrate decreases with decreasing the gas flow rate. The stand-off distance strongly controls the heat flux to the substrate at the center and is loosing its control on the heat flux to the substrate along the radial direction. The stand-off distance strongly influences the temperature distribution in the substrate. Modeling has been done with respect to above three coatings and for good quality of coating, the optimization of various process parameters has been done. This study is useful to understand the thermal exchange between the plasma jet and the substrate which decides the selection of TBC and optimization of process parameters for the production of a given surface condition based on engineering requirement.

REFERENCES

[1] Dr.J.Fazlur rahman and Mohammed Yunus, "Benefits of TBC Coatings on Engine applications", in *Proc. International conference, INCAM,* 2009 at Kalsalingam University, Tamil Nadu.

[2] R. Bolot, M. Imbert, and C. Coddet, "Use of a Low-Reynolds Extension to the Chen-Kim(k-ε) Model to Predict Thermal Exchanges in the Case of an Impinging Plasma Jet", *Int. J. Heat Mass Transfer, 44,* 2001, pp.1095-1106.

[3] Dr.J.Fazlur rahman and Mohammed Yunus, "Study of Mechanical and Tribological Characteristics of Tungsten Carbide-Cobalt HVOF Coating", *Proc. International conference, MEMS,* 2008 at AEC Bhatkal, Karnataka.

[4] Dr.J.Fazlur rahman and Mohammed Yunus, "Micro wave irradiation effect on ceramic composite coatings and carbide cutting tools", *Proc. National conference and proceedings, FTME- 2010,* at Gurunanak Dev college of Engineering, Ludhiana, Punjab.

[5] H.P. Li and E. Pfender, "Three Dimensional Modeling of the Plasma Spray Process", *J. Thermal Spray Technology, 16,* 2007, pp. 245-260.

[6] A.B. Murphy, "Transport Coefficients of Air, Argon-Air, Nitrogen-Air, and Oxygen-Air Plasmas" *Plasma Chem. Plasma Process.*, 1995, 15, pp. 279-307.

[7] Numerical Modelling of Ar-N_2 Plasma Jet Impinging on a Flat Substrate by B. Selvan, K. Ramachandran, B. C. Pillai and D. Subhakar *Journal of Thermal Spray Technology* ASM International, 2010.

[8] K.A. Khor, S. Jana, Pulse laser processing of plasma sprayed thermal barrier coating, *Journal of Materials Processing Technology, 66,* 1996, pp. 4-8.

[9] A.K. Ray, Characterization of bond coat in a thermal barrier coated super alloy used in combustor liners of aero engines, *conference in Materials Characterization 57,* 2006, pp. 199-209.

[10]G. Moskal, L. Swadźba, T. Rzychoń, "Measurement of residual stress in plasma-sprayed TBC with a gradient of porosity and chemical composition, *Journal of Achievements in Materials and Manufacturing Engineering, 23, 2,* 2007, pp. 31-34.

[11]Gorlach, I.A. High Velocity Thermal Gun for Surface\Preparation and Treatment, *South African Journal for Industrial Engineering,13, 1,* 2002, pp. 131 –143.

[12] Verstak, A., Baranovski, V. - *AC-HVAF Sprayed Tungsten Carbide: Properties and Applications*, *Proceedings of the International Thermal Spray Conference*, Seattle, Washington, USA, 2006.

[13]DeMasi, J.T., Sheffler K.D. and Ortiz M., 1989, "Thermal Barrier Coating Life Prediction Model Development," Phase-I, *Proc. Life Prediction of Functionally Graded Thermal Barrier Coatings*, 1989, NASA-182230.

[14] Xi Chen a, John W. Hutchinson , Anthony Evans G. "Simulation of the high temperature impression of thermal barrier coatings with columnar microstructure", *Journal of Acta Materialia, 52*, 2004, pp. 565–571.

[15] K. Ramachandran, N. Kikukawa, and H. Nishiyama, 3D Modeling of Plasma-Particle Interactions in a Plasma Jet Under Dense Loading Conditions, *Journal of Thin Solid Films, 435,* 2003, pp.298-306.

[16] B. Selvan, K. Ramachandran, K.P. Sreekumar, T.K. Thiyagarajan, and P.V. Ananthapadmanabhan, Three-Dimensional Numerical Modeling of an Ar-N_2 Plasma Arc Inside a Non-Transferred Torch, *Plasma Science Technology*, 2009, 11,pp.679-687.

[17] H. Fukanuma, R. Huang, Y. Tanaka, and Y. Uesugi, Mathematical Modeling and Numerical Simulation of Splat Cooling in Plasma Spray Coatings, *Journal of Thermal Spray Technology,* 2009,18, pp.965-974.

[18] H.R. Salimijazi, L. Pershin, T.W. Coyle, J. Mostaghimi, S. Chandra, Y.C. Lau, L. Rosenzweig, and E. Moran, Effect of Droplet Characteristics and Substrate Surface Topography on the Final Morphology of Plasma-Sprayed Zirconia Single Splats, *Journal of Thermal Spray Technology*, 2007, 16, pp. 291-299.

[19] Mohammed yunus, Dr.J. Fazlur rahman, "optimization of usage parameters of ceramic coatings in high temperature applications using Taguchi design" *International Journal of Engineering science and Technology, Vol.3(8),* 2011, pp.193-198.

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

The General Characteristic of Weak Intermolecular Interactions in Liquids and Crystals

Burhan Davarcioglu
Department of Physics, Aksaray University, Aksaray, Turkey
Email: burdavog@hotmail.com

ABSTRACT

A characteristic feature of weak interactions is the relative insensitivity of the bond energy on interatomic distance. The point is that in studying weak interactions one should not adopt a too stringent distance criterion in deciding what constitutes any given type of interaction. The types of weak interaction can arise from dipole-dipole interactions, quadrupole-quadrupole interactions, halogen-halogen interactions. We have seen that in crystals weak intermolecular interactions are strongly directional and that the mutual orientation of neighboring groups is important in achieving stable packing arrangements. Information about weak interactions in crystals is obtained from packing patterns which provide tests for the quality of atom-atom force fields. Whether a particular group of bonded molecules takes the form of a solid, liquid, or gas depends not only on the bonds that exist within each individual molecule, but also on the presence and type of bonds between molecules. Molecular substances are often soluble in organic solvents which are themselves molecular.

Keywords **– Hydrogen bonding, Interactions, Molecular crystals, Van der Waals, Weak intermolecular**

I. INTRODUCTION

Weak intermolecular interactions are not only important in supramolecular chemistry. Those hold the organic world together and are responsible for the very existence of liquids and crystals. For liquids, reliable structural information is hard to come by, but also extensive thermodynamic data are available for certain classes of compounds. Molecular substances are often soluble in organic solvents by molecular. Both the solute, the substance which is dissolving and the solvent are likely to have molecules attracted to each other by van der Waals forces. Although these attractions will be disrupted when they mix, they are replaced by similar ones between the two different sorts of molecules. Molecular substances will not conduct electricity. Even in cases where electrons may be delocalized within a particular molecule, there is not sufficient contact between the molecules to allow the electrons to move through the whole liquid or crystal.

Condensed media in which the distance between molecules is smaller than in gases, invariably show substantial interaction between molecules with saturated chemical bonds. The features of the intermolecular interaction determine the thermodynamic properties of liquids and the kinetics and mechanisms of the elementary chemical acts. The intermolecular interaction also control the formation of donor or acceptor complexes, and are responsible for the formation of colloidal systems. In biological systems intermolecular forces control the stability of all the compounds which are essential to life.

A phenomenon which has long been of interest and use to the coordination or acid-base chemist is the alteration of donor and acceptor vibrational frequencies upon formation of the coordinate bond. In addition to the structural applications of these shifts, qualitative, and in selected instances quantitative, estimates of interaction strengths have been determined from the magnitudes of the frequency shifts of normal vibrations involving the donor or acceptor atom. A few acids, those in which the acceptor site is a hydrogen atom, would seem to allow a quantitative and linear relationship between hydrogen stretching frequency shifts and enthalpy of adduct formation or base strength [1-3].

A molecular parameter more meaningful than frequency shifts as a criterion of the strength of a coordinate bond is the force constant for stretching of that bond. An alternative is to study the change in force constant of a bond adjacent to the coordinate bond. In the case of a donor frequency shifts, this potential constant is anticipated to be sensitive to the acidity of the acid and the nature of the bond which it forms with the donor [4]. A rough estimate of this sensitivity is indicated by the magnitude of the frequency change upon coordination as noted above. The typical interaction induced modification of a vibrational spectrum upon solution of a probe molecule in a solvent consists of vibrational frequency shifts appropriate to the selected vibrational mode and the interaction partners. Hydrogen bonded or weakly bound van der Waals complexes, being perturbed by a relatively inert environment, are known to exhibit in some cases pronounced intramolecular vibrational shifts which might enable one to derive information on the nature of the intermolecular perturbation. In recent years, a wealth of information has been collected on vibrational effects caused in monomeric units by the formation of weakly

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

bound dimmers in the gas phase. An even wider body of data exists which has been obtained by measuring vibrational frequency shifts of individual molecules or van der Waals and hydrogen bonded complexes trapped in low temperature matrices. Recently, the spectra of individual molecules and weakly bound complexes deposited on rare gas host clusters in molecular beams have become available.

Hydrogen bonding has emerged as the most important organizing principle not only for the structures of biologically important molecules but also for crystal engineering [5]. The hydrogen bonding forces a rather open structure on the ice; if you made a model of it, you would find a significant amount of wasted space. When ice melts, the structure breaks down and the molecules tend to fill up this wasted space. This means that the water formed takes up less space than the original ice. Ice is a very unusual solid in this respect most solids show an increase in volume on melting. When water freezes the opposite happens, there is an expansion as the hydrogen bonded structure establishes. Most liquids contract on freezing. Remnants of the rigid hydrogen bonded structure are still present in very cold liquid water, and don't finally disappear until 4 °C. From 0 °C to 4 °C, the density of water increases as the molecules free themselves from the open structure and take up less space. After 4 °C, the thermal motion of the molecules causes them to move apart and the density falls. That is the normal behaviour with liquids on heating. The conversion of a solid to a liquid is called either fusion or melting; the temperature at which this change occurs is the melting point. The quantity of heat required to melt a given amount of a solid is the enthalpy (heat) of fusion. A plot of temperature versus time as a solid is slowly heated is known as a heating curve; a similar plot for a liquid that is slowly cooled is known as a cooling curve. In some cases, it is possible to cool a liquid below its freezing point without having a solid form, a process known as super cooling. The conversion of a solid directly to a gas (vapor) is called sublimation. A plot of the vapor pressure of a solid versus temperature is known as a sublimation curve. The quantity of heat required to convert a given amount of solid directly to a gas is the enthalpy (heat) of sublimation.

The fluctuations in electron charge density in a molecule produce an instantaneous dipole, which in turn creates induced dipoles in neighboring molecules. The ease with which this occurs in a substance is known as its polarizability. Attractions between instantaneous and induced dipoles, called dispersion forces, are found in all substances. Polar molecules also have dipole-dipole and dipole-induced dipole intermolecular forces, arising from permanent dipoles in the molecules. Collectively known as van der Waals forces, dispersion forces, dipole-dipole forces, and dipole-induced dipole forces affect such physical properties as melting points and boiling points. A series of compounds with regularly varying structures and formulas also has regularly varying properties; this is the principle of homology [6].

When a substance is in the liquid state, its molecules or atoms are held together by mutual attraction; without these forces, the molecules or atoms would expand to fill all the space available, becoming a gas. However, these forces are not as strong as those holding solids together, giving liquid molecules or atoms freedom to move about. We can measure the strength of the intermolecular or interatomic forces in liquids through the boiling point: the more tightly the units are held together, the more heat energy will be needed to separate them into a gas. Conversely, if intermolecular attractions are weak, the boiling point will be low. There are four main types of intermolecular forces, from strongest to weakest: ion-dipole, dipole-dipole, dipole-induced dipole, and induced dipole-induced dipole. Collectively, they are referred to as van der Waals' forces, after the scientist who also investigated their effects on gases. When we refer to dipoles, we mean an electrically asymmetrical molecule. If a molecule is not electrically symmetrical, a positive charge will accumulate on one side, and a negative will build up on the other. Molecules with dipoles are said to be polar, nonpolar molecules do not have dipoles. Water is a polar molecule because the oxygen atom wants electrons more than the hydrogen, pulling the molecule's electrons towards the oxygen and creating a charge imbalance. Since electrical charges can attract or repel each other, dipoles are important in intermolecular forces.

The first type of attraction is ion-dipole. In these situations, a charged ion is attracted to the dipole of a polar molecule. These are by far the most powerful types of attraction. Examples include the dissolution of salt in water; the negatively charged Cl ions will be attracted to the positive dipole near the hydrogens, while the positively charged Na ions will seek the negative dipole of the oxygen atom. These attractions are powerful enough to tear apart the NaCl crystal when it enters water, meaning that salt dissolves. If the crystal structure is too strong to be broken by attractions between a solvent and the ionic solid, then the solid will not dissolve. The next type of attraction is dipole-dipole, in which the dipoles of two molecules are mutually attracted. For example, the molecule FI has a permanent dipole because fluorine wants electrons more than iodine, which leads to a positive charge on the iodine and a negative charge on the fluorine. These molecules will attract each other, because the negatively charged fluorine will be drawn to the positive iodine atom of another molecule. A subtype of dipole-dipole interactions is hydrogen bonding, which occurs when a nitrogen, oxygen, or fluorine atom is bonded with one or more hydrogens. Since each of these atoms wants electrons more than hydrogen, a dipole will result. These dipoles result in strong attractions between molecules. The intermolecular attractions in water (H_2O), methanol (CH_3OH), ammonia (NH_3), and hydrogen fluoride (HF) are all examples of hydrogen bonding. All

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

of these substances have conspicuously high boiling points, due to the unusual strength of hydrogen bonding. Another type of intermolecular attraction is dipole-induced dipole. This occurs when a polar molecule, such as water, is attracted to a nonpolar molecule, such as pentane (C_5H_{12}). This type of attraction occurs because the charge on the water molecule distorts the electron clouds of pentane's component atoms, causing them to be either attracted to the positive dipole or repelled by the negative dipole. These interactions are very weak, but explain the solubility of nonpolar compounds in polar solvents, such as oxygen dissolved in water. Note that the first three types of forces generally decrease with size; the larger the atoms or molecules, the less attractive the intermolecular forces will be. The final type of interaction, the induced dipole-induced dipole forces, is also known as dispersion forces or London forces. Since the electron clouds whirling about an atom are not always perfectly symmetrical, there will be occasional attractions between nonpolar molecules because of the anomalies in their electron clouds. These forces are usually weak, but increase with the atom's or molecule's sizes, and can become fairly strong in large molecules or atoms (because there are more opportunities for electron cloud distortion). Therefore, dispersion forces occur between all molecules, but are most evident in nonpolar molecules, because they are the only attractions holding the molecules together. Ion-ion attractive forces also occur when ionic solids are melted, and are even stronger than ion-dipole interactions. These forces only occur at very high temperatures; trying to melt an ionic solid requires a lot of energy. These interactions lead us to postulate a solubility law for liquids, like dissolves like. For example, polar liquids will dissolve other polar liquids because of dipole-dipole interactions, and nonpolar liquids will also usually dissolve nonpolar liquids due to dispersion forces. However, nonpolar liquids are not sufficiently attracted to polar liquids to break the strong dipole-dipole interactions between solvent molecules. Therefore, liquids with like polarities will dissolve, whereas a nonpolar liquid will not dissolve in a polar liquid. Intermolecular attractions have some other effects on liquids. Viscosity is a measure of how fluid, or runny, a liquid is. For example, water has low viscosity and runs easily. Cold maple syrup flows slowly, so it has high viscosity. Intermolecular forces play some role in viscosity, because stronger attractions between molecules cause them to resist flow more strongly. Molecule size is also an important factor in viscosity; longer molecules can become tangled and flow slowly. Surface tension is also a result of intermolecular forces. Molecules at the surface of a liquid are attracted to the molecules beneath and beside them, leading to an inward force on the liquid and a kind of skin on the surface. This tension also causes drops of water to contract into spheres, minimizing surface area [7].

All the different types of intermolecular interaction can be classified into two main groups: physical (determined by the physical characteristics of the interacting molecules) and chemical (responsible for the formation of directional chemical or quasi chemical bonds between molecules. Other workers classify the intermolecular interaction into volume interactions (in which each molecule is bound to surrounding molecules by a force inversely related to distance) and local interactions (in which two or more molecules are held together fairly strongly but the bonds between these groups of molecules have only secondary importance). Other distinguish between short range and long range order when classifying intermolecular forces. The former involves interactions described in terms of the physical parameters of the molecules, whereas the latter involves valency or chemical forces arising from overlap of the electron clouds of the interacting molecules [8-11].

Electrostatic interactions are dominant in ionic crystals, but are also very important in molecular crystals. The magnitudes of any localized charges in the latter are small, but the electrostatic energy is often still large relative to the energies of the van der Waals interactions. To a first approximation electrostatic interactions can be optimized by avoiding like-like interactions in favor of like-unlike interactions, variously depicted as bumps against hollows, donors against acceptors, positive ends of bond dipoles against negative ends. Favorable interactions among overall molecular dipoles are usually much less important than local interactions among bond dipoles [7]. In any case, the dimensions of most molecules are much larger than the shortest distances between molecules so that inferences based on the dipole-dipole approximation are invalid [12].

II. INTERMOLECULAR FORCES

Potential energy is stored whenever work must be done to change the distance between two objects. The attraction between the two objects may be gravitational, electrostatic, magnetic, or strong force. Chemical potential energy is the result of electrostatic attractions between atoms. Differences in the physical and chemical properties of substances are explained by the arrangement of the atoms, ions, or molecules of the substances and by the strength of the forces of attraction between the atoms, ions, or molecules. Atoms within a molecule are attracted to one another by the sharing of electrons, this is called an intramolecular force. The electrostatic forces that held molecules together are called intermolecular forces, and are in general much weaker than the intramolecular forces. Intermolecular forces are electrostatics in nature. It can be divided into: (permanent) dipole-(permanent) dipole interaction, dipole-induced dipole interaction (induction forces), instantaneous dipole-induced dipole interaction (London forces or dispersion forces), and hydrogen bonding. Typically, the first three forces are grouped together and called van der Waals' forces. Van

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

der Waals' forces exist between all molecules. Most of the intermolecular forces are identical to bonding between atoms in a single molecule. Intermolecular forces just extend the thinking to forces between molecules and follow the patterns already set by the bonding within molecules.

- **Ionic forces:** The forces holding ions together in ionic solids are electrostatic forces. Opposite charges attract each other. These are the strongest intermolecular forces. Ionic forces hold many ions in a crystal lattice structure.

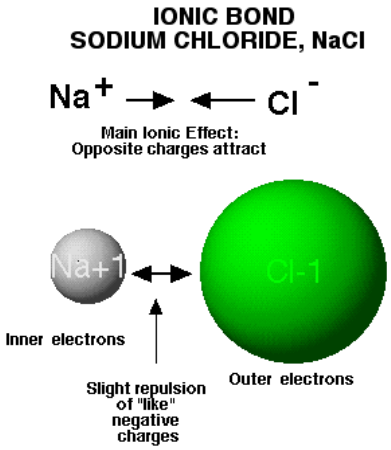


Figure 1. Ionic forces

- **Dipole forces:** Polar covalent molecules are sometimes described as dipoles, meaning that the molecule has two poles. One end (pole) of the molecule has a partial positive charge while the other end has a partial negative charge. The molecules will orientate themselves so that the opposite charges attract principle operates effectively. At the example in Fig. 2, hydrochloric acid is a polar molecule with the partial positive charge on the hydrogen and the partial negative charge on the chlorine. A network of partial + and – charges attract molecules to each other.

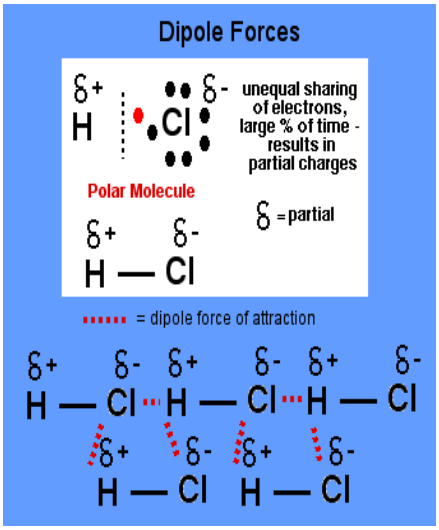


Figure 2. Dipole forces

- **Hydrogen bonding:** The hydrogen bond is really a special case of dipole forces. A hydrogen bond is the attractive force between the hydrogen attached to an electronegative atom of one molecule and an electronegative atom of a different molecule. Usually the electronegative atom is oxygen, nitrogen, or fluorine. In other words, the hydrogen on one molecule attached to O or N that is attracted to an O or N of a different molecule. In Fig. 3, the hydrogen is partially positive and attracted to the partially negative charge on the oxygen or nitrogen. Because oxygen has two lone pairs, two different hydrogen bonds can be made to each oxygen. This is a very specific bond as indicated. Some combinations which are not hydrogen bonds include: hydrogen to another hydrogen or hydrogen to a carbon.

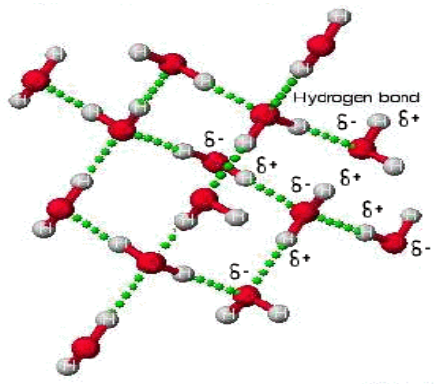


Figure 3. Hydrogen bonding

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

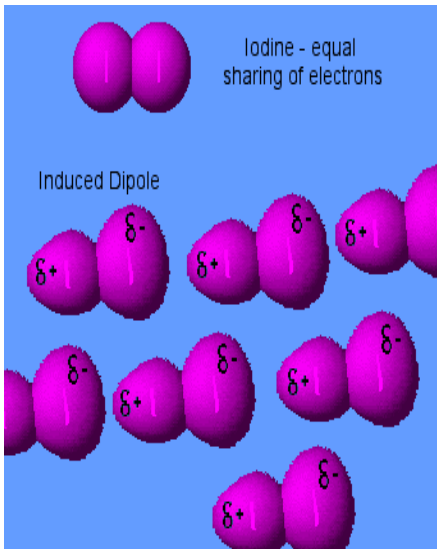


Figure 4. Induced dipole forces

- **Induced dipole forces:** Forces between essentially nonpolar molecules are the weakest of all intermolecular forces. "Temporary dipoles" are formed by the shifting of electron clouds within molecules. These temporary dipoles attract or repel the electron clouds of nearby nonpolar molecules. The temporary dipoles may exist for only a fraction of a second but a force of attraction also exists for that fraction of time. The strength of induced dipole forces depends on how easily electron clouds can be distorted. Large atoms or molecules with many electrons far removed from the nucleus are more easily distorted.

II.1. CLASSIFYING INTERMOLECULAR FORCES

In general, intermolecular forces can be divided into several categories:

1. Strong ionic attraction: Recall lattice energy and its relations to properties of solid. The more ionic, the higher the lattice energy. Ionic bonds are the result of electrostatic attraction between positive and negative ions. Ionic bonding is directly proportional to ionic charge and inversely proportional to ionic size.

2. Intermediate dipole-dipole forces: Substances whose molecules have dipole moment have higher melting point or boiling point than those of similar molecular mass, but their molecules have no dipole moment. Dipole-dipole interaction is the attraction between a partially negative portion of one molecule and a partially positive portion of a nearby molecule. Dipole-dipole interaction occurs in any polar molecule as determined by molecular geometry.

3. Weak London dispersion forces or van der Waal's force: These forces always operate in any substance. The force arisen from induced dipole and the interaction is weaker than the dipole-dipole interaction. In general, the heavier the molecule, the stronger the van der Waal's force or interaction. For example, the boiling points of inert gases increase as their atomic masses increases due to stronger London dispersion interactions. London dispersion forces result from instantaneous non permanent dipoles created by random electron motion. London dispersion forces are present in all molecules and are directly proportional to molecular size.

4. Hydrogen bond: Certain substances such as H_2O, HF, and NH_3 form hydrogen bonds, and the formation of which affects properties (melting point, boiling point, solubility) of substance. Other compounds containing OH and NH_2 groups also form hydrogen bonds. Molecules of many organic compounds such as alcohols, acids, amines, and amino acids contain these groups, and thus hydrogen bonding plays an important role in biological science [13]. Hydrogen bonding is significantly stronger than the dipole-dipole interactions which are in turn stronger than London dispersion forces. Hydrogen bonding exists only in molecules with an N-H, O-H, or F-H bond.

5. Covalent bonding: Covalent is really intramolecular force rather than intermolecular force. It is mentioned here, because some solids are formed due to covalent bonding. For example, in diamond, silicon, quartz etc., the all atoms in the entire crystal are linked together by covalent bonding. These solids are hard, brittle, and have high melting points. Covalent bonding holds atoms tighter than ionic attraction.

6. Metallic bonding: Forces between atoms in metallic solids belong to another category. Valence electrons in metals are rampant. They are not restricted to certain atoms or bonds. Rather they run freely in the entire solid, providing good conductivity for heat and electric energy. These behaviors of electrons give special properties such as ductility and mechanical strength to metals.

The division into types is for convenience in their discussion. All types can be present simultaneously for many substances. Intermolecular forces also play important roles in solutions, a discussion of which is given in hydration, solvation in water. A summary of the interactions is illustrated in the following Fig. 5.

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

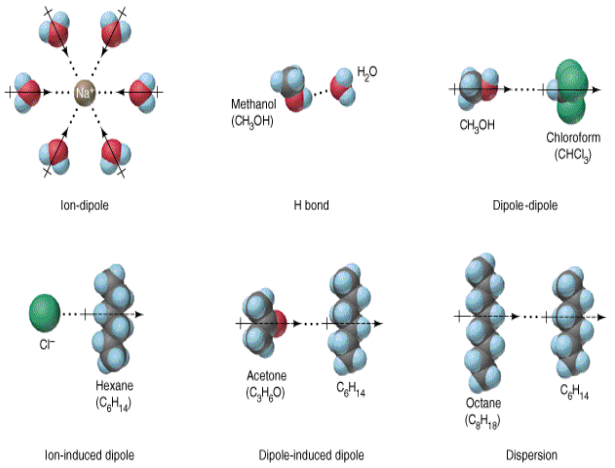


Figure 5. The classifying intermolecular forces

II.2. SUMMARY OF TYPES OF INTERMOLECULAR FORCES

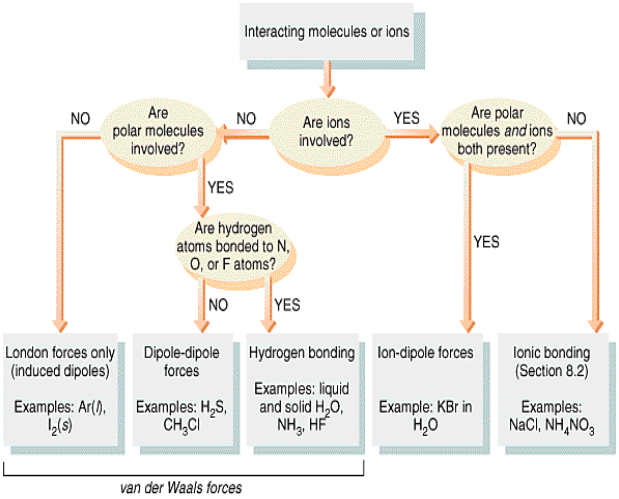
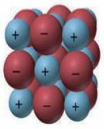
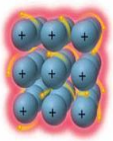


Figure 6. The types of intermolecular forces

The physical properties of melting point, boiling point, vapor pressure, evaporation, viscosity, surface tension, and solubility are related to the strength of attractive forces between molecules. These attractive forces are called intermolecular forces (Fig. 6). The strength of intermolecular forces present in a substance is related to the boiling point and melting point of the substance. Stronger intermolecular forces cause higher melting and boiling points.

Three types of force can operate between covalent molecules: dispersion forces also known as London forces (named after Fritz London who first described these forces theoretically 1930) or as weak intermolecular forces or as van der Waal's forces (named after the person who contributed to our understanding of non ideal gas behavior), dipole-dipole interactions, and hydrogen bonds.

Force	Model	Basis of Attraction	Energy (kJ/mol)	Example
Bonding				
Ionic		Cation–anion	400–4000	NaCl
Covalent		Nuclei–shared e^- pair	150–1100	H—H
Metallic		Cations–delocalized electrons	75–1000	Fe

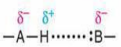
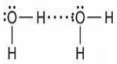
Nonbonding (Intermolecular)				
Ion-dipole		Ion charge–dipole charge	40–600	
H bond		Polar bond to H–dipole charge (high EN of N, O, F)	10–40	
Dipole-dipole		Dipole charges	5–25	I—Cl····I—Cl
Ion–induced dipole		Ion charge–polarizable e^- cloud	3–15	Fe^{2+}····O_2
Dipole–induced dipole		Dipole charge–polarizable e^- cloud	2–10	H—Cl····Cl—Cl
Dispersion (London)		Polarizable e^- clouds	0.05–40	F—F····F—F

Figure 7. Relative strength of intermolecular forces

Relative strength of intermolecular forces: intermolecular forces (dispersion forces, dipole-dipole interactions and hydrogen bonds) are much weaker than intramolecular forces (covalent bonds, ionic bonds or metallic bonds), dispersion forces are the weakest intermolecular force (one hundredth-one thousandth the strength of a covalent bond); hydrogen bonds are the strongest intermolecular force (about one-tenth the strength of a covalent bond), dispersion forces < dipole-dipole interactions < hydrogen bonds. Since van der Waals' forces are much weaker than covalent bond, ionic bond and metallic bond, only small amount of energy is needed to break the intermolecular forces of molecular substances. Molecular crystals or liquids are volatile, molecular crystals are soft and non conductors since there is no delocalized electrons.

The earth's crust may be held together mainly by ionic forces, molecules by covalent bonds, but it is weak intermolecular interactions which hold us along with the rest of the organic world together. The hydrogen bond is the best known example: because of its small bond

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

energy and the small activation energy involved in its formation and rupture, the hydrogen bond is especially suited to play a part in reactions occurring at normal temperatures. It has been recognized that hydrogen bonds restrain protein molecules to their native configurations. That as the methods of structural chemistry are further applied to physiological problems it will be found that the significance of the hydrogen bond for physiology is grater than that of any other single structural feature [14].

II.3. EFFECT OF INTERMOLECULAR FORCES ON SOLUBILITY

Solute is soluble in a solvent when there is a strong solute-solvent interactions (a force large enough to pull the solute particles away from each other). The following substances are soluble in water: ammonium, nitrate and sulphate salts, alkanols, carbonhydrates with low relative molecular mass since they form hydrogen bonds with water molecules. Most of the molecular substances are insoluble (or only very sparingly soluble) in water. Those which do dissolve often react with the water. Molecular substances are often soluble in organic solvents which are themselves molecular. Both the solute and the solvent are likely to have molecules attracted to each other by van der Waals' forces. In general like dissolves like:

- nonpolar solutes dissolve in nonpolar solvents paraffin wax ($C_{30}H_{62}$) is a nonpolar solute that will dissolve in nonpolar solvents like oil, hexane (C_6H_{14}) or carbon tetrachloride (CCl_4). Paraffin wax will not dissolve in polar solvents such as water (H_2O) or ethanol (ethyl alcohol, C_2H_5OH).

- polar solutes such as glucose ($C_6H_{12}O_6$) will dissolve in polar solvents such as water (H_2O) or ethanol (ethyl alcohol, C_2H_5OH) as the partially positively charged atom of the solute molecule is attracted to the partially negatively charged atom of the solvent molecule, and the partially negatively charged atom of the solute molecule is attracted to the partially positively charged atom of the solvent molecule. Glucose will not dissolve in nonpolar solvents such as oil, hexane (C_6H_{14}) or carbon tetrachloride (CCl_4).

- Ionic solutes such as sodium chloride (NaCl) will generally dissolve in polar solvents but not in nonpolar solvents, since the positive ion is attracted the partially negatively charged atom in the polar solvent molecule, and the negative ion of the solute is attracted to the partially positively charged atom on the solvent molecule.

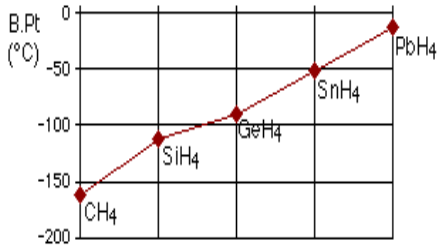


Figure 8. Plot the boiling points of the compounds with hydrogen

Since melting or boiling result from a progressive weakening of the attractive forces between the covalent molecules, the stronger the intermolecular force is, the more energy is required to melt the solid or boil the liquid. If only dispersion forces are present, then the more electrons the molecule has (and consequently the more mass it has) the stronger the dispersion forces will be, so the higher the melting and boiling points will be. Consider the hydrides of Group IV, all of which are nonpolar molecules, so only dispersion forces act between the molecules. CH_4 (molecular mass ~16), SiH_4 (molecular mass ~32), GeH_4 (molecular mass ~77) and SnH_4 (molecular mass ~123) can all be considered nonpolar covalent molecules. As the mass of the molecules increases, so does the strength of the dispersion force acting between the molecules, so more energy is required to weaken the attraction between the molecules resulting in higher boiling point [15]. The increase in boiling point happens because the molecules are getting larger with more electrons, and so van der Waals dispersion forces become greater (Fig. 8). Although for the most part the trend is exactly the same as in Group IV (for exactly the same reasons), the boiling point of the compound of hydrogen with the first element in each group is abnormally high (Fig. 9).

If a covalent molecule has a permanent net dipole then the force of attraction between these molecules will be stronger than if only dispersion forces were present between the molecules. As a consequence, this substance will have a higher melting or boiling point than similar molecules that are nonpolar in nature. Consider the boiling points of the hydrides of Group VII elements. All of the molecules HF (molecular mass ~20), HCl (molecular mass ~37), HBr (molecular mass ~81) and HI (molecular mass ~128) are polar, the hydrogen atom having a partial positive charge (H^+) and the halogen atom having a partial negative charge (F^-, Cl^-, Br^-, I^-) [8].

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

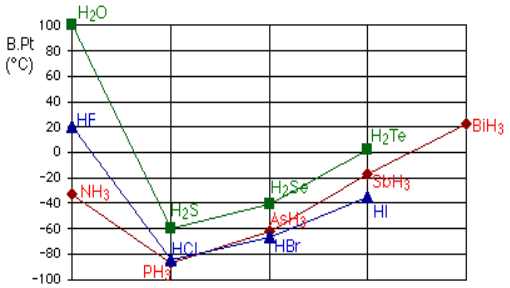


Figure 9. Boiling point of the compound of hydrogen with the first element

As a consequence, the stronger dipole interactions acting between the hydride molecules of Group VII elements results in higher boiling points than for the hydrides of Group IV elements. With the exception of HF, as the molecular mass increases, the boiling points of the hydrides increase. HF is an exception because of the stronger force of attraction between HF molecules resulting from hydrogen bonds acting between the HF molecules. Weaker dipole-dipole interactions act between the molecules of HCl, HBr and HI. So HF has a higher boiling point than the other molecules in this series. In the cases of NH_3, H_2O and HF there must be some additional intermolecular forces of attraction, requiring significantly more heat energy to break (Fig. 9). These relatively powerful intermolecular forces are described as hydrogen bonds.

II.4. ASPECTS OF WEAK INTERACTIONS

Until fairly recently, the term hydrogen bonding was more or less restricted to interactions involving F-H, O-H, and N-H as proton donors, O and N as proton acceptors with typical bond energies in the range 5 to 20 kcal/mol for F···H···F. However, it is now generally recognized that C-H···O interactions also play an important role in determining molecular packing arrangements [6]. They are admittedly weaker with energies typically less than 1 kcal/mol, but they are of common occurrence in molecular crystals containing C, H, and O atoms. Sometimes they have remarkable consequences. For example, eclipsed conformation of a $C(sp^3)$-CH_3 grouping in a crystalline trihydrate has been attributed to the presence of such C-H···O interactions in a cooperative system of O-H···O, O-H···N, and C-H···O hydrogen bond [6, 15].

Hydrogen bonding even the weak C-H···O interactions mentioned above can be described in good approximation as an essentially electrostatic phenomenon, i.e., as a first order coulomb interaction. Other kinds of weak interaction involve mutual polarization of molecules and dispersion forces. A characteristic of weak interactions is the relative insensitivity of the bond energy on interatomic distance. Elongation of a covalent bond by say 0.2 A˚ reduces the binding energy by something of the order of 50 kcal/mol (roughly half the standard bond energy). The same elongation of a typical hydrogen bond reduces the binding energy only by about 1.2 kcal/mol (roughly 25% of the standard binding energy, taken here as 5 kcal/mol). Typical potential energy curves for a OH···O(alcohol) hydrogen bonded interaction with well depth 5 kcal/mol and equilibrium distance 1.8 A˚ and for a non hydrogen bonded H···O interaction with well depth 0.12 kcal/mol and equilibrium distance 2.8 A˚, based on atom-atom potential parameters listed by Gavezzotti [15].

III. TYPES OF WEAK INTERACTION

Hydrogen bonding is undoubtedly the best studied and most important type of weak interaction. We interpret it here in its most general sense, as including all types of X-H···Y interactions. The types of weak interaction can arise from:

- **Dipole-dipole interactions:** The example is the structure of dimethylsulfoxide which is virtually determined by such interactions among the S=O bonds [16]. Interactions among overall molecular dipoles are usually much less important than local interactions among bond dipoles. Indeed, inferences based on interactions among molecular dipoles are not to be trusted. The repulsive nature of local dipole-dipole interactions across rotation axes and mirror planes helps to explain the low frequency of occurrence of these symmetry elements in molecular crystals [4, 7].

- **Quadrupole-quadrupole interactions:** The cubic crystal structures of carbon dioxide and acetylene [17] illustrate the favorable arrangement of like quadrupolar molecules in three dimensions. The molecules sit on the threefold axes of space group Pa3, such that each terminal atom in one molecule is equidistant from the centers of three surrounding molecules.

Benzene and hexafluorobenzene have almost the same molecular quadrupole moments but of opposite sign (around $30x10^{-40}$ Cm^2, negative for C_6H_6, positive for C_6F_6). The molecules are roughly the same shape and size, and since the interaction energy should not depend on the sign of the quadrupole moment, the quadrupole-quadrupole interaction alone should lead to the same stable packing arrangement for both [14]. Even for such small molecules, the packing optimizes local interactions at the expense of the global quadrupole-quadrupole interaction.

- **Halogen-halogen interactions:** The crystal structures of the halogens make an interesting series. They have the same space group and essentially the same layer packing arrangement as low temperature acetylene. In the halogen crystals the interlayer contact distances are roughly equal to the van der Waals diameters, but within the layers there is an increasingly strong tendency towards a specific, highly halogen-halogen interaction. For example, iodine has gone so far that the shortest

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

intermolecular I···I distance 3.50 A˚, or about 0.8 A˚ less than the van der Waals diameter. The iodine structure does not fit simple electrostatic models with charges on the atoms or interactions between molecular quadrupoles [18]. It looks as if the molecule is acting as an electron acceptor along the direction of the I-I bond and as electron donor perpendicular to the bond or vice versa, from the halogen crystal structures alone one can not decide.

Most organic molecules crystallize in low symmetry space groups and most elemental metals and many inorganic salts crystallize in high symmetry groups, but this difference is a consequence of the shapes of the packing units and of the presence or absence of strong electrostatic interactions rather than of any fundamental difference in the packing rules followed by organic and inorganic materials [7].

Highly directional interactions, the first such complex to be studied was 1:1 complex formed by molecular bromine and 1,4-dioxane. The remarkably short O···Br distance of 2.71 A˚ (compared with 3.35 A˚, sum of van der Waals radii) is along the direction of the Br-Br bond (at 2.31 A˚, slightly longer than the distance in gaseous bromine) and roughly in the direction expected for a tetrahedral lone pair on the oxygen atom. Clearly the bromine molecule is acting as an electron acceptor along the direction of the Br-Br bond [14]. From the extensive list of complexes studied, the same conclusion applies to the other halogens and also to halogen-C bonds.

The directionality of nonbonded contacts has been examined for several C-X systems and interpreted in terms of orientation dependent van der Waals surfaces, rotation ellipsoids with the short radius along the C-X bond [19]. However, such a picture would be complicated by the need to construct a different surface for different types of contact atoms, depending on their electron donor abilities. From recent work at the Cambridge Crystallographic Data Centre, C-Cl···O contact distances tend to be shorter in the C-Cl bond direction and longer perpendicular to it, but C-Cl···H distances show practically no orientation dependence. Similarly, electron donors as nucleophiles tend to approach divalent sulfur along one of the X-S directions whereas electron acceptors as electrophiles tend to approach nearly perpendicular to the X-S-Y plane [20].

Allowance for the highly directional nature of weak intermolecular interactions has hardly begun to be made in force fields for atom-atom potential calculations, which adhere, for the most part, to spherical atom models as far as the nonbonded atoms are concerned. Further information about such interactions should be important not only for crystal engineering but also for chemistry. Just as hydrogen bonding can tell us about acid-base relationships, so the wider study of how molecules approach one another in crystals can inform us about the incipient stages of chemical reactions in general.

III.1. THE EXAMPLES TO EXISTENCE OF INTERMOLECULAR FORCES

Examples to show that existence of intermolecular forces are very important.

- **Example 1. Hot pressing hair:** A metal pressing comb is heated and is passed quickly through the hair. The high temperature breaks the biochemical disulfide bonds between and within the keratin proteins and allows the hair to be straightened through the tension applied to the hair during the combing procedure. After the comb has passed through the hair the temperature drops rapidly and this allows the broken biochemical bonds in the hair to reconnect and fix their new position. This reformation of the bonds holds the hair in its new, straightened shape.

In the helical protein of hair, hydrogen bonds within individual helices of keratin and disulfide bridges between adjacent helices, impart strength and elasticity to individual hairs. Also, water can disrupt the hydrogen bonds making the hair limp. When the hair dries, new hydrogen bonding allows it to take on the shape of a curler. Permanent wave solutions induce new disulfide bridges between the helices. Genetically determined, natural curly hair also has a different arrangement of disulfide bridges compared with straight hair.

- **Example 2. Protein:** The primary structure of a protein is a polypeptide which is a polymer of amino acids. Polypeptide chains form a helical structure owing to the hydrogen bonds formed between the N-H and C=O groups. This creates the secondary structure of proteins. In many proteins, including those in hair wool and nails. Hydrogen bonding causes the polypeptide chains to become twisted into tightly coiled helices.

- **Example 3. DNA:** DNA is present in the nuclei of living cells and carries genetic information. The DNA molecule consists of two helical nucleic acid chains which is very stable. Each nucleic acid is made up of three components: a sugar, a phosphoric acid unit and a nitrogen containing heterocyclic base: adenine, cytosine, guanine or thymine. The two nucleic acid chains are held together by hydrogen bonds. These hydrogen bonds are formed between specific pairs or bases on the chains. The two strands coil tightly around each other.

Protein and DNA is very important to our lives: Enzyme controls the metabolic reaction. Proteins also act as a cytoskeleton, membrane proteins, a raw materials for growth, for movement myosin form the basic structure of muscles, osmotic balance and buffering and energy source. The hydrogen bonds between the base pairs tend to drive double helix to reform spontaneously after uncoiling in replication and transcription.

- **Example 4. Plant:** In a narrow capillary of e.g. cellulose there are many oxygen atoms on the surface for hydrogen bonding to the water.

www.ijmer.com Vol.1, Issue.2, pp-443-454 ISSN: 2249-6645

- **Example 5. Calcium sulfate ($CaSO_4$):** Gypsum (hydrated $CaSO_4$, $CaSO_4.H_2O$). Layer of ions are attracted by hydrogen bonds, it is soft and can be cleaved easily. Anhydrite (anhydrous $CaSO_4$) no hydrogen bonds, there are only ionic bonds between ions. It is very hard and very difficult to cleave. Plaster of Paris, a fine white powder is produced by heating gypsum to expel the water. If this powder is moistened and then allowed to dry, it becomes hard or sets. Its major use is in statuary, ceramics, dental plates, fine metal parts for precision instruments and surgical splints.

- **Example 6. Soap and detergent:** Molecules liquid state experience strong intermolecular attractive forces. When those forces are between like molecules, they are referred to as cohesive forces. The molecules of water droplet are held together by cohesive forces and the especially strong cohesive forces at the surface to form surface tension. So, surface tension is a type of intermolecular forces. Soap and detergents help the cleaning of clothes by lowering the surface tension of the water so that it more readily soaks into pores and soiled areas.

IV. LIQUIDS AND CRYSTALS

Walking on water small insects such as the water strider can walk on water because their weight is not enough to penetrate the surface due to the surface tension presented. Washing with cold water the major reason for using hot water for washing is that its surface tension is lower and it is a better wetting agent. But if the detergent lowers the surface tension, the heating may be unnecessary. Example in depth to show the significance of existence of intermolecular forces is water and ice. There are lots of different ways that the water molecules can be arranged in ice. The one below is known as cubic ice or "ice Ic". It is based on the water molecules arranged in a diamond structure. Cubic ice is only stable at temperatures below -80 ºC. The ice you are familiar with has a different, hexagonal structure. It is called "ice Ih".

The hydrogen bonding forces a rather open structure on the ice. When ice melts, the structure breaks down and the molecules tend to fill up this wasted space. This means that the water formed takes up less space than the original ice. Ice is a very unusual solid in this respect most solids show an increase in volume on melting [6, 7].

Why does ice float on water? Hydrogen bonding as a water molecule is composed of two hydrogen atoms and one oxygen atom. The atoms of hydrogen and oxygen are bound by sharing their electrons with one another. This bond is called a "covalent bond". However, since oxygen atoms pull electrons more strongly than hydrogen atoms, the oxygen atom in a water molecule has a slightly negative charge and the hydrogen atoms have a slightly positive charge. So adjacent water molecules are attracted to one another through the slightly negatively charged oxygen atoms and the slightly positively charged hydrogen atoms. This interaction is called "hydrogen bonding". Hydrogen bonding is much weaker than covalent bonding, however, this type of bonding has a large total effect because there are so many hydrogen bonds. The bonds in water molecules are inclined at a tetrahedral angle of 109º [9]. The lone pairs occupy the other corners of the tetrahedron. Liquid water contains associations of water molecules. In ice the arrangement of water molecules is similar, but the regularity extends throughout the whole structure. The structure spaces the molecules further apart than they are in the liquid. This is why when water freezes, it expands (by 9%), and ice is less dense than water.

Table 1. Types of crystals and general properties.

Type of Crystal	Force(s) Holding the Units Together	General Properties	Examples
Ionic	Electrostatic attraction	Hard, brittle, high melting point, poor conductor of heat and electricity	NaCl, LiF, MgO, $CaCO_3$
Covalent	Covalent bond	Hard, high melting point, poor conductor of heat and electricity	C (diamond), SiO_2 (quartz)
Molecular	Dispersion forces, dipole-dipole forces, hydrogen bonds	Soft, low melting, poor conductor of heat and electricity	Ar, CO_2, I_2, H_2O, $C_{12}H_{22}O_{11}$
Metallic	Metallic bond	Soft to hard, low to high melting point, good conductor of heat and electricity	All metallic elements (Na, Mg, Fe, Cu)

A crystalline solid possesses rigid and long range order. In a crystalline solid, atoms, molecules or ions occupy specific (predictable) positions. An amorphous solid does not possess a well defined arrangement and long range molecular order. A glass is an optically transparent fusion product of inorganic materials that has cooled to a rigid state without crystallizing (Table 1). A unit cell is the basic repeating structural unit of a crystalline solid. We have seen that in crystals weak intermolecular interactions are strongly directional and that the mutual orientation of neighboring groups is important in achieving stable packing arrangements. On melting to a liquid, there is usually only a slight change in packing density but the breakdown of periodicity means that the regular, favorable orientation neighboring molecules is partly lost.

Molecular crystals are bound by intermolecular (van der Waals) forces, and knowledge of such force fields should be sufficient to predict crystal structures. In principle, accurate force fields can be obtained using electronic structure methods, but for reasons discussed below. This has not yet been achievable in practice. Thus, theoretical investigations of crystal structures typically rely on empirical force fields that are parametrized using

experimental information. Unfortunately, the predictive capability of such fields is limited, since a given field can describe well only the system used for its parametrization and thus is often not transferable even to polymorphs of this system. As a result, prediction of crystal structures has been considered an impossible task. This opinion was echoed first by Ball [21] and more recently by Desiraju [22], who wrote that the issue eluded scientists for more than 50 years and emphasized the low success rate of crystal structure predictions in the blind tests conducted by the Cambridge Crystallographic Data Center [23]. One of the key issues in predicting crystal structures is the accuracy of the force fields. This accuracy is also critical for calculations of lattice energies at experimental crystal structures [24]. The force fields can be computed ab initio using wave function based methods, but until recently the accuracy achievable for molecules containing more than a few atoms was far from quantitative and was insufficient for determination of crystal structures. One might have hoped that the problem could be resolved by the development of density functional theory, which can be applied to systems containing hundreds of atoms [25]. Unfortunately, conventional density functional theory methods fail badly in describing intermolecular interactions for which dispersion is the dominant component; such systems include molecular organic crystals.

V. CONCLUSION

The hydrogen is attached directly to one of the most electronegative elements, causing the hydrogen to acquire a significant amount of positive charge. Each of the elements to which the hydrogen is attached is not only significantly negative, but also has at least one active lone pair. An alcohol is an organic molecule containing an O-H group. Any molecule which has a hydrogen atom attached directly to an oxygen or a nitrogen is capable of hydrogen bonding. Such molecules will always have higher boiling points than similarly sized molecules which don't have an O-H or an N-H group. The hydrogen bonding makes the molecules stickier, and more heat is necessary to separate them.

Hydrogen bonding also occurs in organic molecules containing N-H groups in the same sort of way that it occurs in ammonia. Examples range from simple molecules like CH_3NH_2 (methylamine) to large molecules like proteins and DNA. The two strands of the famous double helix in DNA are held together by hydrogen bonds between hydrogen atoms attached to nitrogen on one strand, and lone pairs on another nitrogen or an oxygen on the other one.

Intermolecular forces are forces between molecules that determine the physical properties of liquids and crystals. Molecular substances are often soluble in organic solvents which are themselves molecular. Both the solute and the solvent are likely to have molecules attracted to each other by van der Waals' forces.

For liquids, we have a wealth of thermodynamic data from which averaged interaction energies can be estimated but we lack structural information about the relative orientations of neighboring molecules. The equilibrium vapor pressure is the vapor pressure measured when a dynamic equilibrium exists between condensation and evaporation. The boiling point is the temperature at which the (equilibrium) vapor pressure of a liquid is equal to the external pressure.

For crystals, we need better atom-atom potential energy functions. We have a wealth of information about the directional properties of weak interactions but very little thermodynamic data with which to test packing energy estimates. The phenomenon of polymorphism shows that the crystal form stable at room temperature is not necessarily that with the best packing energy, the entropy cannot be neglected.

ACKNOWLEDGEMENTS

I am grateful to Professor Dr. K. Jyrki KAUPPINEN (Physics Department, Laboratory of Optics and Spectroscopy, University of Turku, Turku-Filland) for the oppurtunity to perform this work and his valuable comments on the manuscript, and to Professor Dr. Erzsebet HORVATH (Analytical Chemistry Department, Faculty of Engineering, University of Pannonia, Veszprem-Hungary) for her useful advice and help stimulating discussions.

REFERENCES

[1] K.F. Purcell and R.S. Drago, Studies of the bonding in acetonitrile adducts, *Journal of the American Chemical Society, 88(5),* 1966, 919-924.

[2] D.F. Shriver, Preparation and structures of metal cyanide-Lewis acid bridge compounds, *Journal of the American Chemical Society, 85(10),* 1963, 1405-1408.

[3] K.F. Purcell, σ and π binding effects in the coordination of carbon monoxide and comparison with cyanide ion, *Journal of the American Chemical Society, 91(13),* 1969, 3487-3497.

[4] F. Huisken, M. Kaloudis and A.A. Vigasin, Vibrational frequency shifts caused by weak intermolecular interactions, *Chemical Physics Letters, 269(3-4),* 1997, 235-243.

[5] M.C. Etter, Encoding and decoding hydrogen-bond patterns of organic compounds, *Accounts of Chemical Research, 23(4),* 1990, 120-126.

[6] G.R. Desiraju, The C-H…O hydrogen bond in crystals: What is it?, *Accounts of Chemical Research, 24(10),* 1991, 290-296.

[7] C.P. Brock and J.D. Dunitz, Towards a grammar of crystal packing, *Chemistry of Materials, 6(8),* 1994, 1118-1127.

[8] P. Kollman, A general analysis of noncovalent intermolecular interactions, *Journal of the American Chemical Society, 99(15),* 1977, 4875-4894.

[9] A.E. Lutskii, V.V. Prezhdo, L.I. Degtereva and V.G. Gordienko, Spectroscopy of intermolecular field interactions in solutions, *Russian Chemical Reviews, 51(8),* 1982, 802-817.

[10] J.S. Chickos, D.G. Hesse, J.F. Liebman and S.Y. Panshin, Estimations of the heats of vaporization of simple hydrocarbon derivatives at 298 K, *The Journal of Organic Chemistry, 53(15),* 1988, 3424-3429.

[11] W.N. Setzer and P.V.R. Schleyer, X-ray structural analyses of organolithium compounds, *Advances in Organometallic Chemistry, 24,* 1985, 353-451.

[12] J.K. Whitesell, R.E. Davis, L.L. Saunders, R.J. Wilson and J.P. Feagins, Influence of molecular dipole interactions on solid-state organization, *Journal of the American Chemical Society, 113(9),* 1991, 3267-3270.

[13] R. Taylor and O. Kennard, Crystallographic evidence for the existence of C-H···O, C-H···N, and C-H···Cl hydrogen bonds, *Journal of the American Chemical Society, 104(19),* 1982, 5063-5070.

[14] J.D. Dunitz, Weak intermolecular interactions in solids and liquids, *Molecular Crystals and Liquid Crystals, 279(3-4),* 1996, 209-218.

[15] A. Gavezzotti, Are crystal structures predictable?, *Accounts of Chemical Research, 27(10),* 1994, 309-314.

[16] R. Thomas, C.B. Shoemaker and K. Eriks, The molecular and crystal structure of dimethylsulfoxide, $(H_3C)_2SO$, *Acta Crystallographica, 21(12),* 1966, 12-20.

[17] R.K. McMullan and A. Kvick, Structures of cubic and orthorhombic phases of acetylene by single-crystal neutron diffraction, *Acta Crystallographica, B48(5),* 1992, 726-731.

[18] S. Aono, Theory of intermolecular interactions, *Bulletin of the Chemical Society of Japan, 75(1),* 2002, 65-70.

[19] S.C. Nyburg and C.H. Faerman, A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br and I bonded to carbon, *Acta Crystallographica, B41(4),* 1985, 274-279.

[20] R.E. Rosenfield, R. Taylor, and J.D. Dunitz, Directional preferences of nonbonded atomic contacts with divalent sulfur. 1. Electrophiles and nucleophiles, *Journal of the American Chemical Society, 99(14),* 1977, 4860-4862.

[21] P.Ball, Scandal of crystal design…, *Nature, 381(6584),* 1996, 648-650.

[22] G.R. Desiraju, Cryptic crystallography, *Nature Materials, 1(2),* 2002, 77-79.

[23] W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, J.P.M. Lommerse, W.T.M. Mooij, S.L. Price, H. Scheraga, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer and D.E. Williams, Crystal structure prediction of small organic molecules: a second blind test, *Acta Crystallographica, B58(5),* 2002, 647-661.

[24] W.B. Schweizer and J.D. Dunitz, Quantum mechanical calculations for benzene dimmer energies: present problems and future challenges, *Journal of Chemical Theory and Computation, 2(2),* 2006, 288-291.

[25] R. Podeszwa, B.M. Rice and K. Szalewicz, Predicting structure of molecular crystals from first principles, *Physical Review Letters, 101(11),* 2008, 115503.1-4.

Image Compression Using Binary Space Partitioning and Geometric Wavelets

Pranob K Charles[1], Dr.Habibulla Khan[2], Vinnakota Harish[3], Mule Swathi[3]
Chedurupalli Deepthi[3], Cherukuri Rajesh Kumar[3], Nikhita Nayudu[3]
*(Associate Professor, Dept. of Electronics and Communication Engineering, KLUniversity.)
** (Professor,HOD, Dept. of Electronics and Communication Engineering, KLUniversity.)
*** (B.Tech Scholars, Dept. of Electronics and Communication Engineering, KLUniversity.)

ABSTRACT

For low bit-rate compression applications, segmentation-based coding methods provide, in general, high compression ratios when compared with traditional coding approaches. During the last years, after JPEG2000, different techniques were developed in the area of Image Compression. Although they outperformed the JPEG2000 algorithm, the partitioning problem persists. In this paper, we present a segmentation based image compression technique which is based on Binary Space Partitioning (BSP) and Geometric wavelets. By using Binary Space partitioning technique the image is segmented recursively into a number of segments until an exit criterion is met and a tree is formed with all these segments. Geometric Wavelets are used to remove the insignificant nodes present in the tree. Finally, this method is compared with various wavelet based and transform based image compression techniques and it is show that this method outperforms all of them.

Keywords **– Binary Space Partitioning, image compression, geometric wavelets, JPEG2000.**

I. INTRODUCTION

The field of image compression is rich in diverse source coding schemes ranging from classical lossless techniques and popular transform approaches to the more recent segmentation-based coding methods. The notion of segmentation-based coding was introduced during the early 1980's.Segmentation-based compression methods usually describe the desired image as a set of regions.

In general, the description of each region requires two types of information: 1) the geometry of the region boundaries and 2) the attributes of the image signal within the region. In order to achieve high compression ratio and good image quality, one needs to segment the image into a minimum number of regions such that the geometric description of the regions' boundaries is simple and the image signal within each region is continuous (or smooth).
Therefore, the most challenging aspect of a segmentation-based coding approach is to balance between a small number of geometrically simple regions and the smoothness (or continuity) of the image signal within these regions.

The main work described in this review is based on the document [1] "an improved image compression algorithm using binary space partition scheme and geometric wavelets" written by G.Chopra, A.K.Pal and published in IEEE transactions on image processing in 2011, at some point it is discussed aspects of [3] "image compression using binary space partitioning trees" written by Hayder Radha, Martin Vetterli and Riccardo Leonardi, and published in IEEE transactions on image processing in 1996. and there is a support document to complete the discussion [2] "image coding with geometric wavelets" written by Dror Alani, Amir Averbuch and Shai Dekel and published in IEEE transactions on image processing in 2007.

The technique used in [1] is similar to [2] but they differ only in the case of selecting the type of partition line. The normal form of the straight line is used to represent the partition line incase of [2] whereas, the slope intercept form of the line is used in [1].

This method is applied to 8 bits gray scale images but it could be extended to color images in the same way that JPG2000 has been applied to different type of images (i.e. 8bits/pixel, 24bits/pixels).

In the following sections the algorithm, pseudo code, Binary Space Partition method, tree encoding, results and conclusions are described.

II. BINARY SPACE PARTITIONING (BSP)

Segmentation techniques partition the digital image into a set of different geometric regions which are approximated by simple functions. Segmentation based image coding methods were introduced during the early 1980[6], [7].Since then, many segmentation techniques have been developed and among them the BSP scheme is a simple and effective method.

The most challenging aspect of a segmentation based coding approach is to balance between a small number of geometrically simple regions and the smoothness of the image signal within these regions.

The BSP can be summarized as follows. Given an image *f*, the algorithm divides Ω into two subsets Ω_0 and Ω_1 using a bisecting line and minimizing a given functional. The algorithm continues partitioning each region recursively until it reaches a given measure or there is no enough pixels to subdivide. The algorithm constructs a binary tree with the partitioning information.

To approximate the image *f* in a given region Ω_i a bivariate linear polynomial is used which is defined by:

$$Q_{\Omega_i} = A_i x + B_i y + C_i \qquad (1)$$

The functional used to find the best subdivision for a given region is the following:

$$\mathrm{F}(\Omega_0, \Omega_1) = \arg\min_{\Omega_0,\Omega_1} \|f - Q_{\Omega_0}\|^2_{\Omega_0} + \|f - Q_{\Omega_1}\|^2_{\Omega_1}. \quad (2)$$

Where Ω_0 and Ω_1 represent the subsets resulting from the subdivision of Ω where Ω_0 and Ω_1 should be considered as children for the father Ω . Fig.1. shows the steps involved in Binary Space Partitioning algorithm. First a line L divides the region Ω into two regions Ω_0 and Ω_1 . The two regions Ω_0 and Ω_1 are further divided into Ω_{00} , Ω_{01} and Ω_{11} , Ω_{10} respectively. These four regions are further divided into eight segments and this is done recursively. Then it is represented in a tree structure as shown in fig.2.

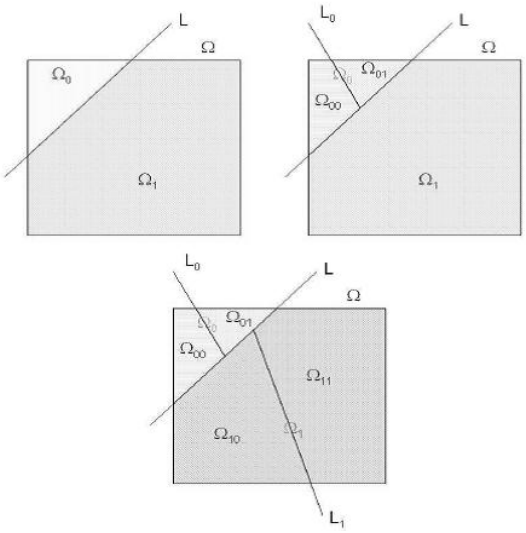


Fig.1. Two partition levels using bisecting lines

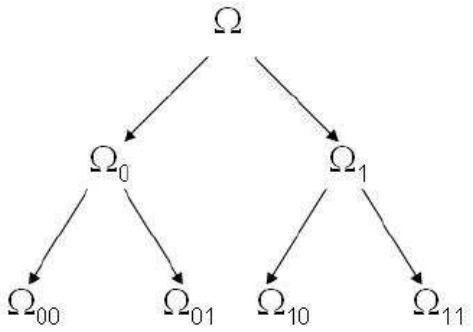


Fig.2. BSP tree representation

The value of the coefficients A,B and C are found by minimizing the function given below:

$$\prod = \sum_{i=1}^{n} [f(x_i, y_i) - (A\, x_i + B\, y_i + C)]^2 \qquad (3)$$

By taking the partial derivatives for A, B ,C and solving the below three equations , we will get the coefficients of the polynomial.

$$\sum_{i=1}^{n} z_i = A\, n + B \sum_{i=1}^{n} x_i + C \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} x_i\, z_i = A \sum_{i=1}^{n} x_i + B \sum_{i=1}^{n} x_i^2 + C \sum_{i=1}^{n} x_i y_i$$

$$\sum_{i=1}^{n} y_i\, z_i = A \sum_{i=1}^{n} y_i + B \sum_{i=1}^{n} x_i y_i + C \sum_{i=1}^{n} y_i^2$$

III. SPARSE GEOMETRIC WAVELETS (GW) REPRESENTATION

In [1] they use the local difference to define the geometric wavelets. The local difference computes the difference between the actual partition and the previous giving us an idea of the degree of change, if the difference is large then the new partition is capturing new details, and if the difference is small, the new partition does not add new information. The GW is defined as follows:

$$\Psi\Omega_0(f) \triangleq 1\Omega_0(Q_{\Omega_0} - Q_\Omega) \qquad (4)$$

Where $1_{\Omega 0}$ is the function that gives us 1 in Ω_0 and 0 in the rest. Ω_0 here means one of the children.

We show that it is possible to reconstruct the function *f* using GW due to the term cancelations.

$$f = \sum_{\Omega_i} \Psi_{\Omega_i}(f) \qquad (5)$$

But using the BSP tree we can compute the norm of each $\psi_{\Omega i}$ (*f*), which is a measure of the degree of change, then sorting these numbers it is possible to approximate the function by the n-term geometric wavelet sum defined as

$$f \approx \sum_{j=0}^{n} \Psi_{\Omega k_j}(f) \qquad (6)$$

The BSP tree that is generated may contain a large number of GW nodes. Yet, for low bit-rate coding, only few of them (typically 1-4 %) are needed to obtain a reasonable approximation of the image. Therefore, we apply the 'greedy' approximation methodology; sort the geometric wavelets according to their energy 'contribution' and extract a global n-term approximation from the joint list of all the geometric wavelets over all the image tiles.

We found that for the purpose of efficient encoding it is useful to impose the additional condition of a tree structure over each image tile. Namely, we require that if a child appears in the sparse representation, then so does the parent. Once a parent is encoded in this hierarchical representation, then we only need to encode the (quantized) BSP line that creates the child. This significantly saves bits when the geometry of the sparse representation is encoded. On the other hand, the penalty for imposing the connected tree structure is not significant, since with high probability, if a child is significant, then so are his ancestors. Fig. 3 illustrates an n-term GW collection whose graph representation includes some unconnected components and Fig. 4 illustrates the final GW tree after the missing ancestors were added.

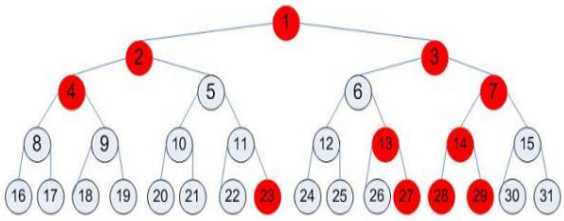


Fig.3. Example of a greedy selection.

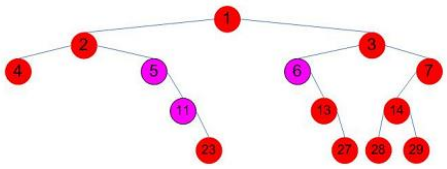


Fig.4. The final GW tree with the additional nodes.

Finally, we apply a rate-distortion (R - D) optimization process. Empirical results show that this rate-distortion mechanism increases the PSNR by 0.1 dB in some cases.

IV. ENCODING THE SPARSE GW REPRESENTATION

Once the BSP tree is generated the next step is to discard the irrelevant information by computing $\psi_{\Omega ki}$ (f) .

Before the actual BSP Tree is encoded, a small header is written to the compressed file. This header contains the minimum and maximum values of the coefficients of the wavelets Q_Ω participating in the sparse representation. These values are used by the decoder to decode the coefficients. In addition, the header contains the minimum and maximum values of the gray levels in the image. The coefficients extremal values are encoded with four bytes each and image extremal values with 1 byte each. Therefore, the header size is $3 \times 2 \times 4+2 \times 1= 26$ bytes.

Due to the fact that the leaves are necessary for the reconstruction, in [3] they impose the requirement that if a child appears in the tree then father has to appear too. But it is not necessary that both children appear in the tree, if one is not significant enough, could be excluded from the sparse representation, then all its descendant should be excluded too. This is an improvement with respect to [2] because if a partition is done both the children appear in the tree independent of whether one child is significant or

not.After the tree is pruned it is encoded using the following information:

- Tree structure information.
    - Number of children.
    - Information to distinguish each child node.
- The quantized coefficients Q_{Ω}.
- The bisecting line information of each Ω if it has a child.
- Header information.

The tree-structure is encoded using the fact that with a high probability a significant node does not have a significant child, in a similar way like 'zero-trees'. Therefore using Huffman code to encode the three different values (Zero children='1', One Child='01', Two-Children='00') it is possible to save in some cases 1 bit, due to normally it is necessary 2 bit to encode 3 different states.

The quantized coefficient Q - that represent the wavelet polynomials are determined by three real numbers (A_i ; B_i ; C_i) that can be stored with 12 bytes using the standard 4-bytes float representation. But in [3] they show an algorithm to store at a rate of 1.5 bytes per polynomial on average, using the standard Graham-Schmidt method to obtain the orthonormal base representation. This could be the greatest improvement with respect to [2].

In order to deal with the time consuming algorithm , we tile the image in squares of 128x128 and they apply the BSP algorithm on each tile. The main disadvantage of doing this is that blocking artifacts appears at the tiles' boundaries. Another disadvantage is that connected areas could be disconnected missing the possibility to improve the approximation.Fig.5 shows the tiling of the cameraman image.

Fig.5. Tiling of cameraman image of size 256 x 256

V. DECODING

In this step compressed bit stream is read to find whether the participating node is the leaf node, has 1 child or 2 children. If one child is participating then by using bit stream, it is found that whether it is left or right. If at least one of the children belongs to the sparse representation, then the coefficients of the bisecting line are calculated. Thereafter, using this optimal cut, domain is partitioned into two sub-domains; and depending upon the situation vertex set of only one child or both children is found. An orthogonal basis was used during the encoding of the coefficients of the coefficients of geometric wavelet. Thus, before using the decoded geometric wavelets in n-term sum, its representation in the standard basis is found. This process is repeated until entire bit stream is read.

VI. ALGORITHM

A. Binary Space Partitioning.

1. Read the Image.
2. Tile the image using tiles of size 128 X 128.
3. Select a title.
4. Select a partition line using least square error based criteria.
5. Divide the title into two segments.
6. Repeat steps 4 and 5 until a minimum threshold value is reached for a particular segment.
7. Repeat steps 4,5 and 6 for the next tile.
8. Taking tile as a parent and the corresponding segments as child we form a tree with the leaf nodes as the final segments.

B. Sparse Geometric Wavelet Representation

9. We apply greedy approximation method for the tree obtained in 'A'.
10. If any child has a missing parent we include it.

C. Encoding:

11. A header containing the minimum and maximum values of the coefficients of the wavelets as well as the

minimum and maximum of the grey levels of the image is added to the compressed files.
12. Encoding the tree structure.

13. Encoding the bisecting line.

14. Encoding the coefficients of the wavelet polynomials.

15. Quantizing the coefficients in an orthogonal polynomial basis representation.

16. Bit allocation of polynomial coefficients.

D. Decoding:

17. Compressed bit stream is read.
19. Whether the selected node is leaf node, has 1 child or 2 children is found.

20. If one child is present it is found whether it is left or right.

21. The coefficients of the polynomial and the bisecting line are decoded.

22. This process is repeated until the entire bit stream is read.

VII.PSEUDO CODE

```
I = image_read('image_name');
        //reads the image to I

I_new = image_tile(I);

        //tiles the image to I_new

For i=1:1:4

    I = select_tile(I_new);

    While (threshold is not reached)

        Select_partition_line();

        Divide_image();

    End

End

Tree = Construct_Binary_Tree( );

        //creates the BSP tree

GD_tree = Greedy_approximate( Tree );

        //applies greedy approximation

I_encode = Encode_tree( GD_tree );

        //encodes the tree structure

I_decode = Decode( I_encode );

        //reconstructs the image
```

VIII. RESULTS

The PSNR (peak signal to noise ratio) based on MSE (mean square error) is used as a measure of "quality." MSE and PSNR are given by the following relations:

$$MSE = \frac{1}{m * n} \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{i,j} - y_{i,j})^2$$

$$PSNR = 10 \log[(255)^2 / MSE]$$

Where n x m is the image size, $x_{i,j}$ is the initial image $y_{i,j}$ is the reconstructed image. MSE and PSNR are inversely proportional to each other and higher value of the PSNR produces better image compression.

Fig 6. Original Image[1]

Fig. 7. Reconstructed image using the proposed method[1]

Table 1. Comparing the PSNR values of various methods[1]

Method	128:1	64:1	32:1
SPIHT	22.8	25	**28**
Kakadu	21.15	24.11	27.29
GW	22.93	25.07	27.48
Proposed method	**23.04**	**25.29**	27.62

IX. CONCLUSIONS

The key idea behind this work is the fact that it is possible to subdivide a region using a bisecting line and to encode this line only with a few bits depending on the quantization schema. This new approach to encode images seems to be better to encode images at a very low bit-rate.

The algorithm was exhaustively analyzed to reduce to the minimum the number of bits encoded, we can see the effort of the authors in 2.3.1 in [2], in order to reduce at most one bit in some cases, to encode the tree-structure.

Although it seems to be a good technique there exists a few things that are not clearly specified neither in [1] nor in [2].The authors in [1] state that the algorithm is computationally intensive. They do not show how much intensive it is. Due to the fact that in a brute force algorithm like this, it is not easy to show the order, but they should compare the performance of this algorithm in such a way some-one can estimate the encoding and decoding time, like running the algorithm in different computers or comparing times with the standard JPG2000.

One important idea, ones can infer from this is that a complex partition can give a better approximation but it means that more information is needed to store the partition, therefore a better storing algorithm is needed.

REFERENCES

[1] Garima Chopra and A.K.Pal , "An Improved Image Compression Algorithm Using Binary Space Partition Scheme and Geometric Wavelets",*IEEE Transactions on Image Processing.VOL 20,NO.1,January 2011.*

[2] D. Alani, A. Averbuch and S. Dekel, "Image coding with geometric wavelets", *IEEE Transactions on Image Processing, 16(1), 69-77, 2007.*

[3] H. Radha, M. Vetterli and R. Leonardi "Image Com-pression Using Binary Space Partitioning Trees", *IEEE Transactions on Image Processing, vol. 5,num. 12, pp. 1610-1624, 1996.*

[4] R. Shukla, L. Dragotti, M.N. D and M. Vetterli, "Rate-Distortion Optimized Tree-Structured *Compression Al-gorithms for Piecewise Polynomial Images", IEEE Transactions on Image Processing, vol. 14,num. 3, pp. 343-359, 2005.*

[5] M.Shapiro, "An embedded hierarchical image coder using zerotrees of wavelet coefficients", *IEEE Transac-tions on Signal Processing, vol. 41, pp.3445-3462,1993.*

[6] M. Kocher and M.Kunt , "A contour-texture approach to image coding,", *in Proc. ICASSP, 1982, pp.436-440*

[7] M.Kunt, A.Lkonomopoulos, and M.Koche," Second generation image coding techniques", Proc. IEEE ,vol.73, no.4,pp,549-574, Apr.1985.

[8] M. A. Losada, G. Tohumoglu, D. Fraile, and A. Artes, "Multi-iteration wavelet zerotree coding for image compression," *Sci. Signal Process., vol. 80, pp. 1281–1287, 2000.*

[9] G.K.Wallace,"The JPEG still-picture compression standard," *Commun. ACM, vol. 34, pp. 30–44, Apr. 1991.*

[10] MPEG-2video,ITU-T-Recommendation H.262-ISO/IEC 13818-2, Jan. 1995.

[11] K. R. Rao and P. Yip, Discrete Cosine Transform:*Algorithms,Advantages,Applications. New York: Academic, 1990.*

[12] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3445–3462, Dec. 1993.*

[13] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set portioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 3, pp. 243–250, Jun. 1996.*

[14] A. Islam and W. A. Pearlman, "An embedded and efficient low complexity ierarchical image coder," *in Proc. SPIE, Jan. 1999, vol. 3653, pp. 294–305.*

[15] D. Tauban, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process., vol. 9, no. 7, pp. 1158–1170, Jul. 2000.*

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-461-469 ISSN: 2249-6645

Implementation of query optimization in OODBMS-Review paper

[1] Prof.S.N. Sawalkar, [2]Prof S.S.Dhande, [3]Dr.G.R.Bamnote
[1] Lecturer, Computer science & IT Dept,Sipna's C.O.E.T,Amravati(M.S.)
[2] Asstt. Professor, Computer science & IT DeptSipna's C.O.E.T,Amravati(M.S.)
[3]Head & Professor,Computer Science&Engg. DeptPRMITR,Badnera Amravati(M.S.)

Abstract

Data management and organization have become so complex and challenging in today's electronic age of information. Databases, be it bibliographic or textual, ought to have the capability of storing graphics, video, audio and other highly structure data. The database technologies have constantly evolved to meet these changing requirements by adopting object oriented programming concepts.

One of the relational databases successes are that the optimizations of its queries have been studied well and showed its proofs in the speed of query execution. However, the relational model only permits the alphanumeric data management. Nowadays, the necessity to support complex data in databases is intensified. Data are represented in the basis as of objects, associations and object identification that permit a fast navigational access between the different objects. Models trying to answer to these needs appeared as the object-oriented and the object relational models. One which provides database capabilities and other extend the relational model with object-oriented features like powerful data abstractions and modeling framework for query processing and optimization. Some of features of object query optimization include object identity and nesting /unnesting of query expressions

Keywords- OODBMS, Object Query Language OQL, Object identity (OID), ODMG Standards etc.[1],[2].

I. NEED OF ODBMS OVER RDBMS

The relational model is the basis of many commercial relational DBMS products (e.g., DB2, Informix, Oracle, Sybase) and the structured query language (SQL) is a widely accepted Standard for both retrieving and updating data.

The basic relational model is simple and mainly views data as tables of rows and columns.

The types of data that can be stored in a table are basic types such as integer, string, and decimal.

Relational DBMSs have been extremely successful in the market. However, the traditional RDBMSs are not suitable for applications with complex data structures or new data types for large, unstructured objects, such as CAD/CAM, Geographic information systems, multimedia databases, imaging and graphics. The RDBMSs typically do not allow users to extend the type system by adding new data types. They also only support first-normal-form relations in which the type of every column must be atomic, i.e., no sets, lists, or tables are allowed inside a column.

Due to the new needs in database systems, a number of researches for OODBMS have begun in the early 80.s. Object-oriented database systems began developing in the mid-80 out of a necessity to meet the requirements of applications beyond the data processing applications which were [are] served by relational database systems.

There would be performance degradation due to RDBMS technology used. The limitations of Relational database for Geographical Information System, CAD, Multimedia, Engineering etc. [2].

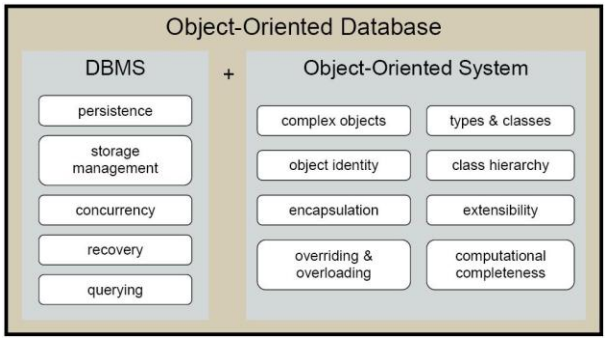
II.INTRODUCTION

A relational database system has a clear specification given by Codd, no such Specification existed for object-oriented database systems even when there were already products and in the market. A consideration of the features of both object-oriented systems database management systems has lead to adefinition of an object-oriented database; this distinguishes between the mandatory, optional and open features of an object-oriented database.

A. Mandatory Features and OODBS

Database is a product that provides a means of storing data persistently and retrieving that data again at a later point in time must address issues that are far more involved than simple storage and retrieval of data.

The mandatory features, which must be present if the system is to be considered



PERSISTENCE
Storage and (random) retrieval of data

CONCURRENCY
The ability to support multiple users simultaneously (lock granularity is often an issue here)

RECOVERY
Ability of database to recover integral data safely under hazardous situations(feature of Auto commit)

STORAGE MANAGEMENT
Involves storage issues including data management, clever caching policies, data, availability, data scalability, integrity, data clustering, system log file s maintaining objects communications.

QUERING
Predictable performance as the number of users or the size of the database increases depending on scalability, query processing, query optimization, query execution.[3]

Query processing = Query optimization + Query execution

B.COMPARATIVES OF RDBMS AND ODBMS DATA PROCESSING

(RDBMS), the performance depending on notion of data clustering is straightforward. Sequential rows of a particular table are stored together on disk, page-by-page. This makes particular sense with the relational architecture as most database activity is based on some form of sequential data access of a table. The physical clustering of data in this environment is extremely important to database performance since the architecture of RDBMS systems is heavily server-centric. ("Server-centric" implies that the majority of actual database operations occur on the server proper, even in a distributed client/server environment.)

This physical collocation of data on disk allows for optimal access of data within a particular table because only data from that table is stored on that particular page. The benefits of clustering in an RDBMS are mostly realized by the server because it performs all of the direct access to disk and virtually all database operations. This clustering plays no real part in client operation as only data requested by the client is actually passed there from the server. In an RDBMS, a query that is not able to benefit from clustering will be slower, but client operation will be unaffected because even in this case only requested data is passed from server to client. Actual data pages remain on the server.

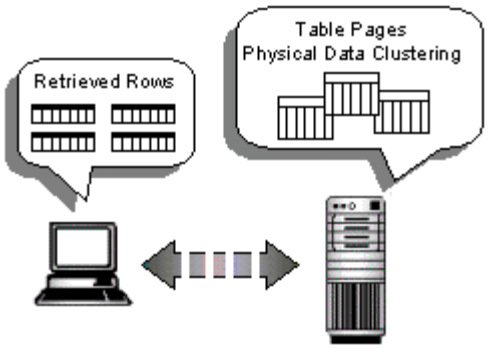


Fig. 1 – Server Centric Technology (RDBMS)

In contrast to relational systems, most—but not all—Object-Oriented Database Management Systems (ODBMS) are actually very client-centric in nature. As with an RDBMS, all disk activity occurs on the server. Unlike an RDBMS, however, most database activity occurs within the client application itself. Pages are retrieved from disk and passed directly from server to client for processing.,

(ODBMS): Physical = Logical Clustering

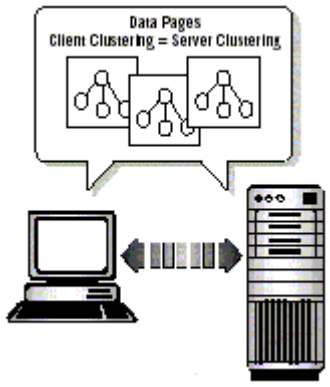


Fig 2 – Client Centric Architecture

In these systems, even such basic database operations as queries and index maintenance are actually client-side operations. The server simply reads and forwards pages to the client, and then waits for those pages to be modified and returned for update to disk.[3].

III.LITURATURE RIVIEW

A. Query optimization: Focus of performance improvement

Query optimization is the process of selecting the most efficient query-evaluation plan from many strategies usually possible for processing a given query if the query is complex. One aspect of optimization occurs at the relational-algebra level, where the system attempts to-find an expression that is equivalent to given application, but more efficient to execute.

Another aspect is selecting a detailed strategy for processing the query, such as choosing the algorithm to use for executing the operation, choosing the specific indices to use, and so on.

In either case the problem boils down to parsing, estimating complexity of the algorithms which minimize cost, or time as the case may be. Object-oriented databases integrate object orientation with database capabilities. Object orientation allows a more direct representation and modeling of real world problems. Today Oracle, Microsoft, Borland, Informix, and others incorporated object-oriented features into their relational systems. Most current OODBs are still not full-fledged database systems comparable to current relational database systems (RDBs) [8].

The Query optimization consist of three major components in OODBS

A) SQL Transformation

B) Execution Plan Selection

C) Cost Model and Statistics

A)SQL Transformation

The purpose of SQL Transformation is to transform the original SQL statement into semantically equivalent OQL statement that can be processed more efficiently.

B)Execution Plan Selection

In Execution Plan Selection, the optimizer selects an execution plan. That describe all the steps when the OQL is processed, such as order in which objects are accessed, when the table are join together.[1],[2]

C) Cost Model and Statistics

The Cost Estimates are based Upon I/O, CPU and Memory Resources Required by each query operation, and The Statistical Information about the database object such as table, indexes and Views and also selecting best plan among many all possible strategies

Query evaluation may takes place by following process

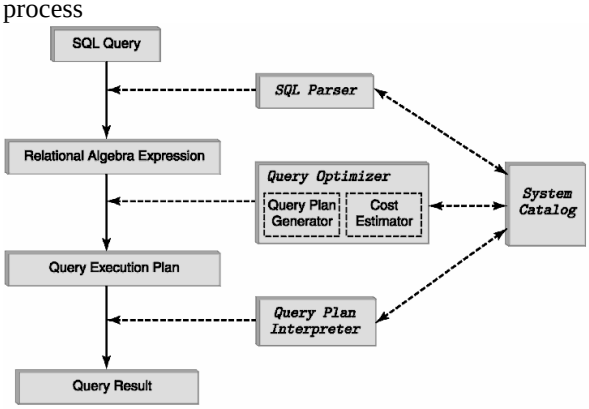


Fig. 3–Query evaluation

System Catalogs provides information about objects relations and their index relations used for communications.

IV.OBJECT QUERY LANGUAGE (OQL)

The Object Query Language (OQL) is the extended SQL syntax associated with ODL of ODMG.

Principles and assumptions of the design:

a) OQL is not computationally complete, but queries can invoke methods and methods can include queries.

b) OQL provides declarative access to objects.

c) OQL assumes the object model defined by ODL.

d) Formal semantics and optimization procedure can be defined for OQL (work is still pending in this area).

e) The syntax of OQL can be extended to merge the language with programming languages such as C++, Java or Smalltalk.

f) OQL provides primitives to deal with sets, structures, and lists.

g) OQL does not define its own update operators but uses the update operators defined on objects.
OQL allows methods to be used in queries in the same way as attributes[1],[2],[3].

V. RELATED WORK

Several methods of query optimization have been proposed, Gemstone [4], O2 [5], Orion [6] and Blackboard [7].

A. The GemStone Object-Oriented Database, from GemStone Systems, Inc.

First introduced in 1987, GemStone is the oldest commercial ODBMS available today. GemStone is particularly well suited for use in complex multi-user, multi-platform client/server applications. It supports concurrent access from multiple external languages, including Smalltalk (VisualWorks, Visual Age, and Visual Smalltalk), C++ and C. GemStone also provides Smalltalk as an internal DML, which can execute methods or the Entire application in the database.

B. A query language for O2.in Building an Object-Oriented Database System

A complete description of the O_2 system, an object-oriented database system, is presented. It has the functionality of a DBMS (persistence, disk management, sharing and query language) and of an object-oriented system (complex objects, object identity, encapsulation, typing, inheritance, overriding, extensibility, and completeness). It also includes a set of user interface generation tools and a complete programming environment. O_2 supports a Multilanguage paradigm and a dual mode of operation (development and execution) and it operates on a workstation/server configuration. The system is described as seen from the programmer's point of view and as seen through the programming environment

C. Query processing in distributed ORION

In this query processing strategies developed and implemented are in the distributed version of the ORION object-oriented database system. The ORION query model is based on the ORION object-oriented data model. Further, we have adopted the response time as the primary objective function for query optimization. The query-processing strategies we have developed reflect our solutions to these requirements. In particular, our strategies are based on a dataflow execution model which represents a plan for executing a query concurrently at multiple sites. One important observation we bring out in our description of the ORION query-processing strategies is that most of the important techniques developed for optimizing and processing a relational query apply directly to an object-oriented query, despite the differences in the underlying data models

D. A blackboard architecture for query optimization in object bases

Adopting the blackboard architecture from the area of Artificial Intelligence, a novel kind of optimizer enabling two desirable ideas will be proposed. Firstly, using such a well-structured approach back propagation of the op-timized queries allows an evolutionary improvement of (Crucial) parts of the optimizer. Secondly, the A search strategy can be applied to harmonize two contrary properties: Alternatives are generated whenever necessary, and straight-forward optimizing is performed whenever possible, however. The generic framework for realizing a blackboard op-timizer is proposed. Then, in order to demonstrate the viability of the new approach, a simple example op-timizer is presented. It can be viewed as an incarnation of the generic framework.

E. Open Source Database DB4o with performance evaluated query optimization

Db4o is the open source object database that enables Java and .NET developers to store and retrieve any application object with only one line of code, eliminating the need to predefine or maintain a separate, rigid data model. db4o enables compelling new features and achieving unprecedented performance and flexibility. db4o excels in a wide range of applications due to its performance, transparency, flexibility and ease of use.

VI. ANALYSIS OF PROBLEM

Relational DBMSs have been extremely successful in the market. However, the traditional RDBMSs are not suitable for applications with complex data structures or new data types for large, unstructured objects, such as CAD/CAM, Geographic information systems, multimedia databases, imaging and graphics. The RDBMSs typically do not allow users to extend the type system by adding new data types. They also only support first-normal-form relations in which the type of every column must be atomic, i.e., no sets, lists, or tables are allowed inside a column.

Due to the new needs in database systems, a number of researches for OODBMS have begun in the early 80.s. Object-oriented database systems began developing in the mid-80's out of a necessity to meet the requirements of applications beyond the data processing applications which were [are] served by relational database systems.

There would be performance degradation due to RDBMS technology used. The limitations of Relational database for Geographical Information System, CAD, Multimedia, Engineering etc. have led to the development of Object-oriented Database Systems.[8]

A. Query Optimization: Implementation

The query optimization is the process of selection of the best path of access to data in a database .Process of optimization is summarizes in three steps (see Fig.). *Rewrite step* consists in a syntactic and semantic rewrite of the query in the goal to determine simpler equivalent queries The result of this step is the generation of a query graph. *Ordering operations step* is takes place in two phases: generation and assessment of the plans which determined in the first phase. *Execution step* permits to choose the optimal execution plan and to execute it. Two approaches present themselves, by materialization or by pipeline.

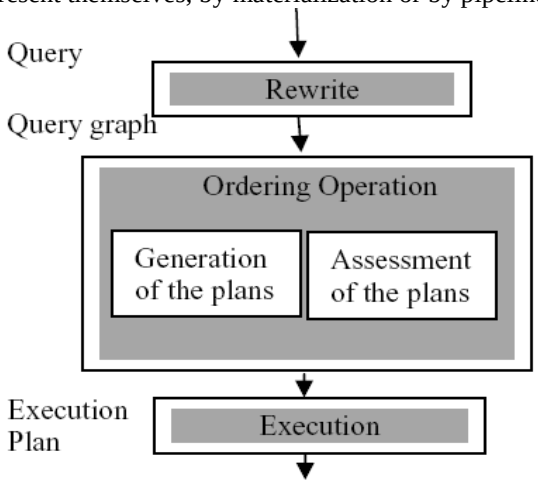


Fig.Summerization of Query Optimization

Features of Query optimization includes

- Search space must be able to include plans that have a low cost.
- The costing technique must be accurate.
- The enumeration algorithm that searches through the execution space must be efficient.

B. Problems of Object Model

Minimal query optimization: One of the biggest problems in OODBs is the optimization of queries. The additional complexity of the object-oriented data model (OODM) complicates the optimization of OODBs queries. This additional complexity is due to Additional data types

- Complex objects
- Methods and Encapsulation

ODBs query languages support the use of nested structures, which may again highly complicate the optimization process. Due to these problems optimization of object-oriented queries is extremely hard to solve and is still in the research stage. The optimization of joins is also another issue that needs more attention.

Lack of query facilities: The OODB query language is not ANSI sql compatible. The query language do not support Nested sub-queries, Set queries like Union, Intersection, Difference

Security concerns with OODBs: RDBs support authorization, OODBs do not support authorization. RDBs allow users to grant and revoke privileges to read or change the definitions, this feature has to be improved by OODB. No support for dynamic class definition changes with OODBs: Most OODBs do not allow dynamic changes to the database schema, such as Adding a new attribute or method to a class,

- Adding a new super class to a class,
- Dropping a super class from a class,

C. Plan explanation

The Plan is the sequence of operations performed by oracle to execute the statement. By examining the explain plan, we can identify inefficient SQL statements The plan gives the following information: An ordering of the tables referenced by the statement An access method for each table mentioned in the statement Data operations like filter, sort, or aggregation Optimization, such as the cost and cardinality of each operation In order to get the result of PLAN execute the following query[1],[2].

SELECT id, object name, operation, options FROM PLAN_TABLE order by id

The table shows the result of plan:--

ID	OBJECT_NAME	OPTIONS	OPERATION
0		SELECT STATEMENTS	
0		SELECT STATEMENT	
0		SELECT STATEMENT	
1		SORT	GROUP BY
1		SORT	GROUP BY
2	RBRANCH	BY INDEX ROWID	TABLE ACCESS
2	RBRANCH	BY INDEX ROWID	TABLE ACCESS

The fig consist of:
ID: is the number assigned to each step in the execution plan.
OBJECT NAME: is the name of table or index, OPTIONS: Options tell more about an operation. For example, the operation TABLE ACCESS can have the options: FULL or BY ROWID. Full means, the

entire table is accessed whereas BY ROWID means, Oracle knows from which block the rows are to be retrieved, which makes the time to access the table shorter.
OPERATION: Provides methods for retrieving and processing rows from a table.[9]

VII. PROPOSED WORK

Object models are descended of the semantic networks and object programming languages. They aim to permit the reuse of structures and operations to construct some more complex entities. In this we improves database manipulation performance by implementing model in ODBMS which use query optimization using *rewriting queries*, *ordering operations* includes plans generations and their assessment, and *execution of plans* to show optimum results. For query processing and optimization ODMG-OQL used to query complex types of data.[9]

A. Application

We are considering an example of retail banking system. The bank is organized into various branches and branch each branch located in a particular city and monitors the assets. Bank customers are identified by their cust-id values. Bank offers two type of accounts i.e. saving account & checking account with loan facility Thus the relation and attributes in the schema are:

Customer (cust_name, cust_street, cust_city)
Branch (branch_city, branch_name, assets)
Account (acct_no, branch-name, and balance)
Depositor (cust_name, acct_no)
Loan (loan_no, branch_name, amount)
Borrower (cust_name, loan_no)

B. Transformations

We make the key observation that since a group-by reduces the cardinality of a relation; an early evaluation of group-by could result in potential saving in the costs of the subsequent joins. We present an example that illustrates a transformation based on the above observation. An appropriate application of such a transformation could result in plans that are superior to the plans produced by conventional optimizers by an order of magnitude or more.

Example: Let us consider the query that computes branches located in a particular city and total count of branches in each city .The following alternative plan is possible. First, group-by clause applied after condition and hence search time is more and CPU cost is high. In other words we first check the condition and then grouped on branch city. Second, group-by clause applied before condition hence search time is less and CPU cost is less. Here we grouped on branch city first and then check the condition

Transformation that enables pushing the group-by past joins. Their approach is based on deriving two queries, one with and the other without a group-by clause, from the given SQL query. The result of the given query is obtained by joining the two queries so formed. Thus, in their approach, given a query, there is a unique alternate placement for the group-by operator. Observe that the transformation reduces the space of choices for join ordering since the ordering is considered only within each query. Prior work on group-by has addressed the problem of pipelining group-by and aggregation with join [5, 6] as well use of group-by to flatten nested SQL queries [7, 5, 8, and 9]. But, these problems are orthogonal to the problem of optimizing queries containing group-by clause. E Preliminaries and Notation We will follow the operational semantics associated with SQL queries [10, 11]. We assume that the query is a single block
SQL query, as below.

Select All <columnlist> AGG1 (bl) AGG2 (bn)
From <tablelist>
Where cond1 And cond2 . . . And condn
Group By col1,..col2

The WHERE clause of the query is a conjunction of simple predicates. SQL semantics require that <columnlist> must be among col1,.. colj. In the above notation,AGGi…..AGGn represent built-in SQL aggregate functions. In this paper, we will not be discussing the cases where there is an ORDER BY clause in the query. We will also assume that there are no nulls in the database. These extensions are addressed in . We refer to columns in {b1, ..bn} as the aggregating columns of the query. The columns in (col1, ..colj} are called grouping columns of the query. The functions{AGG1, ..AGGn} arecalled the aggregating functions of the query. For the purposes of this paper, we included Avg and Count as well as cases where the aggregate functions apply on columns with the qualifier[1],[2].
Optimization To illustrate the object oriented query optimizations consider the same example of Retail Banking system.

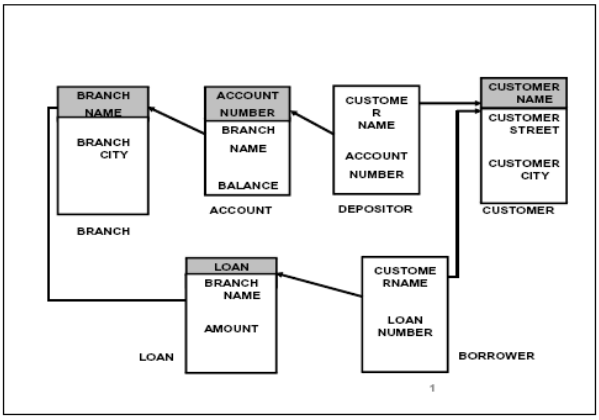


Fig-Class Diagram

C.Query Optimization in OODB

As an example, consider the same Query suppose we want to find the number of branch in each city except pune

Creation of Object Oriented Type

Create Type Branchdet_Ty as Object
(Branch_City Varchar2 (30),
Assets Number (26, 2));

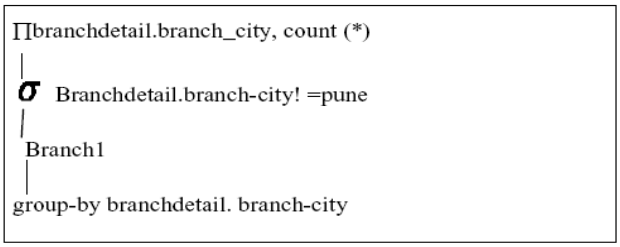
Create Type Accountdet_Ty as
Object (Branch_Name Varchar (30),
Balance Number (12, 2));

Creation of Object Oriented Table

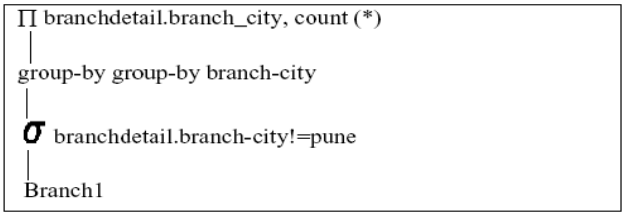
Create Table Branch1 (Branch_Name Varchar2 (30) Primary Key,
Branchdetail Branchdet_Ty);

Create Table Account1
(Account_Number Varchar(15),
Accountdetail Accountdet_Ty);
As an example, consider the above Query suppose we have to find the number of branch in each city except pune

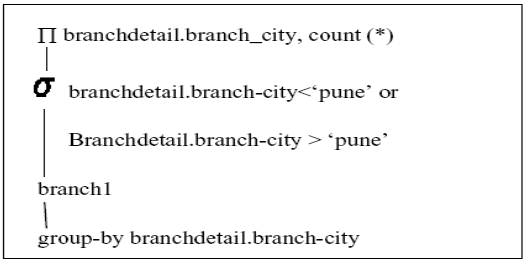
The query evaluation plans for OODB are: -



Plan1



Plan 2



Plan 3

D. Cost Estimation

Given a query there are many equivalent alternative algebraic expression for each expression there are many ways to implement them as operators. The cost estimates are based upon I/O, CPU and Memory resources required by each query operation, and the statistical information about the database object such as Table, Indexes and Views. In a large number of systems, information on the data distribution on a column is provided by *histograms*.

Fig. Shows a histogram for query performance for OODB .A histogram divides the values on a column into k buckets. In many cases, k is a constant and determines the degree of accuracy of the histogram. However, k also determines the memory usage, since while optimizing a query; relevant columns of the histogram are loaded in memory. There are several choices for "bucketization" of values. In many database systems, equi-depth (also called equi height) histograms are used to represent the data distribution

on a column. If the table has n records and the histogram has k buckets, then an equi-depth histogram divides the set of values on that column into k ranges such that each range has the same *number* of values, i.e., n/k. compressed histograms place frequently occurring values in singleton buckets. The number of such singleton buckets may be tuned. It has been shown in that such histograms are effective for either high or low skew data. One aspect of histograms relevant to optimization is the assumption made about values within a bucket. For example, in an equi-depth histogram, values within the endpoints of a bucket may be assumed to occur with uniform spread. A discussion of the above assumption as well as a broad taxonomy of histograms and ramifications of the histogram structures on accuracy appears in . In the absence of histograms, information such as the *min* and *max* of the values in a column may be used. However, in practice, the second lowest and the second highest values are used since the *min* and *max* have a high probability of being outlying values. Histogram information is complemented by information on parameters such as number of distinct values on that column although histograms provide information on a single column they do not provide information on the *correlations* among columns. In order to capture correlations, we need the *joint* distribution of values. One option is to consider 2-dimensional histograms. Unfortunately, the space of possibilities is quite large. In many systems, instead of providing detailed joint distribution, only summary information such as the number of distinct pairs of values is used.

For example, the statistical information associated with a multi-column index may consist of a histogram on the leading column and the total count of distinct combinations of column values present in the data

We achieved statistically significant improvement in the quality of plans with a modest decrease in the optimization cost. The experiments were conducted using on oracle database Table: 1 shows the Query Performance of OODBMS Based on Cost, Cardinality & No of Bytes. From experimental setup we observed that there is significant improvement after query optimization in object oriented database.[1],[2].

Fig: Query Performance of OODB for GROUP BY Clause

Object Oriented Database (OODB)			
Group By Clause			
Plans	Query Performance		
	Cost	Card	Bytes
Plan1	11	8	102
Plan2	11	123	2091
Plan3	11	12	204

Fig 6: Query performance Histogram for OODB

VIII. IMPLECATIONS

We propose a new approach that permits to enrich technique of query optimization existing in the object-oriented databases and the comparative analysis of query optimization for relational databases and object oriented database based on cost, cardinality and no of bytes. Seen the success of query optimization in the relational model, our approach inspires itself of these optimization techniques and enriched it so that they can support the new concepts introduced by the object databases.

We also proposed study of ***Search space*** must be able to include plans that have a low cost. The ***costing technique*** must be accurate. The ***enumeration algorithm*** that searches through the execution space must be efficient. [1],[2],[3],[11]

IX. APPLICATIONS

Query optimization in case of OODB is being successfully implemented in various areas of applications where need of complexity is major issue in case of performance, data is efficiently manipulated and evaluated in case of large complex databases where RDBMS degrades in performance. Areas of applications are[1],[2].[10]

a) CAD/CAM, Geographic information systems
b) Multimedia databases
c) Imaging and Graphics Applications
d) Modeling based engineering applications(Eg.Aircraft Simulator)

X. CONCLUSION

One of the biggest problems in Object Oriented Database is the optimization of queries. Due to these problems optimization of object-oriented queries is extremely hard to solve and is still in the research stage. This proposed work is expected to be a significant contribution to the Database Management area which will not only reduce time or efforts but will also improve the quality and will reduce the cost.[9],[10]

XI. REFERENCES

[1]" Query Optimization in Object-Oriented Database Management Systems: A short review" Abhijit Banubakode et al. / International Journal of Computer Science & Engineering Technology (IJCSET**),[2010],**volume 1.

[2] "Contribution to the Query Optimization in the Object-Oriented Databases" Minyar Sassi, and Amel Grissa-Touzi Volume 6 ISSN June**[2005]** 1307-6884

[3] DATABASE SCALABILITY AND CLUSTERING-How Data Clustering Can Benefit Performance ,A Versant Whitepaper,Database Thery by David Maier**,[2007]**

[4] R. Breitel, D. Maier, A. Otis, J.Penney, B. Schuchardt, J. Stein, H.Wiliams, and M. Williams, "*The Gemstone data magement system. In Object Oriented Concepts, Databases and Applications*", eds. W. Kim and F. H. Lochovsky **[1988].**

[5] F. Bancilhon, S. Cluet, and C. Delobel, "*A query language for O2. In Building an Object-Oriented Database System-The Story of O2*",Morgan Huffman Publishers, San Mateo, Ca., **[1992].**

[6] B.P. Jenq, D. Woelk, W. Kim, and W.-L. Lee, "*Query processing in distributed ORION*. In Proc. EDBT", Venice, Italy, **[1990].**

[7] A. Kemper, G. Moerkotte, and K. Peithner, "*A blackbord architecture for query optimization in object bases*. In Proc. Int. on Very Large Data Bases", Dublin, Ireland, August **[1993].**

[8] R.G.G. Cattell: Object Data Management - Object-Oriented and Extended Relational Database Systems; Addison-Wesley. ISBN 0-201-53092-9

[9]"Query processing in ODBMS"by M.Tamer Ozsu and JoseA. Blakeley,**[1989]**

[10] D. S. Batory. Extensible cost models and query optimization in GENESIS. IEEE Database Engineering,10(4), Nov **[1987]**.

Dynamic Recognition of Malicious Routers

Hemanth S
Assistant Professor, Vel Tech University, Chennai, India
G Sudhakar
Software Engineer,TCS, Hyderabad, India
Chaitanya K
Software Engineer, Infosys, Hyderabad, India

Abstract

In this paper, we considered the problem of detecting whether a compromised router is maliciously dropping packets in the network. Packet dropping from a network of two reasons those is congestion route and malicious attacks. In particular, we are concerned with a simple yet effective attack in which a router selectively drops packets destined for some victim. Unfortunately, it is quite challenging to attribute a missing packet to a malicious action because normal network congestion canproduce the same effect. Modern networks routinely drop packets when the load temporarily exceeds their buffering capacities.Previous detection protocols have tried to address this problem with a user-defined threshold value but in this method we added the buffer size dynamically, because of this congestion get removed as possible. Goal is to differentiate the packet dropping of congestion route from the malicious attacks with protocol X. The proposed method includes broadcasting and also used for large networks.

Keywords: *Malicious attacks, compromise routers, DoS attacks, Broadcasting.*

1. Introduction

The Internet is on the mode of turning the worldwide communication network, and then desires to offer various services with assured quality for all kinds of applications [1]. From last 20 years have been seen an enormous [2] increase of the Internet. Several services of socio-economic interest in societytoday, many of them involving critical considerations, are offered over the Internet. Their exposure to the comprehensive networking environment leaves them susceptible to dissimilar types of computer attacks, amongst which DoS (Denial of Service) attacks, due to their high alike catastrophic index, are decorated [3].

Among these incidents, Denial-of-Service (DoS) attacks cause one of the mostserious threats to internet service applications [2]. Such attacks are not simple theoretical curiosities, but they are vigorously employed in practice. Attackers have continually confirmed their ability to compromise routers, through combinations of social engineering and exploitation of weak passwords or latent software vulnerabilities [4], [5], [6]. This paper addresses the increasing security problem regarding malicious attacks of a particular router in a network.

Ambiguity around packet losses can be resolved [7] using traffic validation protocol, absence of packet be seen as malicious or benign. Three approaches to detect the packet loss are:

1) Static threshold
2) Traffic modeling
3) Traffic measurement.

In the every approach mentioned above packet loss is due to malicious intent and as our proposed model focuses on malicious packet loss, it satisfies the all approaches.

In this paper, we developed a protocol x that dynamically infers the precise number of congestive packet losses that will occur as previous work carried out statically. In the previous work the congestion ambiguity is removed by retransmitting the packets but in our proposed we removed the congestion ambiguity by setting the queue size unlimited. In the previous work link state routing protocol is used to find the shortest path between source and destination for packet transmission in the proposed work we considered distance vector routing protocol. By using the link state routing protocol flooding occurs by using distance vector routing protocol this flooding can be eliminated. As there are number of algorithms in distance vector routing [9] protocol but for efficient purpose we considered DIJKSTRA's algorithm. In this method we broadcast the packet, and neighbor to the router may receive the packet where in previous work unicast is considered. In previous work the proposed protocol was evaluated on small experimental network but in our work we extended for large networks also.

2. Background

Previous works related to the malicious attacks worked [10] on the uni-casting of packets to the redirectors. Broadcasting is not supported, it is not taking [6] into consideration about network parameters like network size, network delay, dynamic routing. Instead, we have focused on the less well-appreciated threat of an attacker subverting the packet forwarding process on a compromised router. Such an attack presents a wide set of opportunities including DoS, surveillance, man-in-the-middle attacks, replay and insertion attacks, and so on. Moreover, most of these attacks can be trivially implemented via the existing command shell languages in commodity routers.

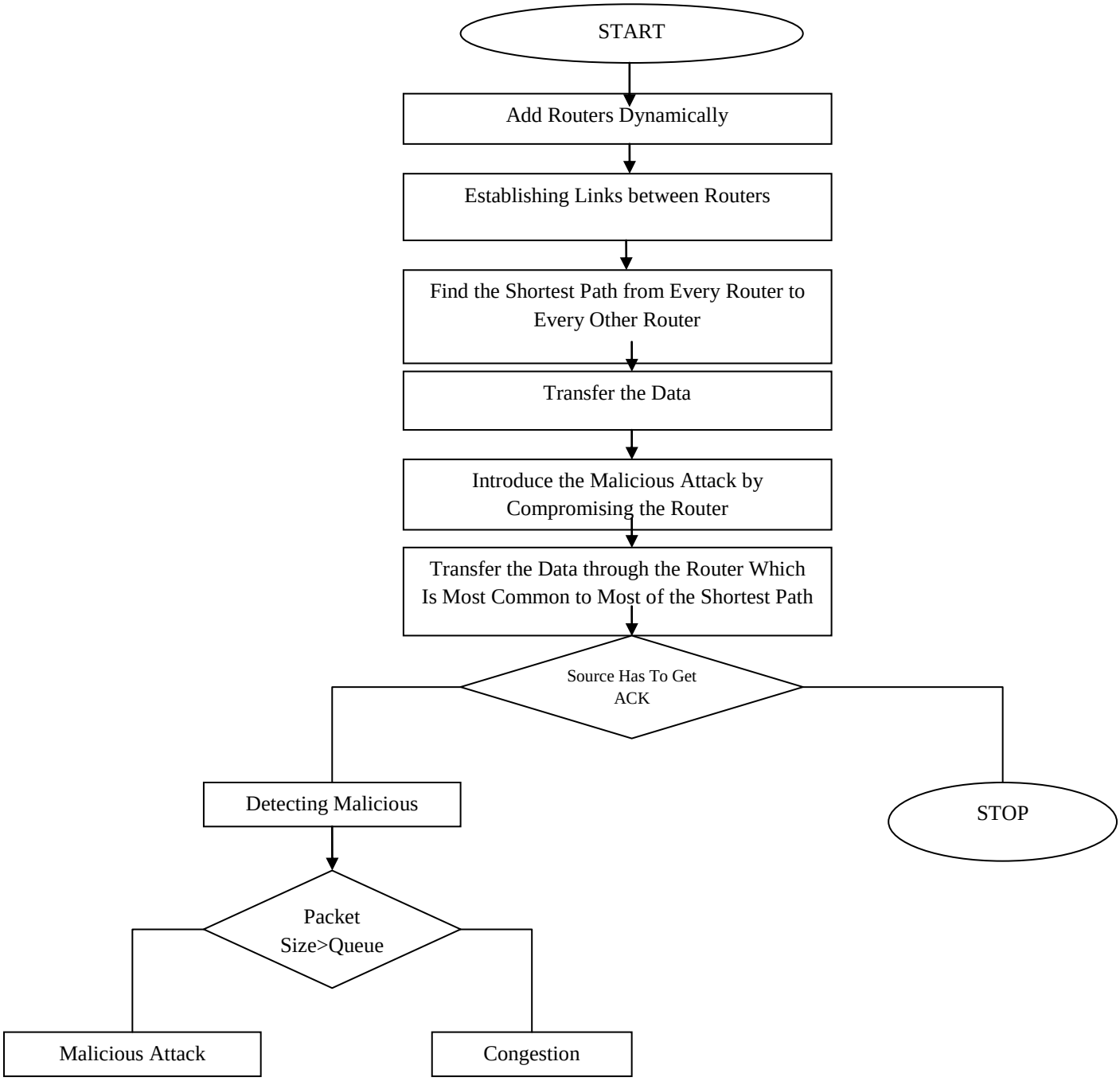
The term routing encapsulates two tasks. These tasks are deciding the paths for data transferred and sending the packets on these paths. The routing is a process that is a function carried out at layer 3 of the OSI reference model. The routing algorithm decides the output line to transfer the incoming packets. The routing algorithms are based [8] on the routing protocol that uses metrics to assess whether a particular path is the optimal path available for transfer of the data packets. The metrics used for evaluating the paths are bandwidth, delay and reliability. The routing algorithms use these protocols to determine an optimal path from the source to the destination. The routing tables maintain all the information related to routing. There are various routing algorithms and depending on these routing algorithms, the information stored in the routing table varies. Every router has its own routing table and it fills this table with the required information to calculate the optimal path between the source router and the destination router.

3. Proposed Model

Packets are forwarded from every router to every other router based on the shortest path via distance vector routing protocol such as DIJKSTRA's algorithm. There is a less possibility to drop the packets due to congestion because in this model every router maintains a queue with some size without limitation. In this model packets are forwarded in broadcast manner to its neighbors. This model can be used to biggest networks also.

3.1 Methodology

The following flow chart shows detecting the malicious intent or congestion route:



The steps in the flow chart are described as:

1. Build a network using direct point to point links between routers.
2. Add routers dynamically in a network and links also.
3. Finding the shortest paths from every router to every other router.
4. Maintains a Queue with some size at every router for stores the incoming packets.
5. Transfers the data through packets from some sources to particular router.
6. Introduce the malicious attacks by compromising node.
7. Transfer the data through the router which is common to most of the shortest paths.
8. If source router will get the Acknowledgement from the destination router then stop.
9. Otherwise detect the malicious intent.
10. At that particular router, consider the sizes of incoming packets and Queue.
11. If PS is less than Q, then the loss can be considered as malicious attack.
12. Otherwise loss can be considered due to Congestion.

PS = Incoming packets size.
Q = Queue size.

The following are different modules in the algorithm:

3.2 Network Model

Consider a network that having individual homogeneous routers connected via point to point links using digraph.

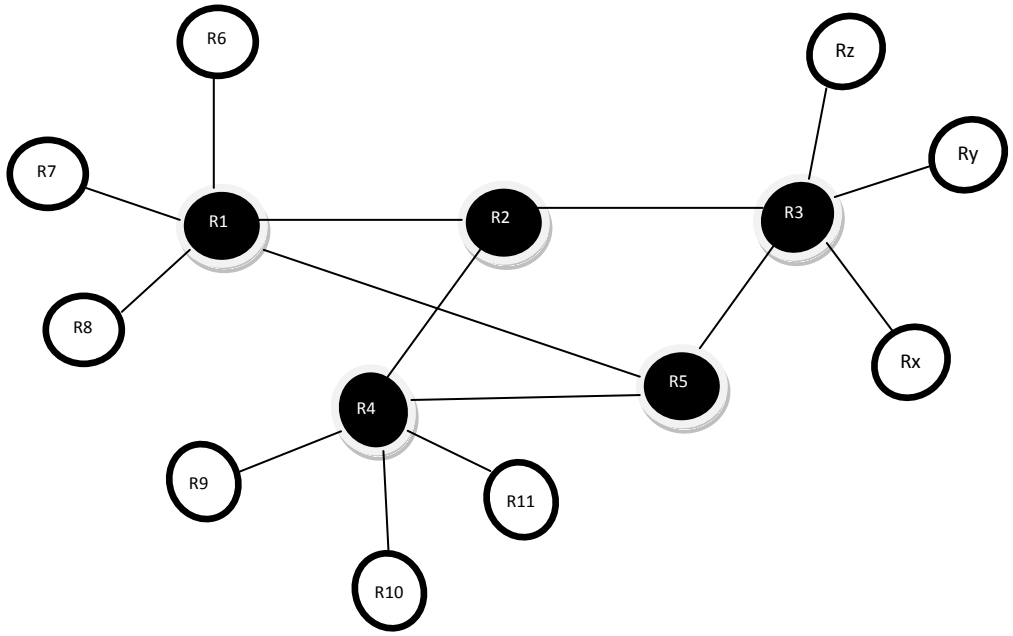


Figure 1: Representation of Network Model

The figure 1 shows the graph of network model using direct point to point links between the routers. That is represented as G= (V, E), where V is the set of homogeneous routers and E is the set of directed links among routers. We can add the routers and links among routers on demand.

In network model routers are connected using direct point to point links in a star topology manner. These star topologies of different networks are connected using different LANs.
We define a path to be a sequence **<R_1,R_2,…….,R_n>** of adjacent routers. A path defines a sequence of routers a packet can follow. In this path first router is source and the last router is sink both routers are called terminal routers. If a network consists of a single path **<R_1,R_2,R_3,R_4>** , then **<R_1,R_2>** and **<R_2,R_3>** are two path segments. But **<R_2,R_4>** is not a path segment because **R_2** and **R_4** are not adjacent.

3.3 Protocol X

Packet dropping can be detected using Protocol X. Initially every router will maintain a Queue with some size. If the packet can be transferred from a source to some destination, then many redirectors can be participated. If more than one router will feed the data into the routers Queue, then packets may be forwarded or dropped. If that router is compromised then it will be blocked and drop the packets or it will misguide the route. In that case calculate the incoming packet size with the Queue size. If Queue size is less than the incoming packet size at a particular time't', then find out that whether the packet is dropped due to congestion, or due to malicious attack. Protocol x detects the traffic faulty routers by validating the Queue of each output interface for each router. Given the buffer size and the rate at which traffic enters and exits a Queue, the behavior of the Queue is determined. If the actual behavior is deviates then the failures occurred.

In traffic validation (TV): what information is collected about traffic and how it is used to determine that a router has been compromised.

Consider the Queue Q in a router r associated with the output interface of link <R, r_d>. the neighbor routers r_{s1} ,r_{s2} ,r_{s3} ,……, r_{sn} feed the data into Q.

$T_{info}(r, Q_{dir}, \pi, t)$ is the traffic information collected by router r that traversed path segment π over the time interval t. Q_{dir} is either Q_{in} or Q_{out} .

Q_{in} is traffic into Q.

Q_{out} is traffic out of Q.

At an abstract level we represent the traffic, a validation mechanism associated with Q, as a predicate TV (Q, q_{pred}(t), S,D), where

q_{pred}(t) is the predicated state of Q at time t.

S= {$\forall i \in \{1,2,\ldots,n\}$: $T_{info}(r_s, Q_{in}, <r_s, r, r_d>, t)$} is a set of information coming into Q as collected by neighbor routers.

D = $T_{info}(r_d, Q_{out}, <r, r_d>, t)$ is the traffic information outgoing traffic from Q collected at router r_d.

TV(Q, $q_{pred}(t)$,S,D) evaluates to false if and only if r was traffic faulty and dropped packets maliciously during time t. T_{info} is represented in different ways. We use three-tuple for each packet traversing Q includes: fp – fingerprint of packet, ps – packet size and the time that

The packet entered or exited based on Q_{dir} , i.e. Q_{in} or Q_{out}.

Practically, the behavior of queue cannot be predicted with complete accuracy. Let $q_{act}(t)$ is the actual length at time t . Based on central limit theorem [11], our assumption tells us that the error, $\mathbf{q_{error} = q_{act} - q_{pred}}$, can be approximated with normal distribution. This suggests the packet loss tests by using this formula.

i.e. $\mathbf{C_{single}}$= Prob (fp is maliciously dropped).

=prob(there is enough space in the queue to buffer fp).

=prob($q_{act} + ps \leq q_{limit}$).

=prob($X + q_{pred}(ts) + ps \leq q_{limit}$). Where X is a random variable $X= q_{act}(ts) - q_{pred}(ts)$.

=prob ($X \leq q_{limit} - q_{pred} - ps$).

=prob($Y \leq (q_{limit} - q_{pred}(ts) - ps - \mu)/\sigma$). Where $Y=(X - \mu)/\sigma$.

=prob($Y \leq y_1$). Where $Y_1= (q_{limit} - q_{pred}(ts) - ps - \mu)/\sigma$.

$\mathbf{C_{single}}= (1+erf(y_{1/\sqrt{2}}))/ 2$. erf is the error function.

3.4 Router Configuration

Every router is having IP address and port number and these are maintained by routing table. Router always must be in listening mode for network sniffing. It will maintain a packet Queue to store incoming packets. Here assume that size of the Queue is fixed. The role of the router is any one of the source, redirector or destination. In general redirectors will be compromised.

3.4.1 ROUTING TABLE

A routing table is a document stored in the router or a network computer. The routing table is stored in the form of a database or is simply a file stored in the router. The data entered in the routing table is referred to when the best possible path to transfer information across two computers in a network is to be determined. The two classifications, viz., static and dynamic routing, are based on the way in which the routing tables are updated every time they are used. The routers in which the data is stored and updated manually are called static routers. On the other hand, the routers, in which the information is changed dynamically, by the router itself, are referred to as dynamic routers.

Cost matrix of routing table:

Distance between two routers is calculated using IP-Address of routers in real networks .In this case we have to use co-ordinates of the routers which are located at different places using Maps.

999.0	999.0	999.0	321.22	999.0
999.0	999.0	999.0	259.16	999.0
999.0	999.0	999.0	307.38	999.0
321.22	259.16	307.38	999.0	218.66
999.0	999.0	999.0	999.0	999.0

Above matrix shows the distances between every router to every other router. This matrix is used for maintains a routing table internally at every router. It is used for selecting a convenient route for transmitting the packet. Based on the adjacent matrix a cost matrix was constructed using the (x,y) coordinates of any two router, then distance would be obtained if both coordinates are known.

THREAT MODEL

Initially find out the shortest path from every router to every other router, later consider the router having maximum incoming sources, then we try to compromise that router, so that there is chance to evaluate the performance of protocol x with effective manner. Our model is easily extended to address other attacks discussed in [12,13,14], such as packet modification or reordering. In a topology any router can harm the incoming packets due to virus attacks, in this case that router changes the data format or encrypt the actual data format and sends that data to the destination router without compromising the router. A threat model is to introduce malicious attacks at a particular router; the following figure shows a sample threat model.

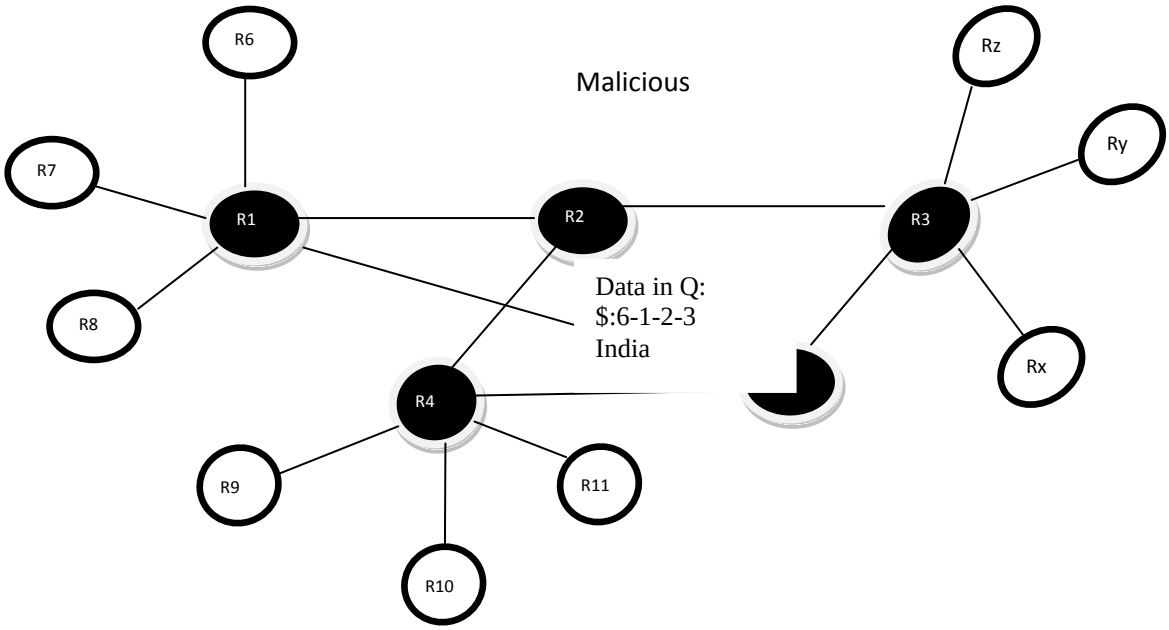


Figure 2: Threat model showing data at malicious router

In figure 2 R2 router represent malicious, if the data comes to that malicious router the further continuation of data is not possible.

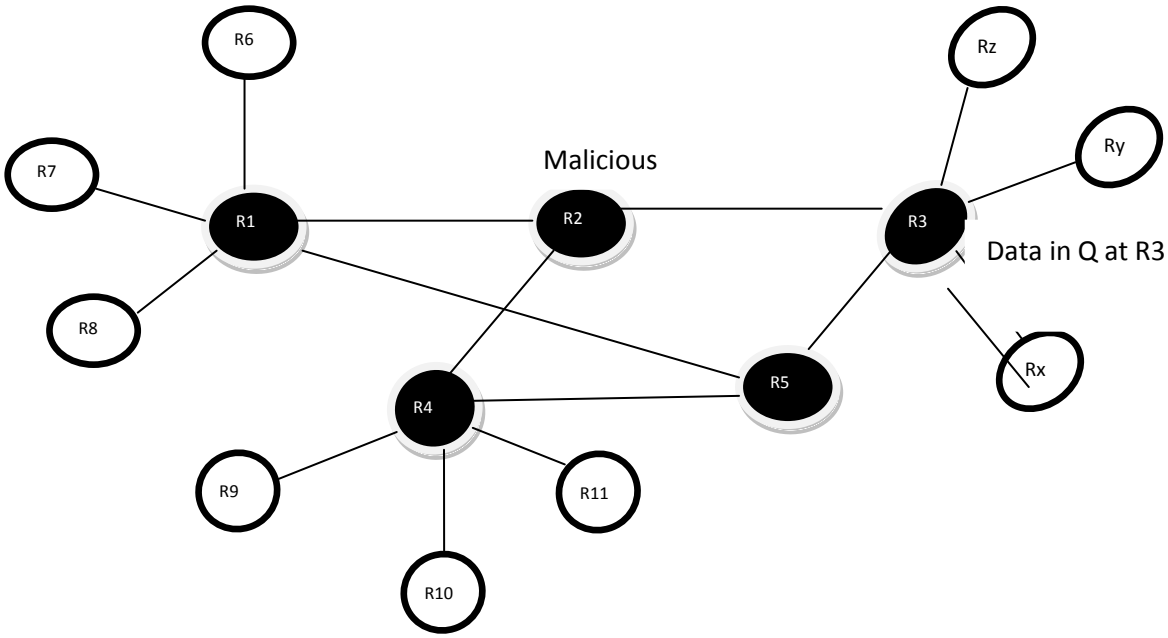


Figure 3: Threat model showing that no data is received from malicious router.

Figure 3 represents that the queue of the router that has receive data from malicious router is empty. If the malicious router appears in the shortest path, then another shortest path needed to be identified to send the packet to destination.

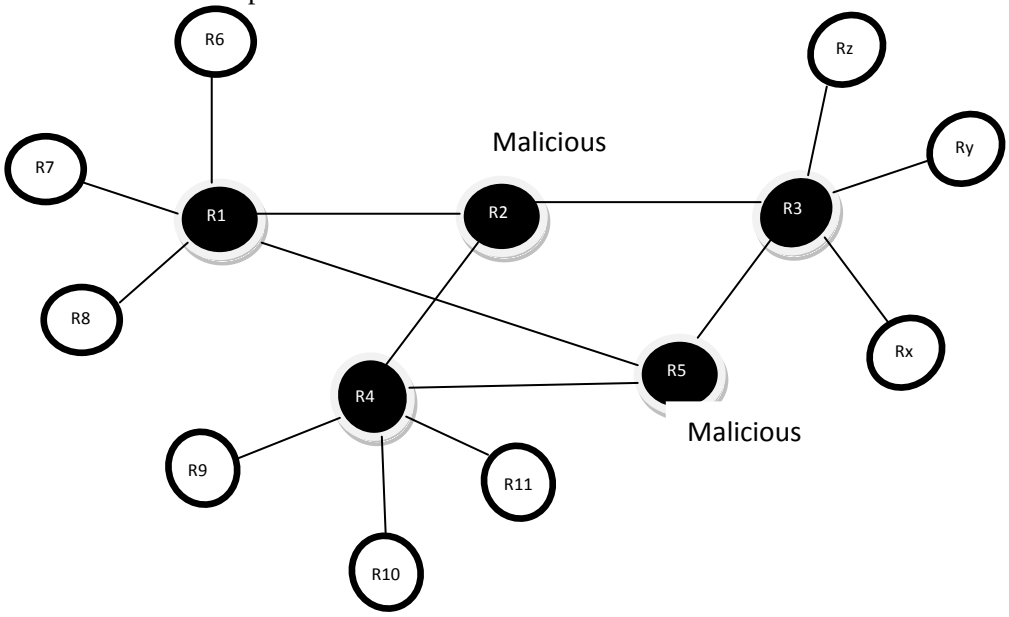


Figure 4: Shows the number of malicious routers in largest networks.

Figure 4 explains in smallest network we have to find the shortest paths from every router to every other router, from those shortest paths select the router which is most common router in those paths. In this network only one router is able to do the compromise for malicious intent. But

in large networks there are many smallest networks which are interconnected with LAN or WAN , for this network we are able to do more than one router as a compromise for malicious intent, why because if only one router is compromised in one LAN then the routers which are in other LANs will be transmitting the packets between the routers. That's why we have to do more than one router as compromise for malicious intent.

BROADCASTING

The previous work does not support broadcasting, it only supports [8] unicasting. Unicasting means communication provides from one source to one destination. But our work supports Broadcasting. Broadcasting provides communication from one host or router to it all neighbor hosts or routers. For this, in our work we have to find the neighbors using adjacency matrix. Using this broadcasting we have to send the data or packet at a time to its all neighbors.

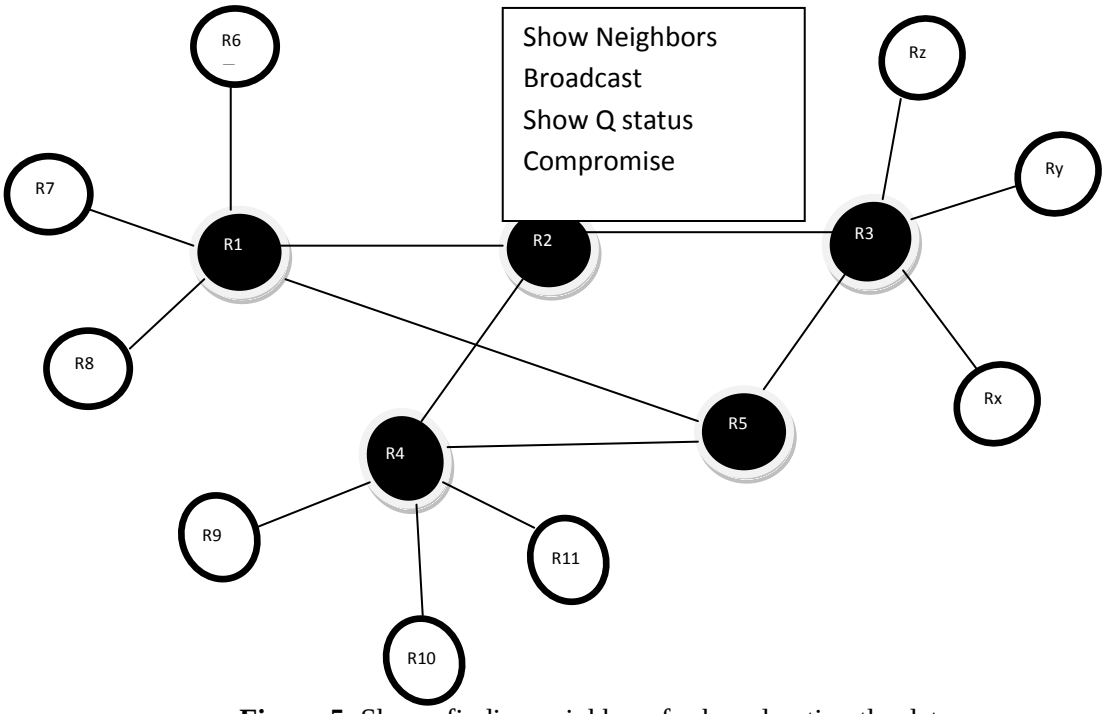


Figure 5: Shows finding neighbors for broadcasting the data.

From figure 5 says:

Show neighbors: find the neighbors for a particular router.

Broadcast: to send the data to its neighbors.

Show Q status: for storing the packets at a router.

Compromise: for malicious attack.

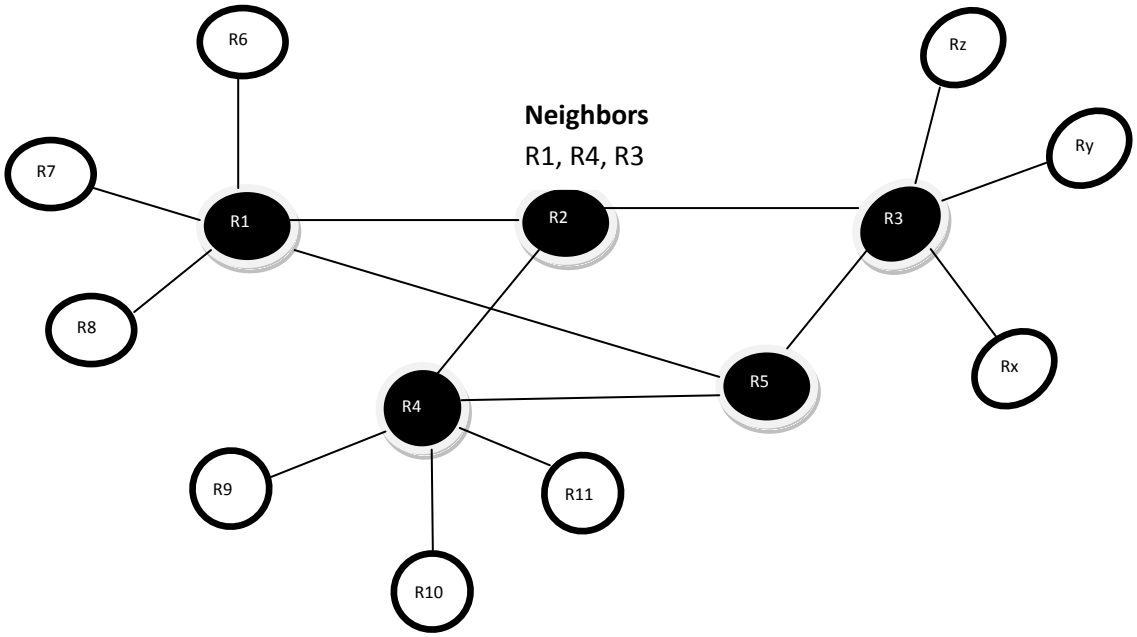


Figure 6: Shows neighbors of any particular router.

The figure 6 shows that neighbors of the particular router, the neighboros are identified using the adjacent matrix.

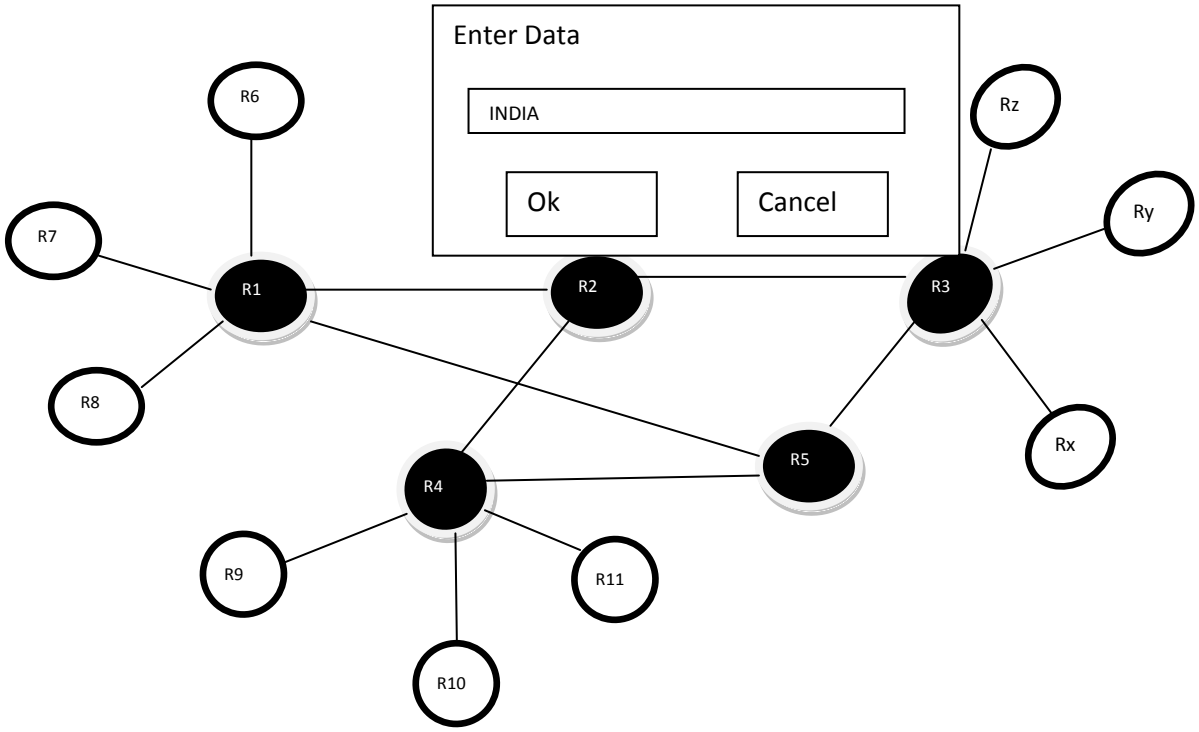


Figure 7: shows the broadcast message

Every router maintains the neighbor's information for broadcasting the message and figure 7 shows the message to be broadcasted.

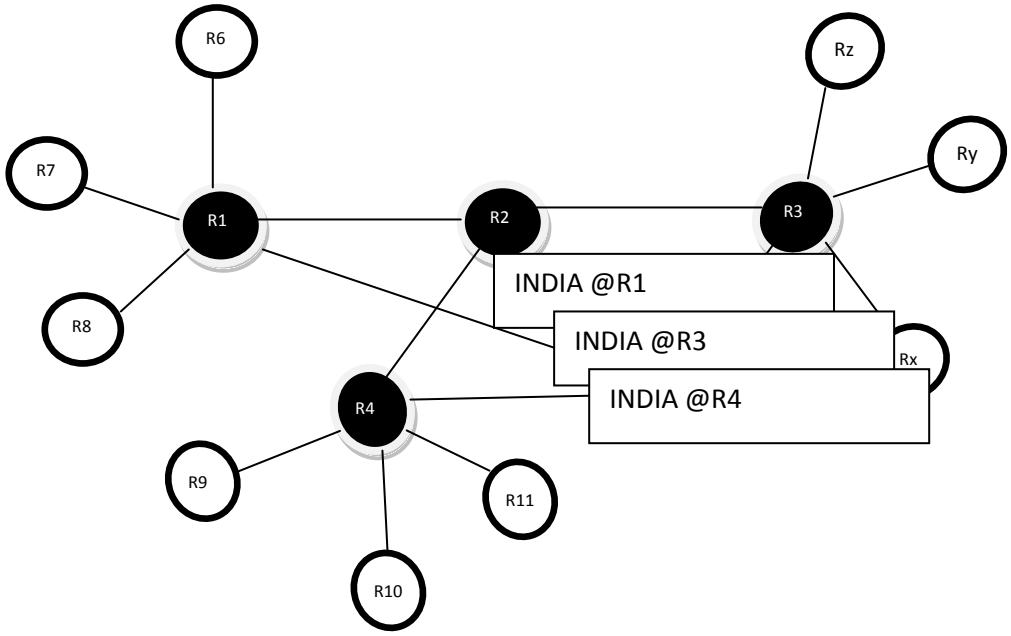


Figure 8: Receive the broadcast message from a router to its all neighbors**.**

Figure 8 show that router R2 broadcasts the data to R1, R3, R4.

The main advantage of proposed model was fast detection, this fast detection can be done by finding the shortest paths among all the routers and prior knowledge about the size of the queue will allow differentiating malicious attack from route congestion quickly. In general malicious attack will be detected if the data size exceeds the queue size of the router. Until that movement we can't imagine that it may be malicious attack. So, the proposed model will wait for an acknowledgement from the receiver. If the ultimate receiver can't acknowledge in mean time, this model will vary the routing path about malicious attacks. If it founds then immediately it will divert the traffic through safety path.

CONCLUSION

In this proposed scheme we consider the scalability of the network i.e. dynamically add the new routers and provide communication with existing network.

We also consider a possibility of attacks in two ways.

1. By making the router which is participating in highest transmission path as compromised router.

2. Based on the selection of any router to compromise.

In both of the situations, if the data is transmitted through that compromised router, further then it cannot forward the packets to the next node in the transmission path.

According to our assumptions there are some refinements are also possible for future work.

Future work:

1. Consideration of mobility of routers. In the sense routers were placed dynamically, but routers cannot move i.e. static.
2. By passing of data transfer from the malicious router after detection i.e., in the transmission path if the malicious router is occurred then find out the alternate path to send the packet to the destination.

3. Intimation about malicious router to the neighboring routers i.e., in our model malicious router can also broadcasting the packets to its neighbors.

REFERENCES

1. P.Owezarski, "On the Impact of DoS Attacks on Internet Traffic Characteristics and QoS" 14th International conference Computer communication and networks, ICCN proceedings,2005.
2. Aye, M.M, "A Queuing Analysis of Tolerating for Denial-of-Service (DoS) Attacks with a Proxy Network" International conference on computer engineering and technology, 2009.
3. D.Pino, M.A.Perez, P.Garcia, P.Fernandez, C.P.Suarez, "Towards self-organizing maps based Computational Intelligent System for denial of Service Attacks Detection", 14th international conference on Intelligent engineering systems, 2010.
4. X.Ao, "Report on DIMACS workshop on large scale Internet Attacks." 2003.
5. K.J.Houle, G.M.Weaver, "Trends in Denial of Service Attack Technology" 2001.
6. C.Labovitz, A.Ahuja, M.Bailey, "Shining Light on Dark Address space" 2001.
7. A.T.Mizrak, Y.C.Cheng, K.Marzullo, S.Savage, "Detecting and isolating Malicious Routers" IEEE transactions on Dependable and secure computing, Vol 3, Iss. 3, pg:230-244, 2006.
8. W.Khan, LB.Le, E.Modiano, "Autonomous routing algorithms for networks with wide-spread failures" Military communications conference 2009.
9. D.C.Lee, "Proof of a modified Dijkstra's algorithm for computing shortest bundle delay in networks with deterministically time-varying links" IEEE transactions of communications letters, Vol:10, iss:10, pg:734-736, 2006
10. A.T.Mizrak, S.Savage, K.Marzullo, "Detecting Malicious Packet Losses" IEEE Transactions on Parallel and distributed systems, Vol.20,No.2,2009.
11. R.J.Larsen, M.L.Marx, "Introduction to Mathematical Statistics and its Applications", 4th edition, Prentice Hall 2005.
12. S. Kent, C. Lynn, J. Mikkelson, and K. Seo, "Secure Border Gateway Protocol (Secure-BGP)," IEEE Journal on Selected Areas inCommunications, vol. 18, no. 4, pp. 582–592, Apr. 2000.
13. Y.-C. Hu, A. Perrig, and D. B. Johnson, "Ariadne: A secure on-demand routing protocol for ad hoc networks," in The 8th ACM Int.Conf. on MobiCom, Sep 2002.
14. S. Cheung, "An efficient message authentication scheme for link state routing," in *ACSAC*, 1997, pp. 90–98.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-482-489 ISSN: 2249-6645

Artificial Neural Network based Image Compression using Levenberg-Marquardt Algorithm

Pranob K Charles[1],Dr. H.Khan[2],Ch.Rajesh Kumar[3],N.Nikhita[3]
Santhosh Roy[3],V.Harish[3],M.Swathi[3]
[1]*Associate professor, Department of ECE, K.L.University, Guntur, A.P, India*
[2]*Professor*,HOD, *Department of ECE, K.L.University, Guntur, A.P, India*
[3]*Project Students , Department of ECE, K L University, Guntur, A.P, India*

ABSTRACT
Uncompressed multimedia (graphics, audio and video) data requires considerable storage capacity and transmission bandwidth. Despite rapid progress in mass-storage density, processor speeds, and digital communication system performance, demand for data storage capacity and data-transmission bandwidth continues to outstrip the capabilities of available technologies. Image compression is one of the popular image processing technologies. We have used an adaptive method for image compression based on complexity level of the image and modification on levenberg-marquardt algorithm for MLP neural network learning is used. In this method different back propagation artificial neural networks are used as compressor and de-compressor and it is achieved by dividing the image in to blocks, computing the complexity of each block and then selecting one network for each block according to its complexity value. The algorithm used has good convergence. This reduces the amount of oscillation in learning procedure. To realise this method practically, multilayer neural (input,hidden and output layers) networks are used.

Keywords – **Artificial neural network, MLP, Training, Compression, Complexity.**

I. INTRODUCTION
Neural networks are inherent adaptive systems, they are suitable for handling non stationeries in image data. Artificial neural network can be employed with success to image compression. The greatest potential of neural networks is the high speed processing that is provided through massively parallel VLSI implementations. The choice to build a neural network in digital hardware comes from several advantages that are typical for digital systems. The crucial problems of neural network hardware are fast multiplication, building a large number of connections between neurons, and fast memory access of weight storage or nonlinear function look up tables.

The most important part of a neuron is the multiplier, which performs high speed pipelined multiplication of synaptic signals with weights. As the neuron has only one multiplier the degree of parallelism is node parallelism. Each neuron has a local weight ROM (as it performs the feed-forward phase of the back propagation algorithm) that stores, as many values as there are connections to the previous layer.

An accumulator is used to add signals from the pipeline with the neuron's bias value, which is stored in an own register.

The aim is to design and implement image compression using Neural network to achieve better SNR and compression levels. The compression is first obtained by modeling the Neural Network in MATLAB. This is for obtaining offline training.

II. NEURAL NETWORKS
2.1 Artificial Neural Network
An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurones. This is true of ANNs as well. Artificial Neural Network is a system loosely modeled on the human brain. The field goes by many names, such as connectionism; parallel distributed

processing, neurocomputing, natural intelligent systems, machine learning algorithms, and aritificial neural networks. It is an attempt to simulate within specialized hardware or sophisticated software, the multiple layers of simple processing elements called neurons. Each neuron is linked to certain of its neighbors with varying coefficients of connectivity that represent the strengths of these connections.

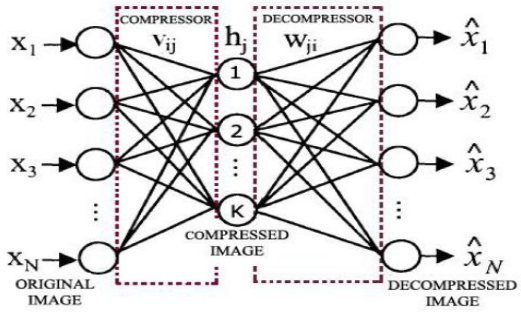


Fig.1.1Basic compression structure

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze.

This expert can then used to provide projections given new situations of interest and answer "what if" questions.

The advantages include Adaptive learning, Self-Organization,Real Time Operation,Fault Tolerance via Redundant Information Coding.

2.2 Layers

Artificial neural network are the simple clustering of the primitive artificial neurons. This clustering occurs by creating layers, which are then connected to one another. How these layers connect may also vary. Basically, all artificially neural networks have a similar structure of topology. Some of the neurons interface the real world to receive its inputs and other neurons provide the real world with the network's outputs. All the rest of the neurons are hidden form view. The input layer consists of neurons that receive input form the external environment. The output layer consists of neurons that communicate the output of the system to the user or external environment. There are usually a number of hidden layers between these two layers; the fig2 below shows a simple structure with only one hidden layer.

When the input layer receives the input its neurons produce output, which becomes input to the other layers of the system.

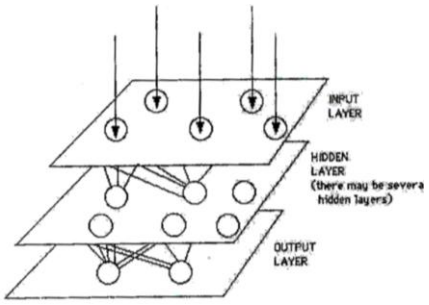


Fig.2.1 *THREE peceptron for image compression*

The process continues until a certain condition is satisfied or until layer is invoked and fires their output to the external environment. To determine the number of hidden neurons the network should have to perform its best, one are often left out to the method trial and error. If the hidden number of neurons are increased too much an over fit occurs, that is the net will have problem to generalize. The training set of data will be memorized, making the network useless on new data sets.

2.3 Learning

The brain basically learns from experience. Neural networks are sometimes called machine-learning algorithms, because changing of its connection weights (training) causes the network to learn the solution to a problem. The strength of connection between the neurons is stored as a weight-value for the specific connection. The system learns new knowledge but adjusting these connection weights. The learning ability of a neural network is determined by its architecture and by the algorithmic method chosen for training. The training method usually consists of one of three schemes:

2.3.1. Unsupervised learning.

Used no external teacher and is based upon only local information. It is also referred to as self-organization, in the sense that it self-organizes data presented to the

network and detects their emergent collective properties. Paradigms of unsupervised learning are Hebbian learning and competitive learning. From Human Neurons to Artificial Neuron other aspect of learning concerns the distinction or not of a separate phase, during which the network is trained, and a subsequent operation phase. We say that a neural network learns off-line if the learning phase and the operation phase are distinct. A neural network learns on-line if it learns and operates at the same time. Usually, supervised learning is performed off-line, whereas unsupervised learning is performed on-line. The hidden neurons must find a way to organize themselves without help from the outside. In this approach, no sample outputs are provided to the network against which it can measure its predictive performance for a given vector of inputs. This is learning by doing.

2.3.2. Reinforcement learning

This incorporates an external teacher, so that each output unit is told what its desired response to input signals ought to be. During the learning process global information may be required. Paradigms of supervised learning include error-correction learning, reinforcement learning and stochastic learning. An important issue concerning supervised learning is the problem of error convergence, i.e. the minimization of error between the desired and computed unit values. The aim is to determine a set of weights which minimizes the error. One well-known method, which is common to many learning paradigms is the least mean square (LMS) convergence**.** This method works on reinforcement from the outside. The connections among the neurons in the hidden layer are randomly arranged, then reshuffled as the network is told how close it is to solving the problem. Reinforcement learning is also called supervised learning, because it requires a teacher. The teacher may be a training set of data or an observer who grades the performance of the network results.

Both unsupervised and reinforcement suffers from relative slowness and inefficiency relying on a random shuffling to find the proper connection weights.

2.3.3. Back propagation

This method is proven highly successful in training of multilayered neural nets. The network is not just given reinforcement for how it is doing on a task. Information about errors is also filtered back through the system and is used to adjust the connections between the layers, thus improving performance. A form of supervised learning.

2.4 Image Compression

Direct transmission of the video data requires a high-bit-rate (Bandwidth) channel. When such a high bandwidth channel is unavailable or not economical, compression techniques have to be used to reduce the bit rate and ideally maintain the same visual quality. Similar arguments can be applied to storage media in which the concern is memory space. Video sequence contain significant amount of redundancy within and between frames. It is this redundancy frame. It is this redundancy that allows video sequences to be compressed. Within each individual frame, the values of neighboring pixels are usually close to one another. This spatial redundancy can be removed fro the image without degrading the picture quality using "Intraframe" techniques.

Also, most of the information in a given frame may be present in adjacent frames. This temporal redundancy can also be removed, in addition to the "within frame" redundancy by "interframe" coding.

III. PROPOSED IMAGE COMPRESSION USING NEURAL NETWORKS (LM ALGORITHM)

3.1 Introduction

A two layer feed-forward neural network and the Levenberg Marquardt algorithm was considered. Image coding using a feed forward neural network consists of the following steps:

An image, F, is divided into rxc blocks of pixels. Each block is then scanned to form a input vector x (n) of size

$$p = r \times c \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad [3.1]$$

It is assumed that the hidden layer of the layer network consists of L neurons each with P synapses, and it is characterized by an appropriately selected weight matrix $\mathbf{W_h}$. All N blocks of the original image is passed through the hidden layer to obtain the hidden signals, h(n), which represent encoded input image blocks, x(n) If L<P such coding delivers image compression.

It is assumed that the output layer consists of m=p=rxc neurons, each with L synapses. Let W_y be an appropriately selected output weight matrix. All N hidden vector h(n), representing an encoded image H, are passed through the output layer to obtain the output signal, y(n). The output signals are reassembled into p=rxc image blocks to obtain a reconstructed image, F_r. There are two error matrices that are used to compare the various image compression techniques. They are Mean Square Error (MSE) and the Peak Signal-to-Noise Ratio (PSNR). The MSE is the cumulative squared error between the compressed and the original image whereas PSNR is the measure of the peak error.

$$MSE = {}^{1}\!/_{MN} \sum_{y=1}^{m} \sum_{x=1}^{n} [I(x,y) - I'(x,y)]^2$$

…… [3.2]

The quality of image coding is typically assessed by the Peak signal-to-noise ratio (PSNR) defined as

PSNR = 20 $\log_{10}$ [255/sqrt(MSE)] …… [3.3]

Training is conducted for a representative class of images using the Levenberg Marquardt algorithm. Once the weight matrices have been appropriately selected, any image can be quickly encoded using the W_h matrix, and then decoded (reconstructed) using the W_y matrix.

3.2Basic Algorithm:

Consider the form of Newton's method where the performance index is sum of squares. The Newton's method for optimizing a performance index F(x) is

$X_{k+1} = X_k - A_k^{-1} g_k$, …….. [3.4]

Where $Ak = \nabla^2 F(x)$ and $g_k = \nabla F(x)$;

It is assume d that F (x) is a sum of squares function:

$$F(x) = \sum_{r=1}^{n} v_i^2(x) = V^T(x) v(x)$$

……[3.5]

Then the j^{th} element of the gradient will would be

$$[\nabla F(x)]_j = \delta F(x) / \delta S_j = 2 \sum_{i=1}^{n} V_i(x) \delta v_i(x) / \delta x_j$$

……[3.6]

The gradient can be written in matrix form:

$\nabla F(x) = 2J^T(x)\, v(x)$ ……[3.7]

Where J(x) is the Jacobian matrix.

Next the Hessian matrix is considered. The k.j element of Hessian matrix would be

$$[\nabla^2 F(x)]_{kj} = \delta^2 F(x) / \delta x_k \, \delta x_j$$

…..…[3.8]

The Hessian matrix can then be expressed in matrix form:

$\nabla^2 F(x) = 2 J^T(x) J(x) + 2 S(x)$; where

$$S(x) = \sum_{i=1}^{n} V_i(x) . \nabla^2 v_i(x)$$

Assuming that S(x) is small, the Hessian matrix is approximated as

$\nabla^2 F(x) \equiv 2 J^T(x) J(x)$ ……[3.9]

Substituting the values of ∇2 F(x) & ∇ F(x), we obtain the Gauss-Newton method:

$X_{k+1} = X_k - [J^T(X_k) J(X_k)]^{-1} J^T(X_k) V(X_k)$ ...…[3.10]

One problem with the Gauss-Newton over the standard Newton's method is that the matrix $H=J^TJ$ may not be invertible. This can be overcome by using the following modification to the approximate Hessian matrix:

$$G = H + \mu I.$$

This leads to Levenberg –Marquardt algorithm

$X_{k+1} = X_k - [J^T(X_k) J(X_k) + \mu_k I]^{-1} J^T(X_k) V(X_k)$…….[3.11]

this algorithm has the very useful feature that as μk is increased it approaches the steepest descent algorithm with small learning rate.

3.3Training Procedure

During training procedure data from a representative image or a class of images is encoded into a structure of the hidden and output weight matrices. It is assumed that an image, F, used in training of size Rx C and consists of rxc blocks.

1. The first step is to convert a block matrix F into a matrix X of size P x N containing training vectors, x(n), formed from image blocks.That is:

$$P= r.c \text{ and } p.N = R.C$$

2. The target data is made equal to the data, that is: D=X

3. The network is then trained until the mean squared error, MSE, is sufficiently small. The matrices W^h and W^y will be subsequently used in the image encoding and decoding steps.

Image Encoding: The hidden-half of the two-layer network is used to encode images. The Encoding procedure can be described as follows:

$$F \rightarrow X, \quad H= (W^h. X) \quad \ldots\ldots\ldots\ldots\ldots\ldots [3.12]$$

Where X is an encoded image of F.

Image Decoding: The image is decoded (reconstructed) using the output-half the two-layer network. The decoding procedure is described as follows:

$$Y = (W^y. H), \quad Y \rightarrow F \quad \ldots\ldots\ldots\ldots\ldots\ldots [3.13]$$

3.4 Algorithm

Step1: Read the test image

Step2: Divide the image into blocks of pixels.

Step3: Scan each block for the complexity level.

Step4: Initialize the neurons.

Step5: Apply scanned vectors to each neuron on the input layer.

Step6: Depending on the weights and the logic involved, perform the operations (TRANSIG).

Step7: Pass them to the hidden layer.

Step8: Again, the same as in step6 (PURELIN).

Step9: Reassemble the outputs.

Step10: Train the neural network and remain the weights.

IV. MATLAB RESULTS AND GRAPHS

In this section, simulation results for different images (64x64) are shown. Their performance measure graphs are also included. Considered the images cameraman.tif and fabric.png form the MATLAB library.

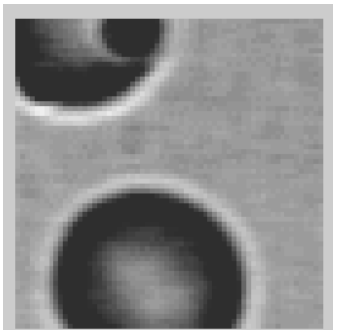


Fig3.1 Original image-1

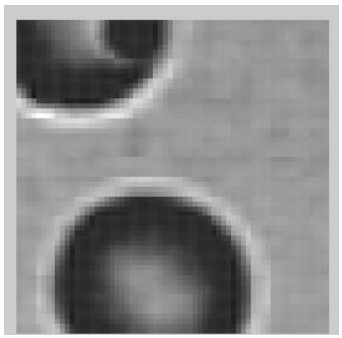


Fig3.2 Encoded (compressed) image-1

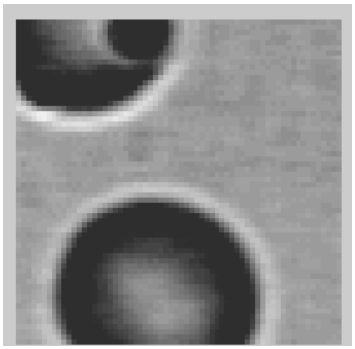


Fig3.3 Decoded (Decompressed) image-2

Fig3.6 Compressed image-2

The performance measure of the applied method can be observed from the graph fig3.4 shown.

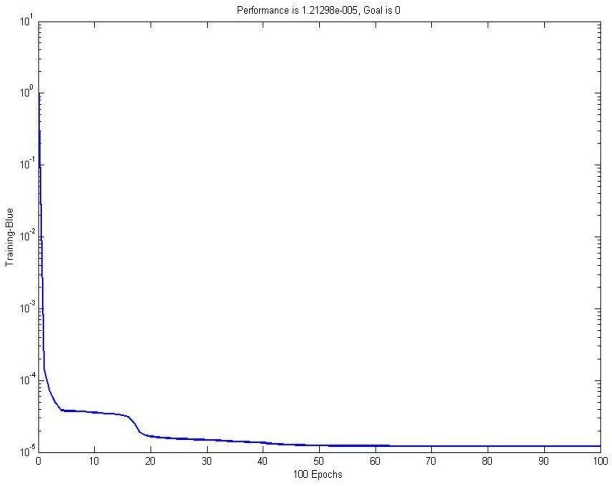


Fig3.4 performance plot

Fig 3.7 Decompressed image-2

Fig3.5original image-2

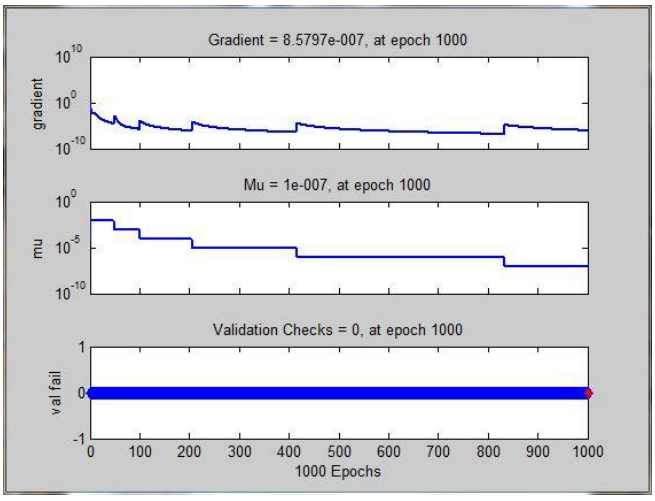


Fig3.8 training state graph

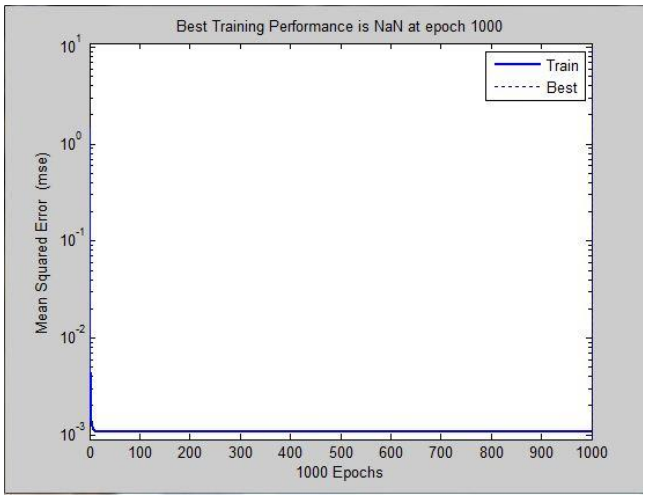


Fig3.9 Performance plot

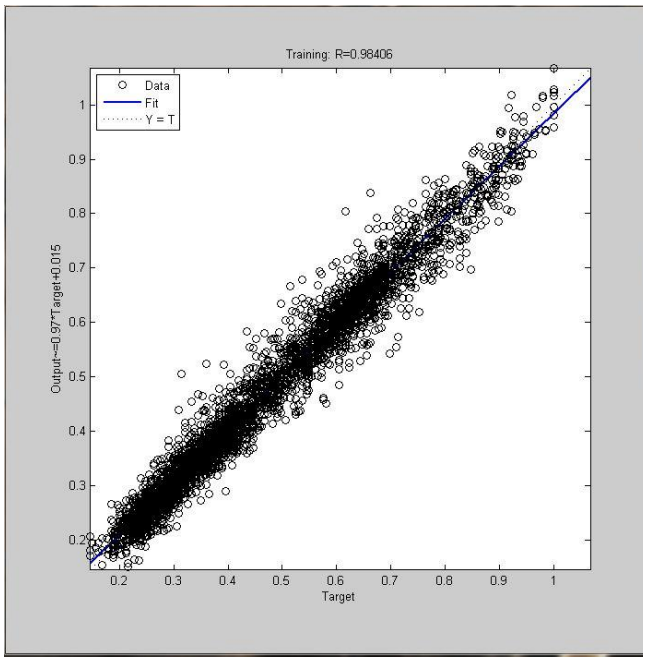


Fig3.10 Regression plot

The default training tool for neural networks provided in MATLAB is fig3.11 below.

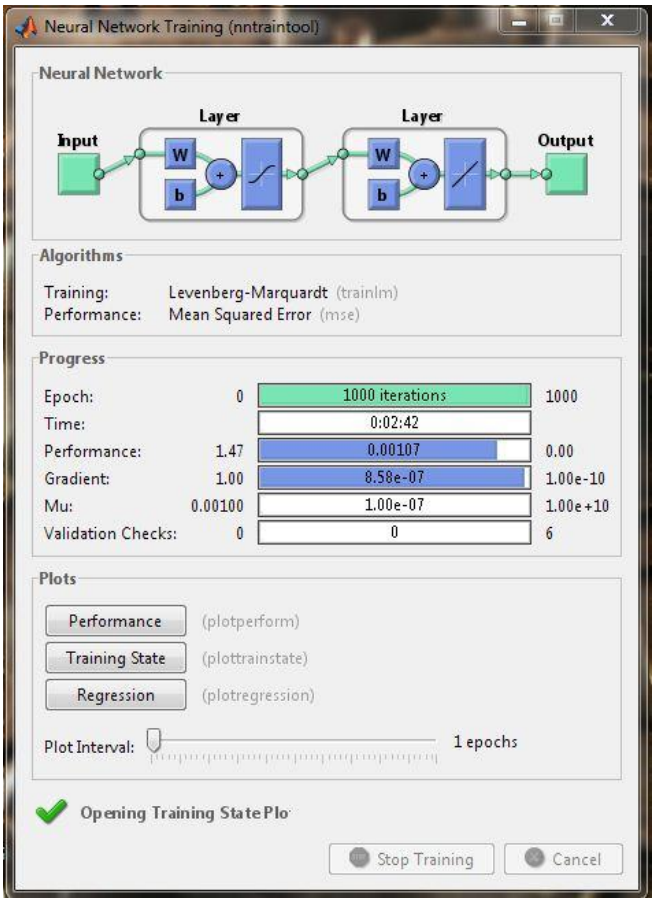


Fig3.11 Neural network training tool(MATLAB)

V. CONCLUSION

The computing world has a lot to gain from neural networks. Their ability to learn by example makes them very flexible and powerful. Furthermore there is no need to devise an algorithm in order to perform a specific task; i.e. there is no need to understand the internal mechanisms of that task. They are also very well suited for real time systems because of their fast response and computational times which are due to their parallel architecture. Neural networks also contribute to other areas of research such as neurology and psychology. They are regularly used to model parts of living organisms and to investigate the internal mechanisms of the brain. Perhaps the most exciting aspect of neural networks is the possibility that some day 'conscious' networks might be produced. There is a number of scientists arguing that consciousness is a

'mechanical' property and that 'conscious' neural networks are a realistic possibility.

Even though neural networks have a huge potential we will only get the best of them when they are integrated with computing, AI, fuzzy logic and related subjects. Neural networks are performing successfully where other methods do not, recognizing and matching complicated, vague, or incomplete patterns.

ACKNOWLEDGEMENTS
The authors like to express their thanks to the management and department of ECE, K L University for their support and encouragement during this work.

REFERENCES

[1] P.J. BURT, E.H. ADELSON, THE LAPLACIAN PYRAMID AS A COMPACT IMAGE CODE, IEEE TRANS. ON COMMUNICATIONS, PP. 532–540, APRIL 1983.

[2] D. SHAO, L.A. MATEOS, W.G. KROPATSCH, IRREGULAR LAPLACIAN GRAPH PYRAMID,CVWW 2010.

[3] S. BECKER AND M. PLUMBLEY, "UNSUPERVISED NEURAL NETWORK LEARNING PROCEDURES FOR FEATUREEXTRACTION AND CLASSIFICATION," 1996, TO APPEAR *INT. J. APPLIED INTELLIGENCE*.

[4] G. W. COTTRELL AND P. MUNRO, "PRINCIPAL COMPONENTS ANALYSIS OF IMAGES VIA BACK PROPAGATION," IN *SPIE VOL. 1001 VISUAL COMMUNICATIONS AND IMAGE PROCESSING '88*, 1988, PP. 1070–1077.

[5] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York, NY: Macmillan, 1994.

[6] A. N. Netravali and J. O. Limb, "Picture coding:A review," *Proc. IEEE*, vol. 68, no. 3, pp. 366–406, March 1980.

[7] A. K. Jain, "Image data compression: A review," *Proc. IEEE*, vol. 69, no. 3, pp. 349– 389, March 1981.

[8] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd ed. San Diego, CA: Academic Press, 1982, vol. I & II.

[9] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice- Hall, 1984.

[10] H. G. Musmann, P. Pirsch, and H.-J. Grallert, "Advances in picture coding," *Proc. IEEE*, vol. 73, no. 4, pp. 523–548, April 1985.

[11]A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression*. New York,NY: Plenum Press, 1988.

[12]A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic Publishers, 1992.

[13]J. A. Storer and M. Cohn, Eds., *Proc. Data Compression Conference*. Snowbird, UT: IEEE Computer Society Press, March 30 - April 2,1993.

[14]N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385–1421, October 1993.

[15] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. New York, NY: Academic Press, 1990.

[16]R. C. Gonzales, R. E. Woods, "Digital Image Processing", Second Edition, Prentice-Hall, 2002.

[17] Veisi H., Jamzad M., "Image Compression Using Neural Networks", Image Processing and Machine Vision Conference (MVIP), Tehran, Iran, 2005.

[18] N.Sonehara, M.Kawato, S.Miyake, K.Nakane, "Image compression using a neural network model", International Joint Conference on Neural Networks, Washington DC,1989.

[19]G.L. Sicuranza, G. Ramponi, S. Marsi, "Artificial neural network for image compression", Electronic letters 26, 477- 479, 1990.

[20]A. Rahman, Chowdhury Mofizur Rahman, "A New Approach for Compressing Color Images using Neural Network", Proceedings of International Conference on Computational Intelligence for Modeling,Control and Automation – CIMCA 2003 ,Vienna, Austria, 2003

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue2, pp-490-494 ISSN: 2249-6645

Human Age Manifold Learning scheme and Curve Fitting for aging features

[1]E.Kannan,[2]Resmi.R.Nair, [3]P.Visu and [4]S.Koteeswaran

[1]Dean, Dept of CSE, Vel Tech Dr.RR & Dr.SR Technical University,Avadi, Chennai-62.
[2]Asst.Professor, Dept of ECE, Vel Tech Engineering College, Chennai – 62.
[3,4]Research Scholar, Dept of CSE, Vel Tech Dr.RR & Dr.SR Technical University,Avadi, Chennai-62.

***Abstract* - The focus of this paper is to estimate the human age automatically via facial image analysis. Age estimation is a type of soft biometrics that provides ancillary information of the users' identity information. It can be used to complement the primary biometric features, such as face, fingerprint, iris, and hand geometry, to improve the performance of a primary (hard) biometrics system. Derived from rapid advances in computer graphics and machine vision, computer-based age synthesis and estimation via faces have become particularly prevalent topics recently because of their explosively emerging real-world applications, such as forensic art, electronic customer relationship management, security control and surveillance monitoring, biometrics, entertainment, and cosmetology. Human faces undergo considerable amount of variations with aging. The Human aging pattern is determined by not only the person's gene, but also by many external factors, such as health, living and weather conditions. Males and females also age differently. Hence, it is a challenging problem for the existing computer vision systems to automatically and effectively estimate human ages.**

In our system we introduce the age manifold learning scheme for extracting face aging features and design a curve fitting and regression method for learning and prediction of human ages. The novel approach improves the age estimation accuracy significantly over all previous methods. Benefits of this proposed approaches for image-based age estimation is shown by extensive experiments on a large internal age database and the public available FG-NET database.

***Keywords* – Face Detection, Face Normalization, Robust Regression, manifold learning,.**

I. INTRODUCTION

Face images convey a significant amount of information including information about the identity, emotional state, ethnic origin, gender, age, and head orientation of a person shown in a face image. The human face conveys important perceptible information related to individual traits. The human traits displayed by facial attributes, such as personal identity, facial expression, gender, age, ethnic origin, and pose, have attracted much attention in the last several decades from both industry and academia since face image processing techniques yield extensive applications in graphics and computer vision fields.

This type of information plays a significant role during face-to-face communication between humans. The use of facial information during interaction is made possible by the remarkable ability of humans to accurately recognize and interpret faces and facial gestures in real time. Current trends in information technology dictate the improvement of the interaction between humans and machines, in an attempt to upgrade the accessibility of computer systems. As part of this effort, many researchers have been working in the area of automatic interpretation of face images so that contact-less human–computer interaction (HCI) [1] based on facial gestures can be developed. In this context, systems capable of identifying faces, recognizing emotions, gender, and head orientation have been developed. Despite the fact that the age of a person plays an important role during interaction, so far no researcher has been involved in designing automatic age estimation systems based on face images. With our work, we aim to produce a system which is capable for estimating the age of a person as reliably as humans.

II. MOTIVATION

The motivation behind our work lies in the important real life applications of the proposed methodology. In summary, those applications include the following.

***Age specific human computer interaction*:** If computers could determine the age of the user, both the computing environment and the type of interaction could be adjusted according to the age of the user [9]. Apart from standard HCI, such a system could be used in combination with secure internet access control in order to ensure that under-aged persons are not granted access to internet pages with unsuitable material. A vending machine, secured by the ASHCI system, can refuse to sell alcohol or cigarettes to the underage people. In image and video retrieval, users could retrieve their photographs or videos by specifying a required

age range. Ad-agency can find out what kind of scroll advertisements can attract the passengers (potential customers) in what age ranges using a latent computer vision system.

***Age-based indexing of face images*:** Automatic age estimation can be used for age-based retrieval of face images from databases. The most common application of this technology is in e-photo albums, where users could have the ability to retrieve their photographs by specifying a required age-range.

***Development of automatic age progression systems*:** Automatic age estimation systems rely on their ability to understand and classify changes in facial appearance due to aging. The methodology required in this task could form the basis of designing automatic age progression systems (i.e., systems with the ability to predict the future facial appearance of subjects). A description of our early work in this area is described elsewhere.

***Understanding the process of age perception by humans*:** Work in the area of automatic age estimation could provide invaluable help to psychologists who study the topic of age perception by humans.

III. RELATED WORK

Automatic image-based human age estimation is an important technique involved in many real-world applications. Estimating human age automatically via facial image analysis has lots of potential real-world applications, it is still a challenging problem to estimate human ages from face images since different individual's age quite differently. The aging process is determined by not only the person's gene, but also many external factors, such as health, living style, living location and weather conditions.

The current age estimation performance is still not good enough for practical use and more effort has to be put into this research direction. The biases in **Age estimation** in many cases are not constant across subgroups of a population.

There are three important methods that can categorize most existing image-based human age estimation technique [7].

- ***Anthropometric Model***
- ***Aging Pattern Subspace***
- ***Age Regression***

A. Anthropometric Model

The cranio-facial development theory and facial skin wrinkle analysis are used to create the anthropometric model [1]. The changes of face shape and texture patterns related to growth are measured to categorize a face into several age groups. These methods are suitable for coarse age estimation or modeling ages just for ***young people*** [7]. However, they are not designed for continuous or refined age classification.

B. Aging Pattern Subspace

To handle highly incomplete data due to the difficulty in data collection, Aging pattern Subspace models [4] a sequence of personal aging face images by learning a subspace. The age of a test face is determined by the projection in the subspace that can best reconstruct the face image. These kinds of methods are designed to deal with the difficulty of utilizing the ***incomplete age databases***.

C. Age Regression

For the regression methods, facial features are extracted by the active appearance models (AAMs) [3] that incorporate the shape and appearance information together. An input face image is then represented by a set of fitted model parameters. The regression coefficients are estimated from the training data with an ***assumption*** of the regression function such as a quadratic model.

IV.PROPOSED SYSTEM

In our paper, refined age estimation technique by age manifold analysis is implemented

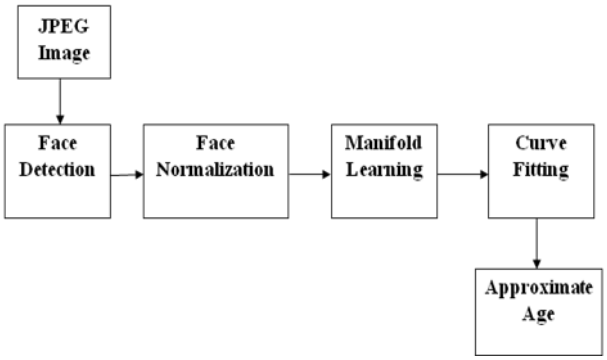


Fig. 1 Age Estimation Process

The age manifold analysis [8] has two advantages to facilitate the age estimation task. First, the manifold analysis is a way to represent the original age data in low dimensionality which is necessary to overcome lack-of-fit of the regression model. Second, the manifold learning captures the underlying face aging structure which is important for accurate modeling and age prediction. Regression gives an efficient methodology for mapping the low dimensionality manifold data into accurate age values.

A. Face Detection

1) ***Color Models*:** Different color spaces used in skin detection previously include HSV, normalized RGB, YCrCb, YIQ and CIELAB [10].HSV gives the best performance for skin pixel detection. We conducted our own experiments independently and converged to the same fact. The experiments also showed the superiority of HSV color space over RGB and YCrCb color spaces [11]. In the HSV space, H stands for hue component, which describes the shade of the color, S stands for saturation component, which describes how pure the hue (color) is while V stands for value component, which describes the brightness. The removal of V component takes care of varying lighting conditions. H varies from 0 to 1 on a circular scale i.e. the colors represented by H=0 and H=1 are the same. S varies from 0 to 1, 1 representing 100 percent purity of the color and S scales.

2) ***Connectivity Analysis:*** Group the skin pixels in the image based on an 8-connected neighborhood i.e. [11] if a skin pixel has got another skin pixel in any of its 8 neighboring places, then both the pixels belongs to the same region. At this stage, we have different regions and we have to classify each of these regions as a human face or not. This is done by finding the centroid, height and width of the region as well as the percentage of skin in the rectangular area defined by the above parameters. The centroid is found by the average of the coordinates of all the pixels in that region. For finding height, the y-coordinate of the centroid is subtracted from the y-coordinates of all pixels in the region. Find the average of all the positive y-coordinates and negative y-coordinates separately. Add the absolute values of both the averages and multiply by 2 [11]. This gives the average height of the region. Average width can be found similarly by using x co-ordinates. Since the height to width ratio of human faces falls within a small range on the real axis, using this parameter along with percentage of skin in a region, the algorithm should be able to throw away most of non face skin regions. So if the height to width ratio falls within the range of well known golden ratio tolerance and if the percentage of skin is higher than a threshold called percentage threshold, then that region is considered a face region. The algorithm works with faces of all sizes and does not assume anything about the scale at which a face appears.

3) Proposed Algorithm

Input: JPEG Image containing Face image.

Step 1: Convert the input RGB image(rgb(i,j)) into HSV image(hsv(i,j)).

Step 2: Get the edge map image (edge(i,j)) from RGB image using Sobel operator.

Step 3: For each pixel (i,j), get the corresponding H,S values.

Step 4: If (colorhistogram(H,S) ϵ Skin Range) and edge(i,j) < edge threshold)
then skin(i,j)=1 i.e. (i,j) is a skin pixel.
else skin(i,j)=0 i.e. (i,j) is a non-skin pixel.
where Skin Range are threshold values

Step 5: Find the difference regions in the image by implementing connectivity analysis using 8-connected neighborhood.

Step 6: Find height and width for each region and percentage of skin in each region.

Step 7: For each region, if (height/width) or (width/height) is within the range and (percentage of skin > threshold value)
then the region is a face,
else it is not a face.

Output: Rectangular cropped face patch.

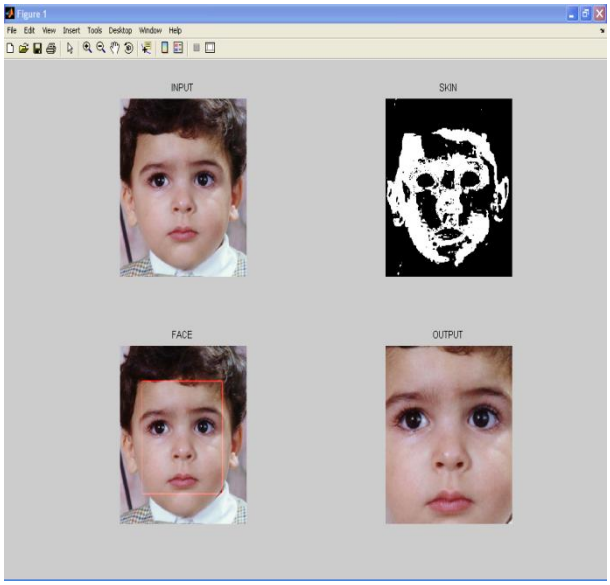


Fig. 2 Face Detection

B. Face Normalization

This method usually increases the global contrast of many images, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to gain a higher contrast without affecting the global contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values.

1) Proposed Algorithm

Input: Rectangular cropped face patch.

Step 1: Convert the facial detected image into a uniform size.

Step 2: Apply illumination normalization to the resized image through Histogram equalization.

$cdf(v) = round((cdf(v) - cdf_{min})*$
$(L-1)/(M*N) - cdf_{min}))$
L = number of gray levels used.
M=width.
N=height.

Output: Normalized face image.

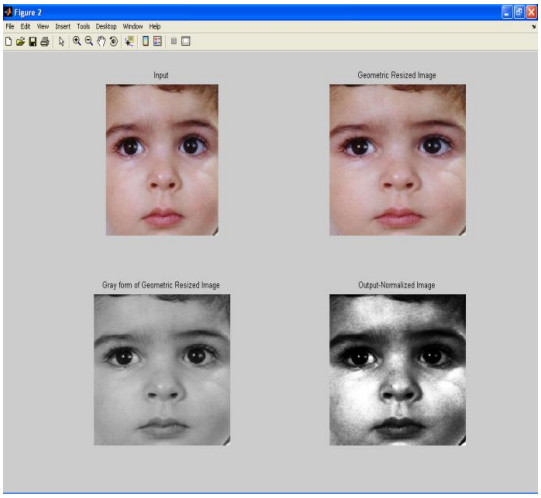


Fig. 3 Face Normalization

C. Manifold Learning

Age is one of the basic attributes in facial images. The objective of manifold embedding is to find a matrix satisfying or directly find n ×d matrix P satisfying $Y=P^TX$, or directly find Y, where Y= {y1, y2,…., yn},X={x1,x2,…..,xn}and P={p1,p2,…. .pn} In a supervised manner, manifold embedding constrains to search nearest neighbors.

Some typical dimensionality reduction and manifold embedding methods are as follows: Principal Component Analysis (PCA), Locally Linear Embedding (LLE) [8] and Orthogonal Locality Preserving Projections (OLPP) [6].

The Locality Preserving Projection (LPP) algorithm, which aims at finding a linear approximation to the eigen functions of the Laplace Beltrami operator on the face manifold. However, LPP is non-orthogonal, and this makes it difficult to reconstruct the data. The orthogonal locality preserving projection (OLPP) method produces orthogonal basis functions and can have more locality preserving power than LPP. The LPP searches the embedding that preserves essential manifold structure by measuring the local neighborhood distance information. The OLPP is expected to have more discriminating power than LPP.

1) Proposed Algorithm

Input: Normalized face image

Step 1: Find the feature points using hough transform
Step 2: Apply OLPP to measure the neighbourhood distance information by graycomatrix function
Step3: Study the = {args various measures and determine a suitable
$P\ (min\sum_{i=}{}^{n}{}_{1}\ \sum_{j=1}^{n}(P^Tx_i-P^Tx_j)^2 * S_{i,j})\}$
P=optimal Laplace Projection.

Output: Feature Pattern for Regression.

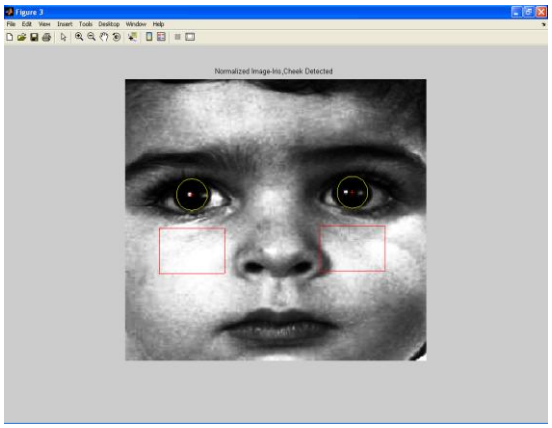


Fig. 4 Manifold Learning

D. Curve Fitting

Curve fitting is finding a curve which has the best fit to a series of data points and possibly other constraints. This section is an introduction to both interpolation (where an exact fit to constraints is expected) and regression analysis. Both are sometimes used for extrapolation. Regression analysis allows for an approximate fit by minimizing the difference between the data points and the curve. Robust regression is also efficiently done using curve fitting.

Regression analysis is the statistical term for curve fitting. We produce a curve that best fits some observed data points. Using regression, we can make predictions as to the behavior of some property in the future. Curve fitting can be performed for any degree, and matlab offers two simple functions for this purpose.

Using function ***polyfit,*** we pass parameters for the range of data(x), the actual values(y), and the degree of the polynomial to which the data is to be fit. Polyfit returns a vector with (DEGREE + 1) elements, corresponding to the polynomial coefficients, starting with the highest degree.

Using ***polyval,*** the vector of coefficients can be evaluated for any data range of X. polyval takes as input parameters a vector of coefficients, and the data range over which a corresponding Y vector is to be performed.

1) Proposed Algorithm

Input: Feature Pattern

Step 1: Find the appropriate polynomial that fits the feature pattern and age details.
Step 2: Map the polynomial into regressive curve.

Regression function:

Age Label L=f(y), where y=manifold data.

The training phase includes all the above processes. During test, an input face image undergoes face detection, Normalization etc except the feature pattern generation (Polynomial fitting).

Output: Appropriate age for the subject in the image is determined.

V. IMPLEMENTATION METHODOLOGY

Our proposed system gets a jpeg image as input. The rectangular facial region is detected from the input image.

First color model threshold analysis is used to find whether a pixel is a skin pixel or not. Then connectivity analysis is applied to find the facial region which is a group of skin pixels that forms the face. Now draw a rectangle over the facial region and it is cropped separately into a variable.

Now the cropped image is resized into uniform assumed size. Then Apply histogram equalization to perform illumination normalization. Now get a normalized face image in which we will perform the Manifold Learning. Next do the ground work for manifold learning i.e. feature point detection by applying Hough transform on the normalized image. Perform Manifold learning (OLPP) by measuring the local neighborhood distance information such as contrast, correlation, homogeneity, entropy etc., These values can be calculated from the gray comatrix of the region(cheek, forehead) chosen for age estimation. Thus our focus shifted from a higher dimension data to lower dimension data. Now find the trend between the features determined and age labels using Regression (curve fitting). This can be done by the curve fitting toolbox present in matlab.

A large number of training images are collected from a broad range of subject ages. The cropped face patches undergo a normalization including illumination normalization (basically histogram equalization). Then the age manifold is learned to map the original face image data into a low-dimensional subspace. A regression function is applied to fit the manifold data.

For test, an input face picture goes through the same process of face detection and normalization. Then the normalized face image is projected on the learned manifold which was computed in the learning stage. Finally, the discrete classification of the age of the input face image is done and an approximate age is found out.

VI. CONCLUSION

The proposed system is tested with FGNET database images which contain only a single face. This is done only in order to preserve the local neighborhood information that helps in determining the feature that varies with ages. The current system first discretely classifies the subject in the image and then estimates the approximate age of the person close to the age mentioned in the database and suggests a range within which the age of the person might be.

REFERENCES

[1] Guodong Guo, Yun Fu, Charles R. Dyer and Thomas S. Huang, "Image based human age estimation by manifold learning and locally adjusted robust regression", IEEE Trans. Image processing, vol. 17,No. 7, July 2008.
[2] The FG-NET Aging Database Available: http://www.fgnet.rsunit.com/
[3] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," IEEE Trans. Multimedia, to be published..
[4] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in Proc. ACMConf. Multimedia, 2006, pp. 307–316.
[5] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human ages by manifold analysis of face pictures and regression on aging features," in Proc. IEEE Conf. Multimedia Expo., 2007, pp. 1383–1386
[6] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp.3608–3614, Nov. 2006.
[7] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proc. IEEE Conf. CVPR*, 2006, pp. 387–394.
[8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
[9] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 34, no. 1, pp. 621–628, Feb. 2004.
[10] Sanjay Kr. Singh, D.S. Chauhan, Mayank Vatsa, Richa Singh, "A Robust Skin Color Based Face Detection Algorithm," Tamkang Journal of Science and Engineering, Vol.6, No. 4, pp. 227-234 (2003).
[11] K. Sandeep and A.N. Rajagopalan, "Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information".

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue2, pp-495-499 ISSN: 2249-6645

Sender and Receiver Based Efficient Broadcasting Algorithm in Mobile Ad Hoc Networks

[1]P.Visu, [2]E.Kannan,[3]S.Koteeswaran and [4]M.Saleembabu

[1,3]Research Scholar, Dept of CSE, Vel Tech Dr.RR & Dr.SR Technical University,Avadi, Chennai-62.
[2]Dean, Dept of CSE, Vel Tech Dr.RR & Dr.SR Technical University,Avadi, Chennai-62.
[4]Professor, Dept of CSE, Vel Tech Dr.RR & Dr.SR Technical University,Avadi, Chennai-62.

Abstract - **This paper develops novel broadcast algorithms for mobile ad hoc networks to improve the efficiency and provide guarantee for full delivery of broadcasting. An important objective is to reduce broadcast redundancy and to avoid the broadcast storm problem. It proposes two broadcasting algorithms such as, Sender based algorithm and Receiver based algorithm. The proposed Sender based algorithm selects subset of forwarding nodes using 1-hop neighbor information. It can reduce both the computational complexity of selecting the forwarding nodes and the maximum number of selected nodes in the worst case. The proposed receiver based broadcasting algorithm can significantly reduce redundant broadcasts in the network. It decides whether or not to broadcast the message. The probability of two neighbor nodes broadcasting the same message exponentially decreases when the distance between them decreases or when the node density increases. The receiver based algorithm uses the Responsibility based scheme which further reduces the redundancy. The proposed algorithms improve the efficiency and also guarantee full delivery.**

Key Words - **Wireless ad hoc, broadcasting, flooding, B-coverage set, Localized algorithm**

I. INTRODUCTION

In wireless communication systems, there will be a need for the rapid deployment of independent mobile users. Significant examples include establishing survivable, efficient, dynamic communication for emergency/rescue operations, disaster relief efforts, and military networks. Such network scenarios cannot rely on centralized and organized connectivity, and can be conceived as applications of Mobile Ad Hoc Networks (MANET). A Mobile Ad Hoc Network (MANET) is a set of nodes communicating with each other via multi-hop wireless links. Each node can directly communicate with only those nodes that are in its communication range. Intermediate nodes forward messages to the nodes that are more than one hop distance from the source. Since the nodes are mobile, the topology of the network is constantly changing. The set of applications for MANETs is diverse, ranging from small, static networks that are constrained by power sources, to large-scale, mobile, highly dynamic networks. Broadcasting is the process in which one node sends a packet to all other nodes in the network. Broadcasting is often necessary in MANET routing protocols. For example, many routing protocols such as Dynamic Source Routing (DSR) [1], Ad Hoc on Demand Distance Vector (AODV) [2], and Zone Routing protocol (ZRP) [3] and Location Aided Routing (LAR) [4] use broadcasting to establish routes. The broadcast is spontaneous. Any mobile host can issue a broadcast operation at any time.

In MANET, broadcasting is used in the route discovery process in several routing protocols, when advising an error message to erase invalid routes from the routing table, or as an efficient mechanism for reliable multicast in a fast moving MANET. In MANETs with the promiscuous receiving node, the traditional blind flooding incurs significant redundancy, collision, and contention, which is known as the broadcast storm problem [5]. Efficient broadcasting in a MANET focuses on selecting a small forward node set while ensuring broadcast coverage. Ad hoc wireless networks are dynamic in nature. Due to this dynamic nature, global information/infrastructure such as minimal spanning tree is no longer suitable to support broadcasting in ad hoc networks.

Broadcasting means one node sends a packet to all other nodes in a network. Efficient broadcasting in a mobile ad hoc network focuses on selecting a small forward node set while ensuring broadcast coverage. The objective is to determine a small set of forward nodes to ensure full coverage. A formal framework is used to model inaccurate local views in MANETs, where full coverage is guaranteed if three sufficient conditions connectivity, link availability, and consistency are met. A MANET consists of a set of mobile hosts that may communicate with one another from time to time. No base stations are supported. Each host is equipped with a CSMA/CA (carrier sense multiple access with collision avoidance) transceiver. In such environment, a host may communicate with another directly or indirectly. In the latter case, a multi hop scenario occurs, where the packets originated from the source host are relayed by several intermediate hosts before reaching the destination. The

broadcast problem refers to the sending of a message to other hosts in the network. The problem considered here has the following characteristics. In a broadcast process, each node decides its forwarding status based on given neighborhood information and the corresponding broadcast protocol. Most existing broadcast schemes assume either the underlying network topology is static during the broadcast process such that the neighborhood information can be updated in a timely manner. The results in show that existing static network broadcast schemes perform poorly in terms of delivery ratio when nodes are mobile. There are two sources that cause the failure of message delivery.

- *Collision:* The message intended for a destination collides with another message.
- *Mobility nodes:* A former neighbor moves out of the transmission range of the current node (i.e., it is no longer a neighbor).

Mobile Ad hoc Network (MANET) consist of a collection of mobile hosts without a fixed infrastructure. Due to limited wireless power a host may not communicate with its destination directly. It usually requires other hosts to forward its packets to the destination through several hops. So in MANET every host acts as a router when it is forwarding packets for other hosts. Because of mobility of hosts and time variability of the wireless medium, the topology of MANET varies frequently. Therefore the routing protocol plays an important role in MANET. There has been extensive research on routing protocols, such as DSR, AODV, LAR and ZRP. A common feature of these routing protocols is that their route discovery all relies on network wide broadcasting to find the destination. Recently, a number of research groups have proposed more efficient broadcasting techniques whose goal is to minimize the number of retransmissions while attempting to ensure that a broadcast packet is delivered to each node in the network.

A MANET consists of a set of mobile hosts that may communicate with one another from time to time. No base stations are supported. Each host is equipped with a CSMA/CA (*carrier sense multiple access with collision avoidance*) transceiver. In such environment, a host may communicate with another directly or indirectly. In the latter case, a multihop scenario occurs, where the packets originating from the source host are relayed by several intermediate hosts before reaching the destination. The *broadcast problem* refers to the sending of a message to other hosts in the network. The problem considered here is assumed to have the following characteristics.

- ***The broadcast is spontaneous:*** Any mobile host can issue a broadcast message at any time. For reasons such as host mobility and lack of synchronization, preparing any kind of global topology knowledge is prohibitive (in fact this is at least as hard as the broadcast problem). Little or even no local connectivity information may be collected in advance.
- *The broadcast is unreliable:* No acknowledgement mechanism will be used. However, an attempt should be made to distribute a broadcast message to as many hosts as possible without paying too much effort. The motivations to make such an assumption are

i. A host may miss a broadcast message because it is off-line, it is temporarily isolated from the network, or it experiences repetitive collisions.

ii. Acknowledgements may cause serious medium contention (and thus, another "storm") surrounding the sender, and receiver.

iii. In many applications (e.g., route discovery), a 100% reliable broadcast is unnecessary. In addition, we assume that a host can detect duplicate broadcast messages. This is essential to prevent endless flooding of a message. One way to do so is to associate with each broadcast message a tuple (source ID, sequence number) as that in [1, 2].

II. RELATED WORKS

Existing broadcasting methods in mobile ad hoc networks are single source broadcasting algorithms in which only one source node can send the broadcast message to all the nodes in the network. Existing broadcasting algorithms are classified into following types.

Simple flooding [8, 9]: requires each node in a MANET to rebroadcast all packets.

Probability based [10]: assigns probabilities to each node to rebroadcast depending on the topology of the network.

Area based [10]: common transmission distance is assumed and a node will rebroadcast if there is sufficient coverage area.

Neighborhood based [11–15]: State on the neighborhood is maintained by neighborhood method, and the information obtained from the neighboring nodes is used for rebroadcast.

A. Simple Flooding Method

In this method, a source node of a MANET disseminates a message to all its neighbors, each of these neighbors will check if they have seen this message before, if yes the message will be dropped, if not the message will redisseminated at once to all their neighbors. The process goes on until all nodes have the message. Although this method is very reliable for a MANET with low density nodes and high mobility but it is very harmful and unproductive as it causes severe network congestion and quickly exhaust the battery power. Blind flooding ensures the coverage; the broadcast packet is guaranteed to be received by every node in the network, providing there is no packet loss caused by collision in the MAC layer and there is no high-speed movement of nodes during the broadcast process. However, due to the broadcast nature of wireless communication media, redundant transmissions in bound flooding may cause the broadcast storm problem, in which redundant packets cause contention and collision.

B. Probability Based Approach

1. Probabilistic scheme: The Probabilistic scheme from [10] is similar to Flooding, except that nodes only rebroadcast with a predetermined probability. In dense networks multiple nodes share similar transmission coverage. Thus, randomly

having some nodes not rebroadcast saves node and network resources without harming delivery effectiveness. In sparse networks, there is much less shared coverage; thus, nodes won't receive all the broadcast packets with the Probabilistic scheme unless the probability parameter is high. When the probability is 100%, this scheme is identical to Flooding.

2. ***Counter-Based scheme:*** Ni et al [10] show an inverse relationship between the number of times a packet is received at a node and the probability of that node being able to reach additional area on a rebroadcast. This result is the basis of their Counter-Based scheme. Upon reception of a previously unseen packet, the node initiates a counter with a value of one and sets a RDT (which is randomly chosen between 0 and Tmax seconds). During the RDT, the counter is incremented by one for each redundant packet received. If the counter is less than a threshold value when the RDT expires, the packet is rebroadcast. Otherwise, it is simply dropped. From [10], threshold values above six relate to little additional coverage area being reached.

C. *Area Based Methods*

Suppose a node receives a packet from a sender that is located only one meter away. If the receiving node rebroadcasts, the additional area covered by the retransmission is quite low. On the other extreme, if a node is located at the boundary of the sender node's transmission distance, then a rebroadcast would reach significant additional area, 61% to be precise [10]. A node using an Area Based Method can evaluate additional coverage area based on all received redundant transmissions. We note that area based methods only consider the coverage area of a transmission; they don't consider whether nodes exist within that area.

1. ***Distance-Based scheme:*** A node using the Distance-Based Scheme compares the distance between itself and each neighbor node that has previously rebroadcast a given packet1. Upon reception of a previously unseen packet, a RDT is initiated and redundant packets are cached. When the RDT expires, all source node locations are examined to see if any node is closer than a threshold distance value. If true, the node doesn't rebroadcast.

2. ***Location-Based scheme:*** The Location-Based scheme [10] uses a more precise estimation of expected additional coverage area in the decision to rebroadcast. In this method, each node must have the means to determine its own location, e.g., a Global Positioning System (GPS). Whenever a node originates or rebroadcasts a packet it adds its own location to the header of the packet. When a node initially receives a packet, it notes the location of the sender and calculates the additional coverage area obtainable were it to rebroadcast. If the additional area is less than a threshold value, the node will not rebroadcast, and all future receptions of the same packet will be ignored. Otherwise, the node assigns a RDT before delivery. If the node receives a redundant packet during the RDT, it recalculates the additional coverage area and compares that value to the threshold. The area calculation and threshold comparison occur with all redundant broadcasts received until the packet reaches either its scheduled send time or is dropped.

D. *Neighbor Knowledge method*

In this method each node will have knowledge of its neighbors and maintains neighbors list. A node that receives a broadcast packet compares its neighbor list to the sender's neighbor list. If the receiving node would not reach any additional nodes then it will not re-broadcast. Otherwise the node rebroadcasts the packet. This is called as self-pruning.

III. PROPOSED SYSTEM

This paper proposes two broadcasting algorithms based on 1-hop neighbor information. The two algorithms namely, Sender based algorithm and receiver based algorithm. The first proposed algorithm is a sender-based algorithm.

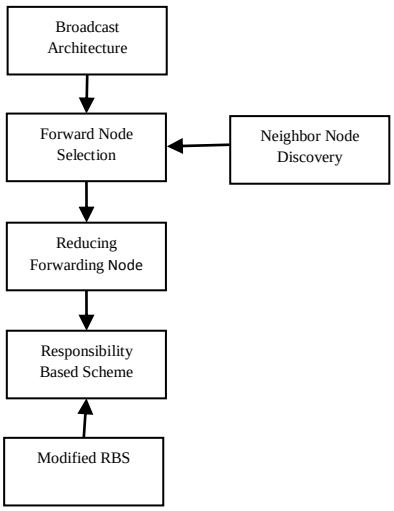


Fig. 1. Broadcasting Process

In sender-based algorithms, the broadcasting nodes select a subset of their neighbors to forward the message and it use 1-hop information. In, Liu et al. propose a broadcasting algorithm that reduces the number of broadcasts and achieves local optimality by selecting the minimum number of forwarding nodes with minimum time complexity O (n log n), where n is the number of neighbors. This optimality only holds for a subclass of sender-based broadcasting algorithms employing 1-hop information and proves that proposed sender-based algorithm can achieve full delivery with time complexity O (n). Moreover, Liu et al.'s algorithm selects n forwarding nodes in the worst case, while this proposed algorithm selects 11 nodes in the worst case. The sender-based algorithm results in fewer broadcasts than does Liu et al.'s algorithm. All these interesting properties are achieved at the cost of a slight increase in end-to-end delay. Thus, the first proposed algorithm is preferred to Liu et al.'s algorithm when the value of n is typically large, and it is important to bind the packet size.

In receiver-based algorithms, the receiver decides whether or not to broadcast the message. The proposed receiver-based algorithm is a novel broadcasting algorithm that can significantly reduce the number of broadcasts in the network. We show that using our proposed receiver based algorithm, two close neighbors are not likely to broadcast the same message. In other words, we prove that the probability of broadcast for a node N_A exponentially decreases when the

distance between N_A and its broadcasting neighbor decreases or when the density of nodes increases. The number of broadcasts using our receiver-based algorithm is less than one of the best known approximations for the minimum number of required broadcasts. It uses Responsibility based scheme to further reduce the redundancy also achieves the efficiency.

IV. IMPLEMENTATION

A. Forwarding Node Selection

In the proposed sender-based algorithm each sender selects a subset of nodes to forward the message. The subset of neighbor is called B-Coverage set. A node can have several B-Coverage set. A forwarding node selection algorithm is called slice based algorithm. Slice-based selection algorithm would be one that selects all of the neighbors as the B-coverage set. Sender-based algorithm can achieve full delivery if it uses any slice-based algorithm to select the forwarding nodes. An efficient slice-based algorithm that selects 11 nodes in the worst case and has computational complexity O (n), where n is the number of neighbors.

B. Reducing forwarding nodes

Each broadcasting node attaches a list of its selected forwarding nodes to the message before broadcasting it. This procedure will increase the band width and power required to broadcast the message. The proposed slice-based selection algorithm reduces the number of selected forwarding nodes to 11 in the worst case. It further reduces the number of selected nodes. Then slice-based algorithm selects a subset of neighbors such that there is at least one selected node in any nonempty bulged slice around A. Node N_A extracts the list of the forwarding nodes from each message which it receives.

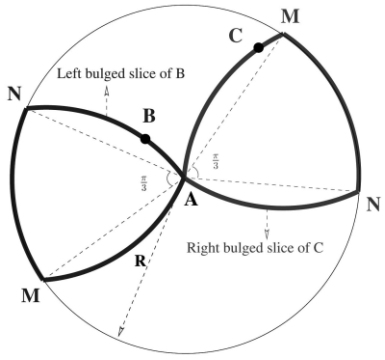


Fig. 2. Bulged Slices

Consider subset of N_A's neighbors that has broadcast the message or been selected by other nodes to forward it. Since all of the selected forwarding nodes are required to broadcast the message, it is sufficient for N_A to find a subset of its neighbors such that any nonempty bulged slice around A contains at least one node from bulged slice. It can be simply extended to achieve this in O (n). Note that the extended algorithm can start with a node N_A and select any node in bulged slice A as soon as it appears in the left bulged slice of the previously selected node. Finally, the extended algorithm removes all of the nodes in bulged slice from the set of selected nodes.

C. Maximizing the minimum node weight of B-Coverage set

Let node N_A assigns a weight to each of its neighbors. The weight can represent the neighbor's battery lifetime, its distance to N_A, the average delay of the node, the level of trust, or a combination of them. In some scenarios, find a B-coverage set such that its minimum node weight is the maximum or its maximum node weight is the minimum among that of all B-coverage sets. For example, assume that the weight of each node represents its battery lifetime in a wireless network. It may be desirable to select the nodes with a higher battery lifetime to forward the message in order to keep the nodes with a lower battery lifetime alive. The slice based algorithm shows how to find a B-coverage set such that its minimum node weight is the maximum among that of all B-coverage sets. The main design challenge is to determine whether or not to broadcast a received message. Although this algorithm is simple to implement, it has limited effect in reducing the number of redundant broadcasts.

D. Responsibility based scheme

The main idea of receiver-based algorithm is that a node avoids broadcasting if it is not responsible for any of its neighbors. A node N_A is not responsible for a neighbor N_B if N_B has received the message or if there is another neighbor N_C such that N_C has received the message and N_B is closer to N_C than it is to N_A. Suppose N_A stores IDs of all its neighbors that have broadcast the message during the defer period. When executed by a node N_A, first uses this information to determine which neighbors have not received the message.

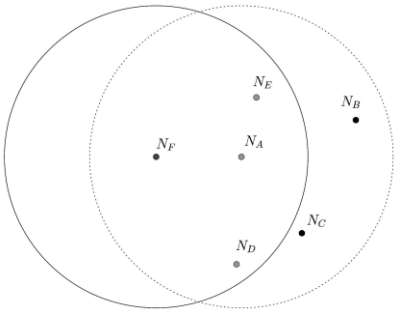


Fig. 3. Example of RBS decision

It then returns false if and only if it finds a neighbor N_B that has not received the message. The output of RBS determines whether or not the broadcast is redundant. So the redundancy is reduced. The modified version of RBS uses the position and transmission range of the broadcasting nodes to determine which neighbors have not received the message.

V. RESULTS

The broadcasting process starts from node 44. The mobility of all nodes is traced. This neighbor information is used to select the optimized forwarding nodes. Slice based algorithm is used to select the forwarding nodes. It reduces the redundancy.

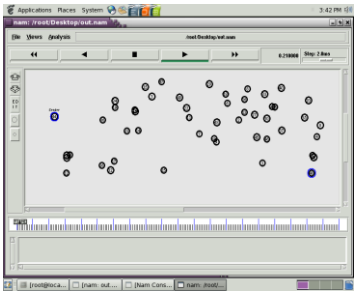


Fig. 4. Initial Node Configuration

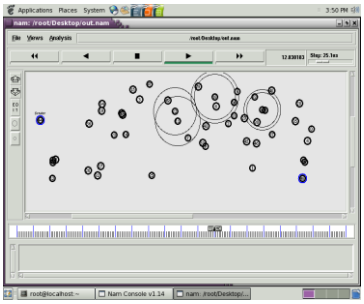


Fig. 5. Broadcasting process

The proposed broadcasting algorithm reduces the redundancy by selecting the minimum number of forwarding nodes and improves the broadcast efficiency.

VI. CONCLUSION

Wireless mobile ad hoc networks present difficult challenges to routing protocol designers. Mobility, constrained bandwidth, and limited power cause frequent topology changes. Broadcasting plays a vital role in mobile ad hoc networks. So this work provides efficient broadcasting by reducing the unnecessary retransmissions. The proposed forwarding-node selection algorithm results in fewer broadcasts in the network. The proposed receiver based algorithm significantly reduces the number of forwarding nodes in the network. So this algorithm is efficient, has minimum number of retransmission and is collision free.

REFERENCES

[1] D. Johnson, D. Maltz and Y. Hu. The Dynamic Source Routing Protocol for Mobile Ad hoc Networks.Internet Draft: draft-ietf-manet-dsr-09.txt, 2003.

[2] C. Perkins, E. Beldig-Royer and S. Das. Ad hoc on Demand Distance Vector (AODV) Routing. Request for Comments 3561, July 2003.

[3] Z. Haas and M. Pearlman. The Performance of Query Control Schemes for the Zone Routing Protocol. IEEE/ACM Transactions on Networking, 9(4):427–438, 2001.

[4] Y. Ko and N. Vaidya. Location-aided Routing (LAR) in Mobile Ad hoc Networks. Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM), 66-75, 1998.

[5] Y.-C. Tseng, S.-Y. Ni, and E.-Y. Shih, "Adaptive Approaches to Relieving Broadcast Storms in a Wireless Multihop Mobile Ad Hoc Network," IEEE Trans. Computers, vol. 52, no. 5, pp. 545-557, May2003.

[6] Majid Khabbazian, and Vijay K. Bhargava," *Efficient Broadcasting in Mobile Ad Hoc Networks*" IEEE Trans. on mobile computing, vol. 8, no. 2, february 2009.

[7] http://www.isi.edu/nsnam/ns

[8] C. Ho, K. Obraczka, G. Tsudik and K. Viswanath. Flooding for Reliable Multicast in Multi-hop Ad hoc Networks. International Workshop in Discrete Algorithms and Methods for Mobile Computing and Communication, 64–71, 1999.

[9] J. Jetcheva, Y. Hu, D. Maltz and D. Johnson. A Simple Protocol for Multicast and Broadcast in Mobile Ad hoc Networks. Internet Draft, draft- ietf-manet-simple-mbcast-01.txt, 2001

[10] S. Ni, Y. Tseng, Y. Chen and J. Sheu. The Broadcast Storm Problem in a Mobile Ad hoc Network. International Workshop on Mobile Computing and Networks, 151–162, 1999

[11] H. Lim and C. Kim. Multicast Tree Construction and Flooding in Wireless Ad hoc Networks. In Proceedings of the ACM International Work- shop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM), 2000

[12] W. Peng and X. Lu. Efficient Broadcast in Mobile Ad hoc Networks using Connected Dominating Sets. Journal of Software, 1999.

[13] W. Peng and X. Lu. On the Reduction of Broadcast Redundancy in Mobile Ad hoc Networks. In Proceedings of MOBIHOC, 2000.

[14] W. Peng and X. Lu. AHBP: An Efficient Broadcast Protocol for Mobile Ad hoc Networks. Journal of Science and Technology, 2002

[15] J. Sucec and I. Marsic. An Efficient Distributed Network-wide Broadcast Algorithm for Mobile Ad hoc Networks. CAIP Technical Report 248 - Rutgers University, September 2000.

www.ijmer.com Vol.1, Issue.2, pp-500-509 ISSN: 2249-6645

Baseband Analysis of Long Term Evolution Systems

Jasvinder Singh Sadana[1] , Neelima Selam[2]

Abstract- *This paper focuses on the universal mobile telecommunications system long term evolution LTE[1 to16] and discusses the requirements for device technologies pertaining to mobile terminals. The LTE[1 to16] represents the next generation cellular phone technology that is intended to achieve a high peak data rate, low latency, and high radio efficiency in addition to low cost and sufficiently high mobility characteristics. Vigorous discussion regarding the specifications for LTE[1 to16] is currently ongoing in the 3rd generation partnership project. This paper also introduces various device technologies that support current mobile terminals besides emphasizes has been given to OFDM[1 to16] with 64 QAM[1 to16] technique while down linking has been done.*

I. INTRODUCTION

What is LTE

Long Term Evolution LTE [1 to 16] describes standardization work by the Third Generation Partnership Project (3GPP) to define a new high-speed radio access method for mobile communications systems.

LTE [1 to 16] is the next step on a clearly-charted roadmap to so-called '4G' mobile systems that starts with today's 2G and 3G networks. Building on the technical foundations of the 3GPP family of cellular systems that embraces GSM, GPRS and EDGE as well as WCDMA and now HSPA (High Speed Packet Access), LTE [1 to 16] offers a smooth evolutionary path to higher speeds and lower latency. Coupled with more efficient use of operators' finite spectrum assets, LTE [1 to 16] enables an even richer, more compelling mobile service environment.

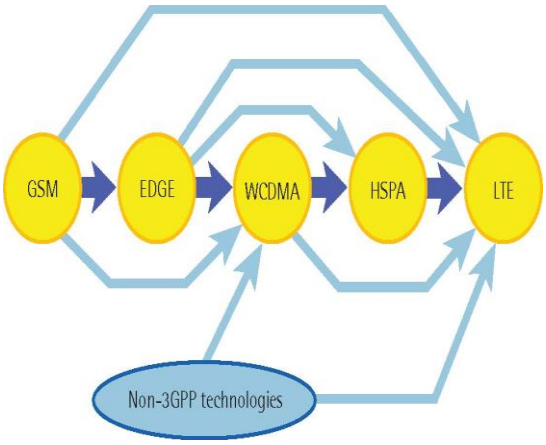


Figure 1

Choice of Upgrade Paths[1]

In parallel with its advanced new radio interface, realising the full potential of LTE [1 to 16] requires an evolution from today's hybrid packet/circuit switched networks to a simplified, all-IP (Internet Protocol) environment. From an operator's point of view, the pay-off is reduced delivery costs for rich, blended applications combining voice, video and data services plus simplified interworking with other fixed and wireless networks.

By creating new value-added service possibilities, LTE [1 to 16] promises long-term revenue stability and growth for around two hundred mobile operators that are already firmly committed to the UMTS/HSPA family of 3G systems. Just as importantly, it provides a powerful tool to attract customers who are provided with an increasing number of technology options for broadband connectivity on the move.

Based on the UMTS/HSPA family of standards, LTE [1 to 16] will enhance the capabilities of current cellular network technologies to satisfy the needs of a highly demanding customer accustomed to fixed broadband services. As such, it unifies the voice-oriented environment of today's mobile networks with the data-centric service possibilities of the fixed Internet.

Another key goal of the project is the harmonious coexistence of LTE [1 to 16] systems alongside legacy circuit switched networks. This will allow operators to introduce LTE's all-IP concept progressively, retaining the value of their existing voice-based service platforms while benefiting from the performance boost that LTE [1 to 16] delivers for data services.

From an operator's perspective, the flexible channel bandwidths and harmonised FDD/TDD modes of LTE [1 to 16] provide a more efficient use of carriers' existing and future spectrum resources. LTE [1 to 16] also provides a more robust platform for operators to offer compelling value-added services in the mobile domain.

From a technical point of view, a fundamental objective of the LTE [1 to 16] project is to offer higher data speeds, for both down- and uplink transmission. Apart from this increase in raw data rates, LTE [1 to 16] is characterised by reduced packet latency; the restriction that determines the responsiveness of gaming, VoIP, videoconferencing and other real-time services.

The key characteristics of LTE [1 to 16] are summarised here, with specific comparison with today's UMTS/HSPA networks:

www.ijmer.com Vol.1, Issue.2, pp-500-509 ISSN: 2249-6645

- Enhanced air interface allows increased data rates: LTE [1 to 16] is built on an all-new radio access network based on OFDM[1 to 16] (Orthogonal Frequency-Division Multiplexing) technology. Specified in 3GPP Release 8, the air interface for LTE [1 to 16] combines OFDM[1 to 16] based modulation and multiple access scheme for the downlink, together with SC-FDMA (Single Carrier FDMA) for the uplink.

- All OFDM[1 to 16] schemes split available spectrum into thousands of extremely narrowband carriers, each carrying a part of the signal. In LTE [1 to 16], the innate spectral efficiency of OFDM[1 to 16] is further enhanced with higher order modulation schemes such as 64QAM [1 to 16] and sophisticated FEC (Forward Error Correction) schemes such as tail biting, convolutional coding and turbo coding, alongside complementary radio techniques like MIMO and Beam Forming with up to four antennas per station.

- The result of these radio interface features is significantly improved radio performance, yielding up to five times the average throughput of HSPA. Downlink peak data rates are extended up to a theoretical maximum of 300 Mbit/s per 20 MHz of spectrum. Similarly, LTE [1 to 16] theoretical uplink rates can reach 75 Mbit/s per 20 MHz of spectrum, with theoretical support for at least 200 active users per cell in 5 MHz.

- As explained in the following paragraphs, the performance of HSPA is itself evolving through the use of technologies like 64QAM [1 to 16] and MIMO. These features are part of 3GPP Release 7, while a combination of 64QAM [1 to 16] and MIMO for HSDPA (FDD) is specified in Release 8. LTE [1 to 16] however, delivers even greater improvements in overall performance and efficiency through the use of OFDM [1 to 16] technology for the air interface, rather than the WCDMA-based UTRAN common to WCDMA and HSPA systems, and through more complex MIMO and beam forming antenna configurations.

- The capabilities of LTE [1 to 16] will also evolve, with improvements specified in forthcoming Releases allowing LTE [1 to 16] (advanced) to fulfill the requirements of IMT-Advanced, the ITU term for so-called '4G' systems that will be the eventual successors to evolved 3G and 3G+ technologies.

- High spectral efficiency: LTE's greater spectral efficiency allows operators to support increased numbers of customers within their existing and future spectrum allocations, with a reduced cost of delivery per bit.

- Flexible radio planning: LTE [1 to 16] can deliver optimum performance in a cell size of up to 5 km. It is still capable of delivering effective performance in cell sizes of up to 30 km radius, with more limited performance available in cell sizes up to 100 km radius. See Section 4 for more information on spectrum for LTE and deployment flexibility.

- Reduced latency: By reducing round-trip times to 10ms or even less (compared with 40–50ms for HSPA), LTE [1 to 16] delivers a more responsive user experience. This permits interactive, real-time services such as high-quality audio/videoconferencing and multi-player gaming.

- An all-IP environment: One of the most significant features of LTE [1 to 16] is its transition to a 'flat', all-IP based core network with a simplified architecture and open interfaces. Indeed, much of 3GPP's standardisation work targets the conversion of existing core network architecture to an all-IP system. Within 3GPP, this initiative has been referred to as Systems Architecture Evolution (SAE) – now called Evolved Packet Core (EPC). SAE/EPC enables more flexible service provisioning plus simplified interworking with fixed and non-3GPP mobile networks.

- EPC is based on TCP/IP protocols – like the vast majority of today's fixed data networks – thus providing PC-like services including voice, video, rich media and messaging. This migration to an all-packet architecture also enables improved interworking with other fixed and wireless communication networks.

www.ijmer.com Vol.1, Issue.2, pp-500-509 ISSN: 2249-6645

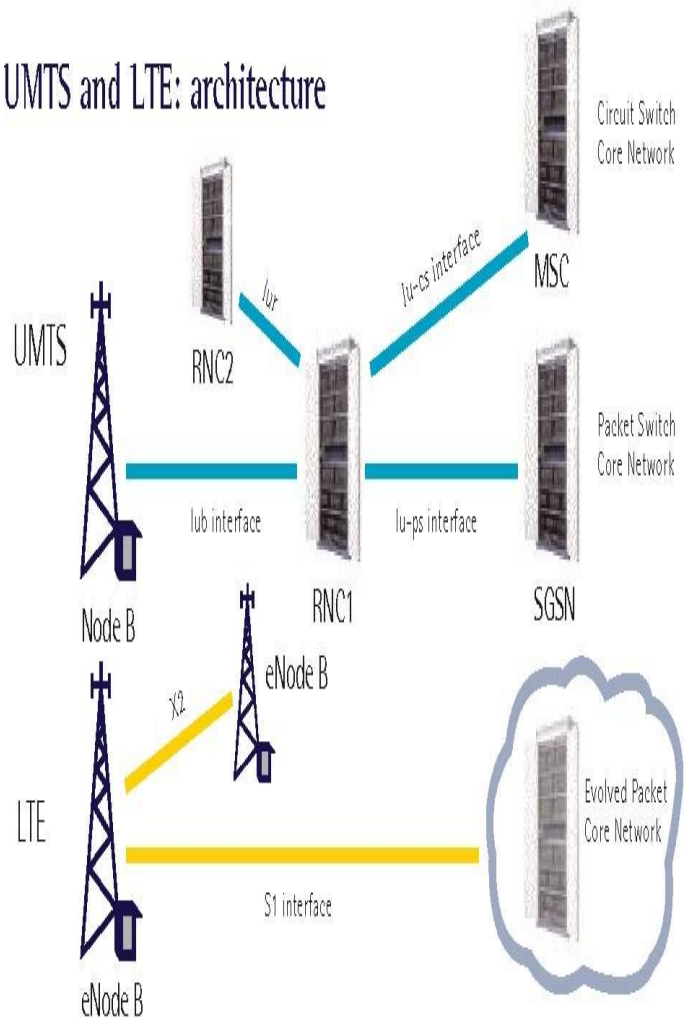


Figure 2

LTE gives operators the benefits of evolution to a simplified, all-IP network architecture.[1]

- Co-existence with legacy standards and systems: LTE [1 to 16] users should be able to make voice calls from their terminal and have access to basic data services even when they are in areas without LTE [1 to 16] coverage.
- LTE [1 to 16] therefore allows smooth, seamless service handover in areas of HSPA, WCDMA or GSM/GPRS/EDGE coverage. Furthermore, LTE/SAE supports not only intra-system and inter-system handovers, but inter-domain handovers between packet switched and circuit switched sessions.
- Extra cost reduction capabilities: The introduction of features such as a multi-vendor RAN (MVR) or self optimising networks (SON) should help to reduce opex and provide the potential to realise lower costs per bit.
- In order to achieve high data rates the first most aspect which should be looked after is the physical layer of the LTE [1 to 16]. This paper focuses on the combination of two efficient modulation and multiplexing schemes used in the LTE [1 to 16].

1. Orthogonal Frequency Division Multiplexing (OFDM)
2. 64- Quadrature Amplitude Modulation (64-QAM)

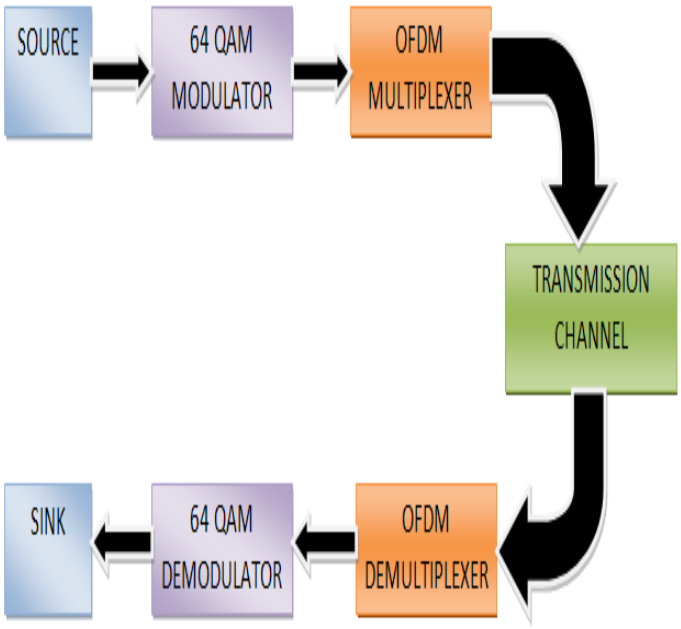


Figure 3

Proposed Baseband Model for Long Term Evolution System

OFDM

Firstly the fundamental difference between modulation and multiplexing shall be discussed.

Modulation - *a mapping of the information on changes in the carrier phase, frequency or amplitude or combination.*

Multiplexing - *method of sharing a bandwidth with other independent data channels.*

- OFDM[1 to 16] is a combination of modulation and multiplexing. Multiplexing generally refers to Independent signals, those produced by different sources. So it is a question of how to share the spectrum with these users. In OFDM[1 to 16] the question of multiplexing is applied to independent

www.ijmer.com Vol.1, Issue.2, pp-500-509 ISSN: 2249-6645

signals but these independent signals are a sub-set of the one main signal. In OFDM[1 to 16] the signal itself is first split into independent channels, modulated by data and then re-multiplexed to create the OFDM[1 to 16] carrier.

- OFDM[1 to 16] is a special case of Frequency Division Multiplex (FDM). As an analogy, a FDM channel is like water flow out of a faucet, in contrast the OFDM[1 to 16] signal is like a shower. In a faucet all water comes in one big stream and cannot be sub-divided. OFDM[1 to 16] shower is made up of a lot of little streams

1) (2)

Figure 4.1 A Regular-FDM single carrier – A whole bunch of water coming all in one stream.[2]

Figure 4.2 Orthogonal-FDM – Same amount of water coming from a lot of small streams.[2]

- Think about what the advantage might be of one over the other? One obvious one is that if I put my thumb over the faucet hole, I can stop the water flow but I cannot do the same for the shower. So although both do the same thing, they respond differently to interference.

- Another way to see this intuitively is to use the analogy of making a shipment via a truck.

- We have two options, one hire a big truck or a bunch of smaller ones. Both methods carry the exact same amount of data. But in case of an accident, only 1/4 of data on the OFDM[1 to 16] trucking will suffer.

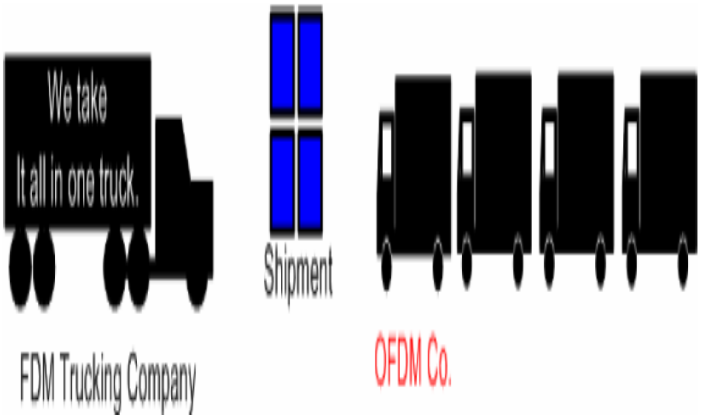


Figure 5

FDM v/s OFDM[2]

- These four smaller trucks when seen as signals are called the sub-carriers in an OFDM[1 to 16] system and they must be orthogonal for this idea to work. The independent sub-channels can be multiplexed by frequency division multiplexing (FDM), called multi-carrier transmission or it can be based on a code division multiplex (CDM), in this case it is called multi-code transmission.

The importance of being orthogonal

The main concept in OFDM[1 to 16] is orthogonality of the sub-carriers. Since the carriers are all sine/cosine wave, we know that area under one period of a sine or a cosine wave is zero. This is easily shown.

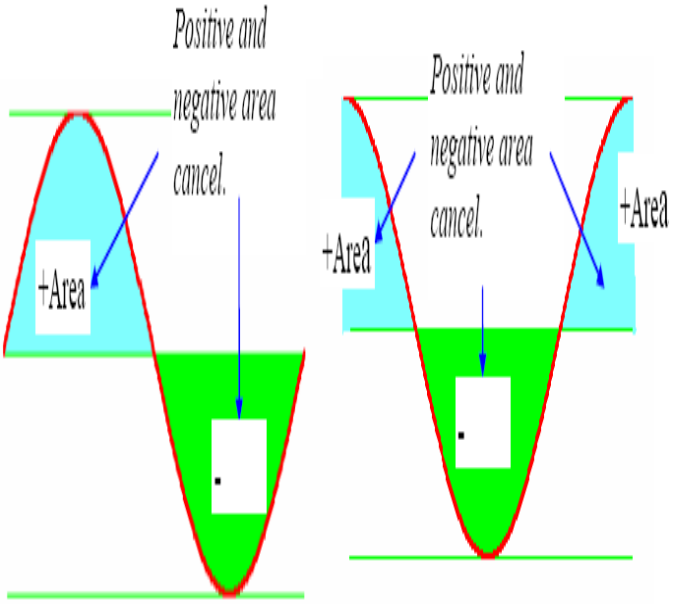


Figure 6

The area under a sine and a cosine wave over one period is always zero. [2]

Let's take a sine wave of frequency m and multiply it by a sinusoid (sine or a cosine) of a frequency *n*, where both *m* and *n* are integers. The integral or the area under this product is given by

$f(t)=\sin mwt \times \sin nwt$

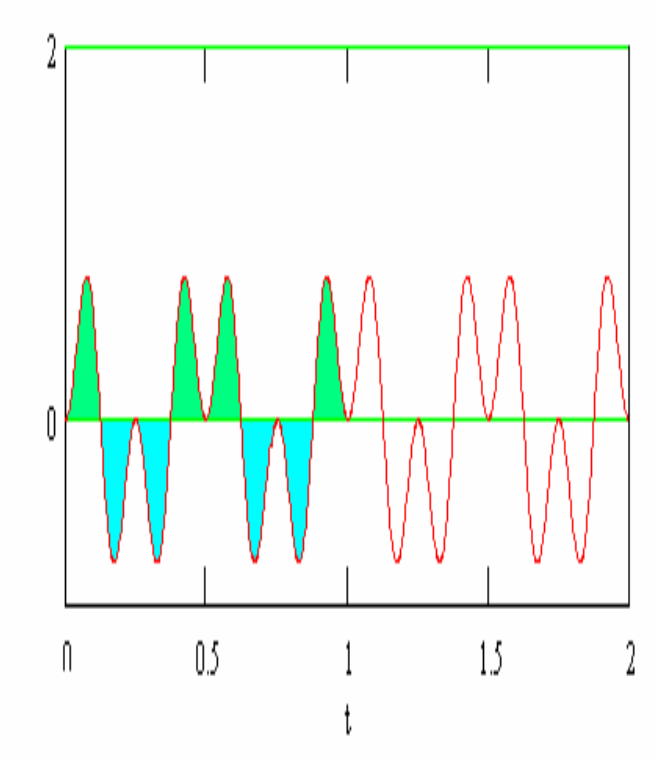


Figure 7

The area under a sine wave multiplied by its own harmonic is always zero. [2]

By the simple trigonometric relationship, this is equal to a sum of two sinusoids of frequencies *(n-m)* and *(n+m)*

$$= \frac{1}{2}\cos(m-n) - \frac{1}{2}\cos(m+n)$$

These two components are each a sinusoid, so the integral is equal to zero over one period.

$$= \int_0^{2\pi} \frac{1}{2}\cos(m-n)\omega t - \int_0^{2\pi} \frac{1}{2}\cos(m+n)\omega t$$
$$= 0 - 0$$

We conclude that when we multiply a sinusoid of frequency *n* by a sinusoid of frequency *m/n*, the area under the product is zero. In general for all integers *n* and *m*, sin*mx*,cos*mx*, cos *nx*,sin *nx* are all orthogonal to each other. These frequencies are called harmonics.

This idea is key to understanding OFDM[1 to 16]. The orthogonality allows simultaneous transmission on a lot of sub-carriers in a tight frequency space without interference from each other. In essence this is similar to CDMA, where codes are used to make data sequences independent (also orthogonal) which allows many independent users to transmit in same space successfully.

Let's first look at what a Frequency Division Multiplexing FDM is. If I have a bandwidth that goes from frequency *a* to *b*, I can subdivide this into a frequency space of four equal spaces. In frequency space the modulated carriers would look like this. OFDM[1 to 16] is a special case of FDM.

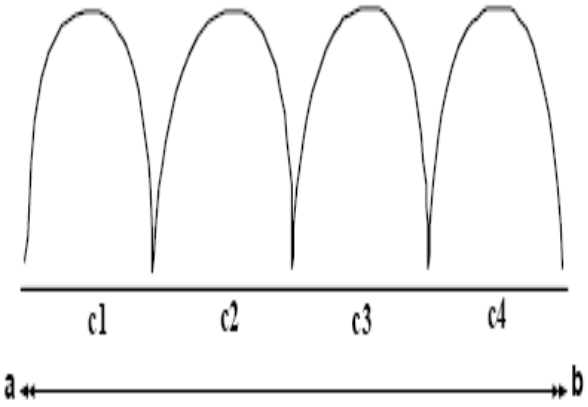


Figure 8

FDM carriers are placed to next to each other. [2]

The frequencies *a* and *b* can be anything, integer or non-integer since no relationship is implied between *a* and *b*. Same is true of the carrier center frequencies which are based on frequencies that do not have any special relationship to each other. But, what if frequency c_1 and c_n were such that for any *n*, an integer, the following holds.

$c_n = n \times c_1$
So that
$c_2 = 2c_1$
$c_3 = 3c_1$
$c_4 = 4c_1$

All three of these frequencies are harmonic to c1. In this case, since these carriers are orthogonal to each other, when added together, they do not interfere with each other. In FDM, since we do not generally have frequencies that follow the above relationship, we get interference from neighbour carriers. To provide adjacent channel interference protection, signals are moved further apart.
The symbols rate that can be carried by a PSK carrier of bandwidth *b*, is given by

$Rs = 2Bl = Bp$

where B_l is lowpass bandwidth and Bp, the passband bandwidth. This relationship assumes a perfect Nyquist

filtering with rolloff = 0.0. Since this is unachievable, we use root raised cosine filtering which for a roll-off of α gives the following relationship.

$$Rs = \frac{Bp}{1+\alpha}$$

So if we need three carriers, each of data rate = 20 Mbps, then we might place our BPSK carriers as shown below. With Rs = 20 and B = 20 x 1.25 = 25 MHz. Each carrier may be placed (25 + 2.5) 27.5 MHz apart allowing for a 10% guard band. The frequencies would not be orthogonal but in FDM we don't care about this. It's the guard band that helps keep interference under control.

The following figure shows the OFDM[1 to 16] signal in time and frequency domains. The subcarriers and guard intervals have been shown. The guard intervals have been placed between two adjacent symbols. Inter Symbol Interference is avoided by means of guard intervals. As a result bandwidth is conserved. Capacity of the channel is increased subsequently.

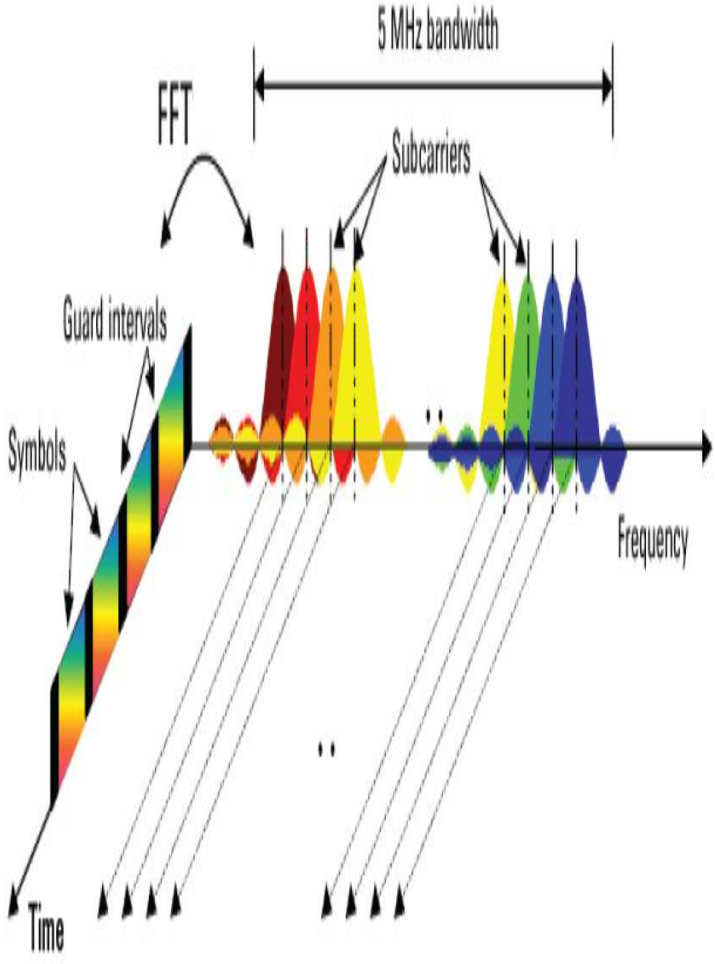


Figure 9

OFDM signal represented in frequency and time[2]

QAM

QAM, Quadrature amplitude modulation is widely used in many digital data radio communications and data communications applications. A variety of forms of QAM[1 to 16] are available and some of the more common forms include 16 QAM, 32 QAM, 64 QAM, 128 QAM, and 256 QAM. Here the figures refer to the number of points on the constellation, i.e. the number of distinct states that can exist.

- The various flavors of QAM[1 to 16] may be used when data-rates beyond those offered by 8-PSK are required by a radio communications system. This is because QAM[1 to 16] achieves a greater distance between adjacent points in the I-Q plane by distributing the points more evenly. And in this way the points on the constellation are more distinct and data errors are reduced. While it is possible to transmit more bits per symbol, if the energy of the constellation is to remain the same, the points on the constellation must be closer together and the transmission becomes more susceptible to noise. This results in a higher bit error rate than for the lower order QAM[1 to 16] variants. In this way there is a balance between obtaining the higher data rates and maintaining an acceptable bit error rate for any radio communications system.

- QAM[1 to 16] is used in many radio communications and data delivery applications. However some specific variants of QAM[1 to 16] are used in some specific applications and standards.

For domestic broadcast applications for example, 64 QAM and 256 QAM are often used in digital cable television and cable modem applications. In the UK, 16 QAM and 64 QAM[1 to 16] are currently used for digital terrestrial television using DVB - Digital Video Broadcasting. In the US, 64 QAM and 256 QAM are the mandated modulation schemes for digital cable as standardised by the SCTE in the standard ANSI/SCTE 07 2000.

In addition to this, variants of QAM[1 to 16] are also used for many wireless and cellular technology applications.

Constellation diagrams for QAM

The constellation diagrams show the different positions for the states within different forms of QAM[1 to 16], quadrature amplitude modulation. As the order of the modulation increases, so does the number of points on the QAM[1 to 16] constellation diagram. Since the number of bits per symbol gets increased so does the packaging efficiency of the QAM increases. The diagrams below show constellation diagrams for a variety of formats of Quadrature Amplitude modulation technique:

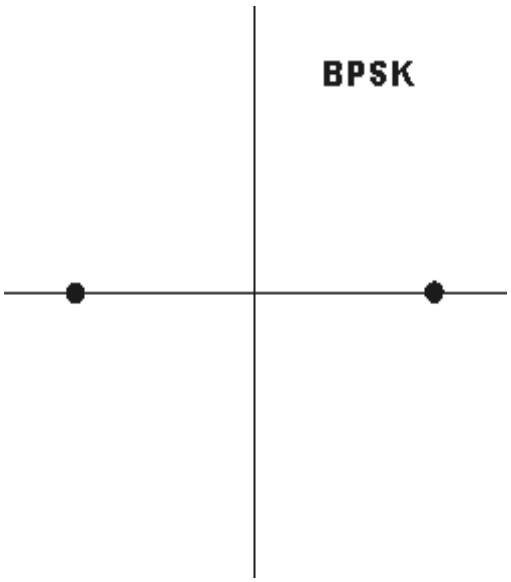


Figure 10

BPSK Constellation Diagram[17]

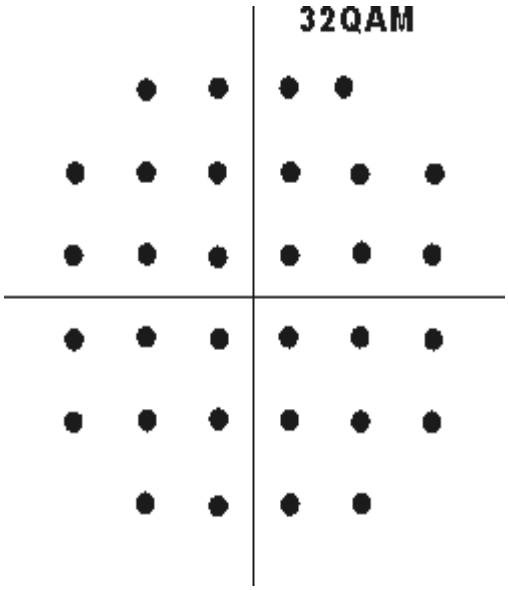


Figure 12

32 QAM Constellation Diagram[17]

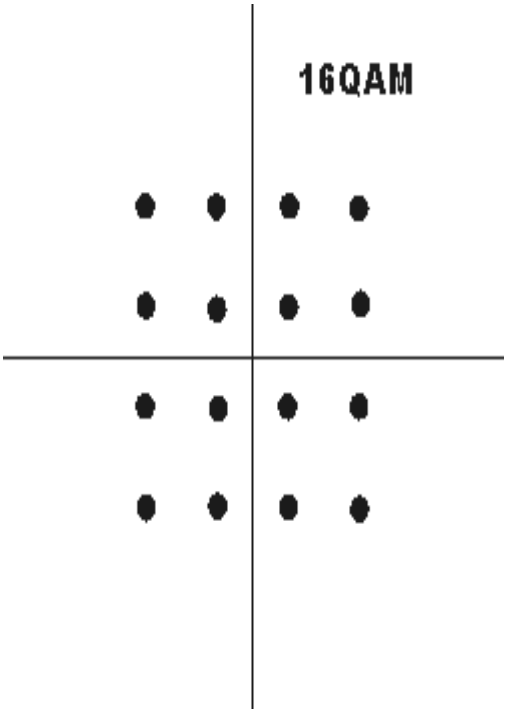


Figure 11

16 QAM Constellation Diagram[17]

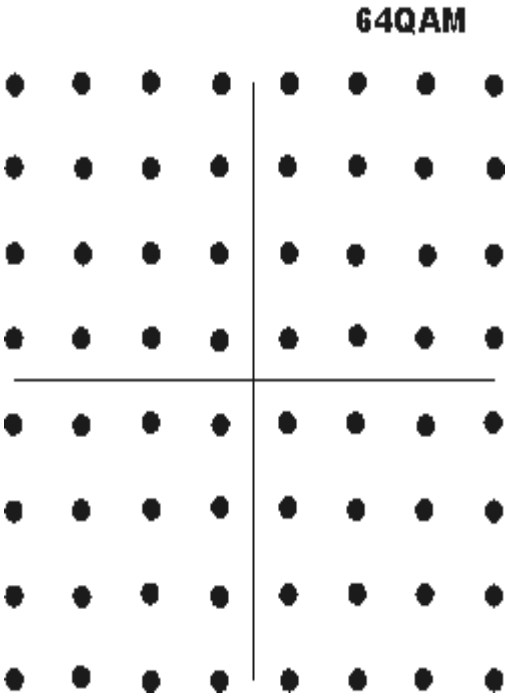


Figure 13

64 QAM Constellation Diagram[17]

www.ijmer.com Vol.1, Issue.2, pp-500-509 ISSN: 2249-6645

QAM bits per symbol

The advantage of using QAM[1 to 16] is that it is a higher order form of modulation and as a result :

1. Bit Packaging Ratio is increased.
2. Bandwidth efficiency is increased.
3. Over all channel capacity is increased.
4. More number of customers can be catered.
5. The trade off between bandwidth and signal to noise ratio can be made i.e. same channel capacity is achievable at a lower signal to noise ratio value because of higher bandwidth.
6. Less transmission power is being required by the transmitter.
7. Power budget of both the transmitter and receiver reduces by many folds thus reducing the over all cost of the system.
8. The data rate of a link can be increased.

The table below gives a summary of the bit rates of different forms of QAM[1 to 16] and PSK.

MODULATION	BITS PER SYMBOL	SYMBOL RATE
BPSK	1	1 x bit rate
QPSK	2	1/2 bit rate
8PSK	3	1/3 bit rate
16QAM	4	1/4 bit rate
32QAM	5	1/5 bit rate
64QAM	6	1/6 bit rate

Table 1

The more the number of bits per symbol increased simultaneously Bit Packaging Ratio increases. One Bit in 64 QAM carries information equivalent to 6 Bits. Thus making it a very efficient Modulation Mechanism

QAM noise margin

While higher order modulation rates are able to offer much faster data rates and higher levels of spectral efficiency for the radio communications system, this comes at a price. The higher order modulation schemes are considerably less resilient to noise and interference.

As a result of this, many radio communications systems now use dynamic adaptive modulation techniques. They sense the channel conditions and adapt the modulation scheme to obtain the highest data rate for the given conditions. As signal to noise ratios decrease errors will increase along with re-sends of the data, thereby slowing throughput. By reverting to a lower order modulation scheme the link can be made more reliable with fewer data errors and re-sends.

The use of 64- QAM[1 to 16] has become more prevalent in cable systems, as both a video and data modulation. With its six bits per digital symbol (6 bits/symbol), it offers the highest bandwidth efficiency available today among digital cable signals. Expectations are that 64- QAM[1 to 16] will evolve to become a dominant modulation format of the digital multiplex. With the value of bandwidth at a premium, particularly based on bandwidth consumption trends of the past, it is of continued interest to find techniques that increase throughput capability. In HFC systems, there are various ways to create more bandwidth, such as increasing fibre counts, implementing equipment segmentation in the plant, adding wavelengths, or improving compression techniques in digital television signals. The migration of 16- QAM[1 to 16], which represents 4 bits/symbol, to 64- QAM[1 to 16] and its 6 bits/symbol, provides 33% more efficient bandwidth usage. Of course, this is at the expense of a higher SNR requirement, as well as increased sensitivity to other impairments, such as phase noise and interference.

Theoretical AWGN Performance

The bit error rate (BER) curves for M-QAM are straightforward to develop in an AWGN-only channel, since symbol-by-symbol, hard-decision decoding is optimal and symbols are uniformly affected. Recognizing that most symbols are bounded on four sides with decision boundaries for large M easily generates upper bounds. However, more accurate solutions are not difficult to develop. Assuming one bit error for every symbol error, a situation that can be assured under what is referred to as Gray encoded mapping of bits to symbols (adjacent symbol differ by one bit only), we can show the following bit error probabilities as a function of SNR, assuming the SNR associated with optimal detection mechanisms.

www.ijmer.com Vol.1, Issue.2, pp-500-509 ISSN: 2249-6645

$\mathbf{P_e(64\text{-}QAM) = (7/12)\ Q\ [(SNR/21)^{1/2}]}$

Recognizing that signal power is related to energy per symbol as $E_s = P_s \cdot T_s$, and that

$E_b = E_s / \log_2 M$, these expression can be written in the form common to digital communication theory, which uses E_b/N_o, as

$\mathbf{P_e(64\text{-}QAM) = (7/12)\ Q\ [(2E_b/7N_o)^{1/2}]}$

In this case, No is the noise power *density*, and the Q( ) function is a well-known, oft-tabulated function associated with the solution to the integral under the upper tail of the Gaussian probability density function (PDF), which is thus accounting for the statistical nature of the AWGN.

II. RESULTS

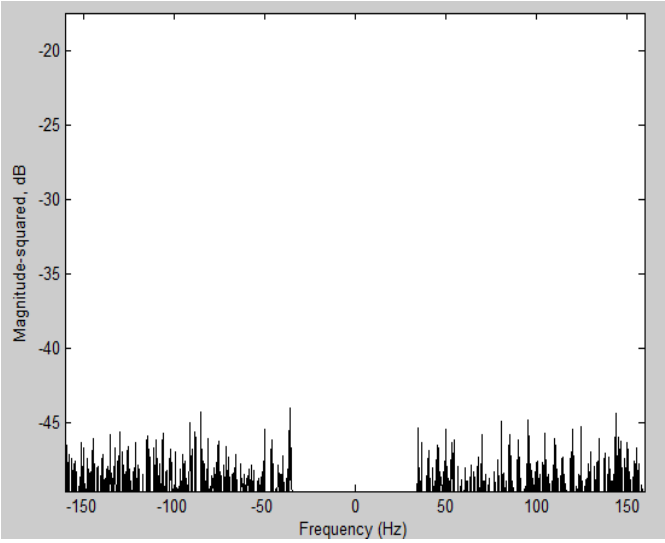


Figure 14
Transmitted Signal

In the above figure the signal generated at the transmitted side is shown

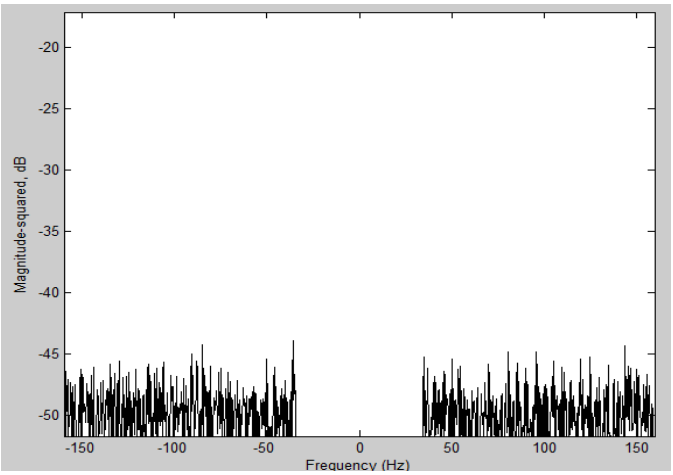


Figure 16
Received Signal

In the above figure the signal received at the receiver side is shown.

- Both the signal have been shown in the frequency domain hence they represent the power spectrum of the signal at both the sides.
- There is a notable difference in both the spectrums incurred due to channel loss.

III. CONCLUSION

Based on the results no Bit Loss and Packet Loss have been recorded. Although simulation environment differs from the real world scenario. Despite Orthogonal Frequency Division Multiplexing and 64 Quadrature Amplitude Modulation Techniques are very powerful transmission mechanisms but there are following few problems which they have to encounter in the real world. The few such problems are:

1. Multipath Fading
2. Space Diversity
3. Path Loss
4. High Bit error rate

Besides we require efficient coding techniques in order to perform Error Detection and Correction.

IV. PRACTICAL SCOPE

The practical scope of this project involves hardware implementation of the block diagram. If the same is implemented very high data rates will be achievable in the mobile environment and implementation of 100Mbps of LAN on mobile phones will become feasible. The higher data rates will support following Value Added Services:

1. **Rich voice**
   a. High Quality Video Conferencing.
   b. VoIP
2. **Photo messaging**
   a. Instant Messaging
   b. Mobile
   c. Email
   d. Video messaging
3. **Browsing**
   a. Super fast browsing.
   b. Uploading content to the social networking sites.
4. **Paid Information**
   a. E-newspapers
   b. High Quality Audio Streaming.
5. **Video on demand**
   a. Broadcast television services.
   b. True on demand television.
   c. High quality video streaming.

6. **Music**
    a. High quality music downloading and storage.
7. **Online gaming**
8. **Content messaging and cross media**
    a. Wide scale distribution of video clips.
    b. Video based mobile advertising.
9. **M-commerce**
    a. Mobile handsets as payment devices, with payment details carried over high speed networks to erasable rapid completion of transaction.
10. **Mobile data networking**
    a. P2P file transfer.
    b. Business Applications
    c. Application sharing
    d. M2M communication
    e. Mobile intranet/extranet

V. REFERENCES

1. A White Paper from the UMTS Forum **Towards Global Mobile Broadband** Standardising the future of mobile communications with LTE.
2. Orthogonal Frequency Division Multiplex (OFDM) Tutorial Intuitive Guide to Principles of Communications www.complextoreal.com.
3. Long Term Evolution (LTE): an introduction October 2007 White Paper.
4. "3GPP Long-Term Evolution / System Architecture Evolution: Overview" by Ulrich Barth at Alcatel
5. Technical Overview of 3GPP LTE | Hyung G. Myung.
6. EDGE, HSPA and LTE continue to lead and innovate mobile broadband Wednesday, 03 September 2008 www.3gamericas.org.
7. 4G systems, Get the skinny on OFDM, MIMO By Sam Jenkins Principal Engineer CTO Office picoChip Designs Ltd.
8. SC-FDMA for 3GPP LTE uplink Hong-Jik Kim, Ph. D.
9. IMT-2000 & Systems beyond: global spectrum perspectives at the WRC-07 Jean-Pierre Bienaimé Chairman, UMTS Forum www.umts-forum.org.
10. Nomor Research: White Paper on LTE Advance Progress on "LTE Advanced" The new 4G standard Eiko Seidel, Chief Technical Officer Nomor Research GmbH, Munich, Germany.
11. UMTS Long Term Evolution (LTE) Technology Introduction by Rohde & Schwarz
12. "The Long Term Evolution of 3G" on Ericsson Review, no. 2, 2005
13. "3G Long-Term Evolution" by Dr. Erik Dahlman at Ericsson Research
14. "3GPP LTE & 3GPP2 LTE Standardization" by Dr. Lee, HyeonWoo at Samsung Electronics
15. "Overview of the 3GPP LTE Physical Layer" by James Zyren and Dr. Wes McCoy, Freescale Semiconductor
16. "Trends in Mobile Network Architectures" by Dr. Michael Schopp at Siemens Networks
17. (http://www.radio-electronics.com/info/rf-technology-design/pm-phase modulation/8qam-16qam-32qam-64qam-128qam-256qam.php)

PERFORMANCE EVALUATION OF ARRAY ANTENNAS

K.Ch.sri Kavya[1] , Y.N.Sandhya Devi[2] ,G.Sudheer Kumar[2] , Narendra Neupane[2]
[1](Associate Professor, Dept. of Electronics and Communication Engineering, KL University.)
[2](B.Tech Scholars, Dept. of Electronics and Communication Engineering, KL University.)

ABSTRACT

This paper will presents an optimization technique to reduce the side lobe level in the case of linear array antennas. There several methods are introduced for the side lobe reduction .But always the basic trade-off occur when implementing amplitude weighting functions is that a trade between low side lobe levels and a loss in main beam directivity always results . Here we made a comparison of three methods namely uniform illumination method , Taylor line source attenuation method and Taylor line source using redistribution method .The paper also presents different source distributions and their respective directivity patterns.

***Keywords*:** Amplitude weighing, uniform illumination, array antennas, normalization.

I.INTRODUCTION

Array antennas offer a wide range of opportunities in the variation of their directivity patterns through amplitude and phase control. Through the use of individual amplitude and phase control, array antennas offer a wide range of directivity pattern shape implementations to the antenna designer .Synthesis of linear array antennas has been extensively used in the last decades.[9]-[10] .Common optimization goals in array synthesis are the side lobe suppression and null control to reduce interference effects. High directivity antennas have defined main beams whose widths are inversely proportional to their aperture extents. High directivity antennas also have side lobes, which are often undesirable as they may permit reception of energy from undesired directions. The energy from the undesired directions may contain interfering sources such as multipath or even deliberate jammers.

Use of these amplitude weighting functions have a well known effect on the peak of the main beam of the directivity pattern. The amplitude tapering for side lobe reduction reduces the spatial efficiency (or aperture efficiency) of the antenna. Along with the reduction of peak directivity, amplitude tapering also results in a broadening of the main beam.

The purpose of the paper is to present a proper normalization technique to obtain a low side lobe level and to avoid loss in main lobe directivity.

II.METHODOLOGIES

A. Uniform Illumination method

Equal illumination at every element in an array referred to as uniform illumination, results in directivity patterns with three distinct features. Firstly, uniform illumination gives the highest aperture efficiency possible of 100% or 0 dB, for any given aperture area. Secondly, the first side lobes for a linear/rectangular aperture have peaks of approximately –13.1 dB relative to the main beam peak; and the first side lobes for a circular aperture have peaks of approximately –17.6 dB relative to the main beam peak Thirdly, uniform weighting results in a directivity pattern with the familiar sinc(x) or sin(x)/x where x=sin(θ) angular distribution, as shown in Figure.

B. Directivity Pattern Calculations

The directivity pattern calculations given by Hansen [2] and Raffoul and Hilburn [4] are become confusing even though the calculations are not complex.

The below equation presents the calculation for the voltage directivity pattern for a linear array of N elements of isotropic radiators, where Δx is the inter-element spacing, and an is the amplitude of element n. Note that this equation is for the simplest array case of uniform phase for that of a broadside fixed beam array.

$$E(\theta)=\sum_{n=1}^{N} a_n \, e^{-j\left(\frac{2\Pi}{\lambda} n\Delta \, x\sin\,\theta\right)} \qquad (1)$$

The term uniform illumination is often used to describe the array amplitude distribution when the amplitude of all the elements is equal. If the voltage amplitudes all equal one Volt, the peak voltage E_{peak} , for the ideal linear array of isotropic elements occurs when θ is zero and has a value given by Equation

$$E_{peak} = \sum_{n=1}^{N} a_n$$

If a_n=1

$$E_{peak} = \sum_{n=1}^{N} 1 = N \qquad (2)$$

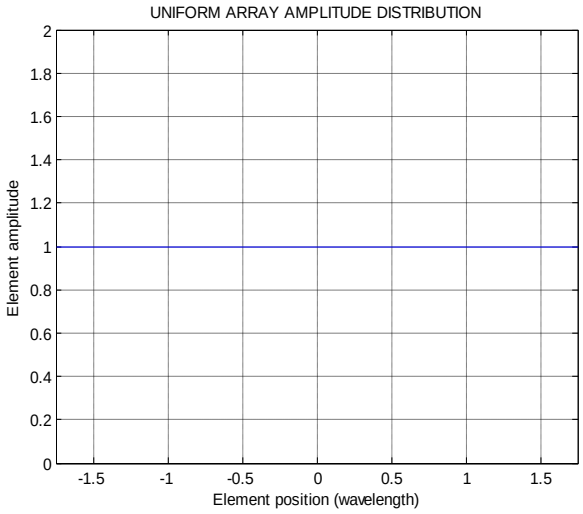


Figure 1.Plot of the uniform amplitude distribution for the eight element array

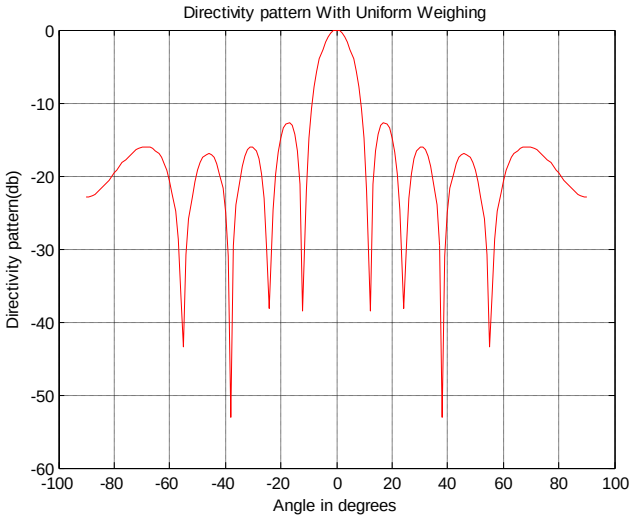


Figure 2.Directivity pattern of a linear array of eight elements with uniform amplitude using MATLAB software

III. TAYLOR SYNTHESIS

Since the 1940s, numerous researchers have contributed varying approaches for synthesizing amplitude distributions for the purpose of side lobe reduction. For this discussion, we will use the Taylor distributions as they are arguably more commonly used for array antenna pattern synthesis The Taylor yields an optimum compromise between beam width and side lode level .The Technique introduced by Taylor to pattern whose first few main lobes(closest to main lobe) are maintained at an equal level .The remaining side lobe levels monotonically decreases [5].The details of the analytical formulation are complex .They are presented in the literature [2][6]. Taylor published his synthesis technique for linear/rectangular [2] and circular [1] apertures in 1955 and 1960, respectively. This method also presents the same directivity pattern as that of Figure in uniform case , except that a Taylor amplitude weighting has been employed to reduce the near in side lobes . The $\bar{n}$ parameter is used to define how many near-in side lobes are held constant at the desired amplitude level. For further detail on this parameter refer to Taylor [1-2].

The Normalized line source which yields the desired pattern is given by

$$I(z') = \frac{\lambda}{l}\left[1 + 2\sum_{p=1}^{\bar{n}-1} SF(p, A, \bar{n}) \cos\left(2\Pi p \frac{z'}{l}\right)\right] \qquad (3)$$

The coefficients SF (p,A,$\bar{n}$) represent samples for Taylor pattern and the can be obtained by

$$SF(p, A, \bar{n}) = \begin{cases} \dfrac{[(\bar{n}-1)!]^2}{(n-1+p)!\,(\bar{n}-1-p)!} \displaystyle\prod_{m=1}^{\bar{n}-1}\left[1 - \left(\frac{\Pi p}{u_m}\right)^2\right] & |p| < \bar{n} \\ 0 & |p| \geq \bar{n} \end{cases}$$

The Taylor space factor is given by

$$SF(u,A,\bar{n}) = \frac{\sin u}{u} \frac{\prod_{n=1}^{\bar{n}-1}\left[1-\left(\frac{u}{u_n}\right)^2\right]}{\prod_{n=1}^{\bar{n}-1}\left[1-\left(\frac{u}{n\Pi}\right)^2\right]} \qquad (4)$$

Where $u = \Pi v = \Pi \frac{l}{\lambda} \cos\theta$

$$u_n = \Pi v_n = \Pi \frac{l}{\lambda} \cos\theta_n$$

Where θ_n represents the locatons of the nulls. From the Figures 2 to 4, the directivity patterns of both are normalized to zero dB. As these patterns are not normalized to a consistent peak, this is limited to no value in assessing the efficiency loss trade-off with side lobe reduction levels.

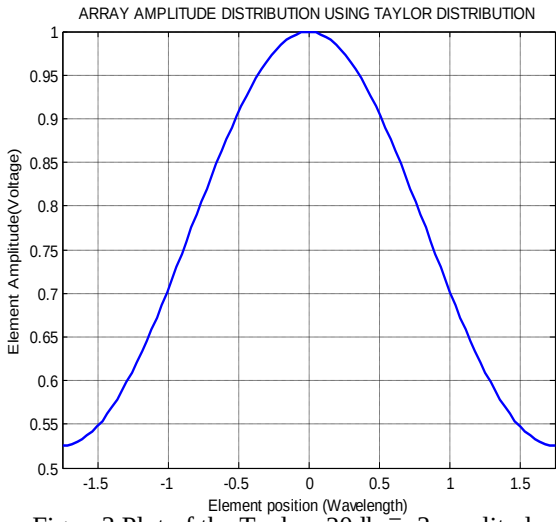


Figure3.Plot of the Taylor -20db $\bar{n}$=3 amplitude distribution.

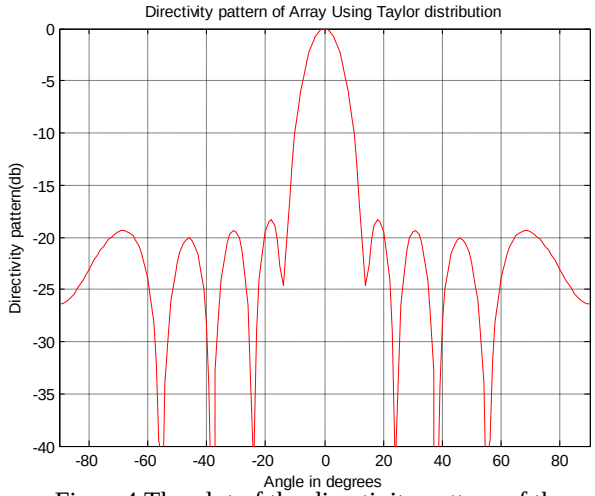


Figure4.The plot of the directivity pattern of the linear array using Taylor -20db $\bar{n}$=3amplitude for near inside lobe reduction.

IV.NORMALIZATIONS

There are two physical methods for generating amplitude distributions for array antennas. Amplitude tapers can be created by either redistributing the power among the elements or by attenuating the power for the outer elements. With the attenuated method, power removed at the outer elements is attenuated in ohmic losses.

A. Attenuation Method

The attenuation method is analogous to achieving the amplitude taper by increasingly resistively attenuating the field energy for radiators toward the periphery of the array to achieve Voltages less than 1. But, this method is the least efficient, and the main beam gain loss is the greatest. Consider the example of a linear array of eight elements having element patterns and an inter-element spacing of 0.695 λ.

The following series of plots were calculated using Equation1, using software written in MATLAB. Routines written to calculate the amplitude weighting coefficients for array side lobe reduction usually provide a_n for each element in Voltage form. Not always, but often, the routines are written to provide a maximum value of 1, The Taylor Voltages calculated for this amplitude illumination function are (from the outer elements to the center) 0.5828,0.7283,0.9147,1 from the figure 3.Where as in the case of uniform illumination every element have the amplitude is equal to 1V.

This approach is further illustrated in Figures 5,6 where the amplitude distributions and resulting directivity patterns are presented for the range of Taylor weightings from –20 to –65 dB. The attenuation method predicts very significant main beam pattern losses

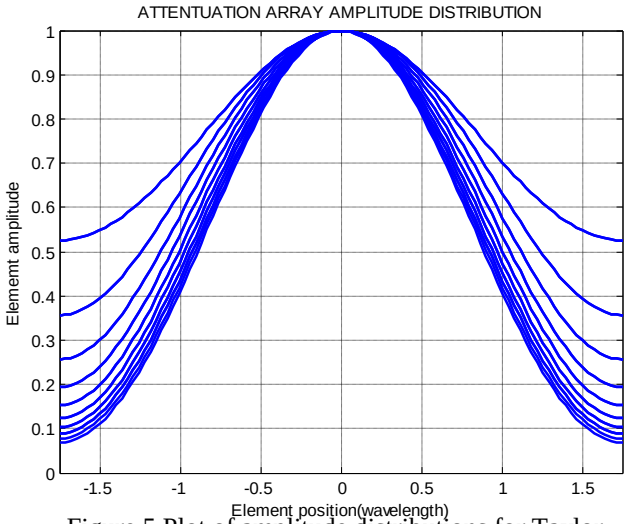


Figure 5.Plot of amplitude distributions for Taylor functions of -20db to -65db using the attenuation normalization.

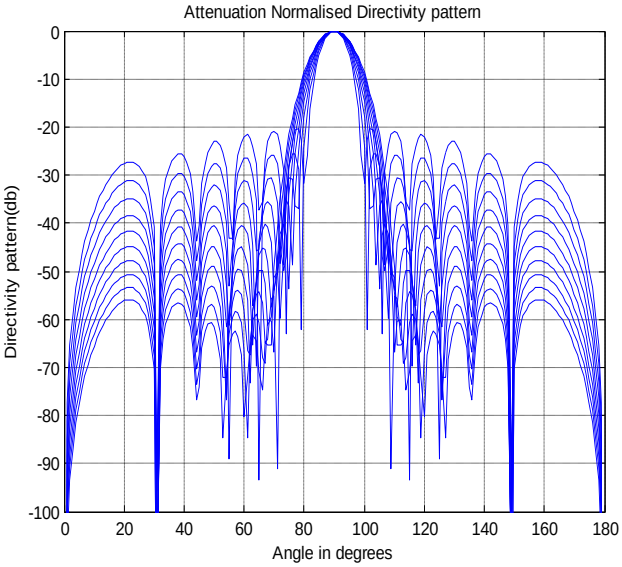


Figure 6.Plot of directivity patterns with amplitude distributions for Taylor functions of -20db to-65db using attenuation normalization

V.UNIFORM ILLUMINATION VS TAYLOR TECHNIQUE

Using Taylor line source (Tchebyshev error) Technique, we get the better side lobe reduction in directivity pattern of linear array but there is a loss in the main beam of the directivity pattern. So a trade is always observed in between low side lobe levels and a loss in main beam directivity always results. In this example the main lobe to first side lobe level in case of Taylor Technique is 18.36 db where as in the case of uniform illumination method is only 12.66db from the figure 7.But the main lobe peak value in the directivity pattern of seven element linear array is 10 db higher than that of Taylor Technique. By Using Taylor, Along with the reduction of peak directivity, amplitude tapering also results in a broadening of the main beam.

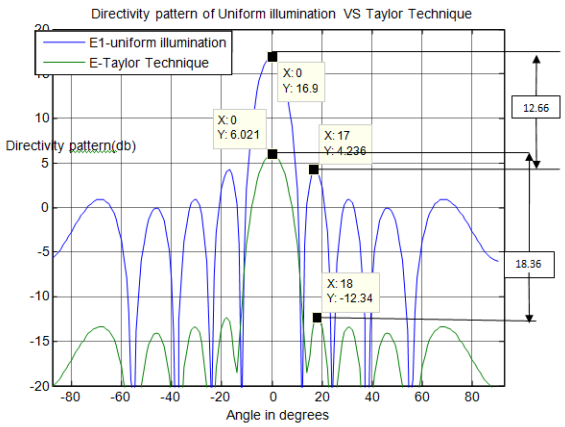


Figure 7.Comparision plot of the directivity patterns of uniform illumination and Taylor using attenuation method

A. Redistribution Method

In contrast to the attenuation method, significantly higher aperture efficiencies can be obtained by redistributing or renormalizing the energy within the amplitude distribution [8]. This can be thought of as conservation of energy as any Voltage removed from outer elements is reallocated to more central elements. The redistributed normalization process can be affected by simply translating the average of the amplitude weights back to one using Equation 5.

$$a_{n,redistributed\quad normalized} = \frac{a_n N}{\sum_{n=1}^{N} a_n} \qquad (5)$$

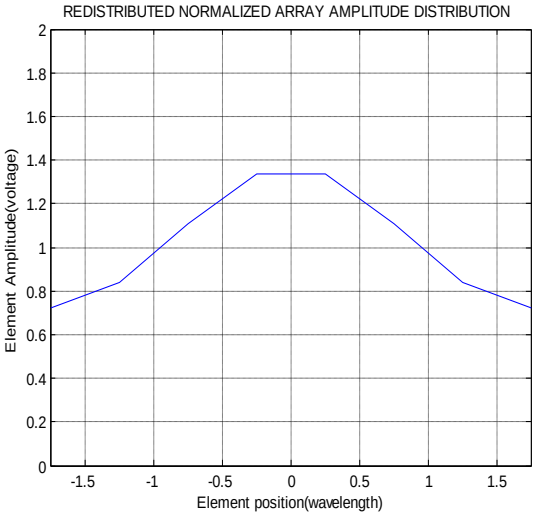


Figure 8.Amplitude Distribution of eight element linear array using taylor redistributded normalizations

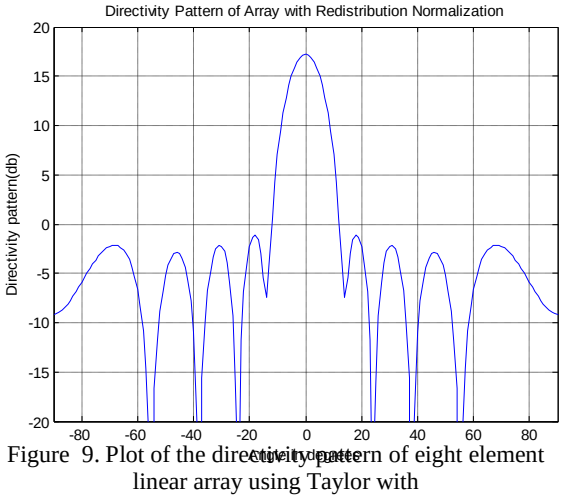


Figure 9. Plot of the directivity pattern of eight element linear array using Taylor with Redistribution normalization

VI.ATTENUATION METHOD VS REDSTRIBUTION

Figure 10 presents a comparison of the directivity patterns of the eight element linear array using Taylor attenuation and redistribution normalizations. Regarding the methods of achieving the tapers, both the methods have same level of reduction in first side lobes but the attenuated method discards the energy while the redistribution method does not lose any energy .There is no loss in the main beam peak value by using Taylor redistribution normalization.

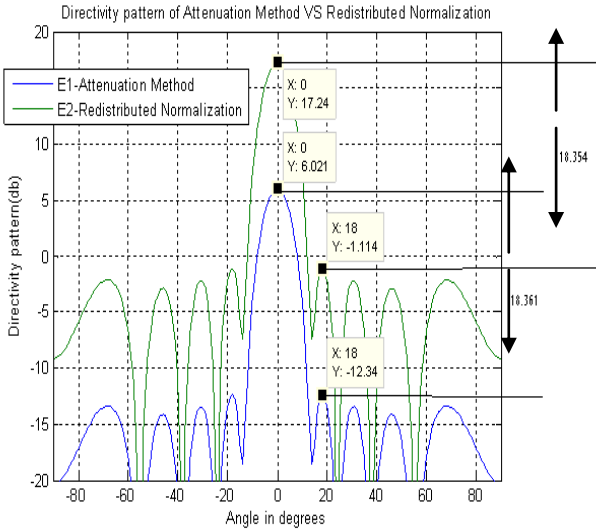


Figure 10.Comparision plot of directivity pattern of eight element linear array using Taylor attenuation and redistribution normalizations.

VII.COMPARISION OF THREE TECHNIQUES

Figure 11 gives the comparison of the three methods like uniform illumination and Taylor attenuation method and redistribution normalization technique.

In the previous section we have seen the comparison of the uniform illumination method and Taylor using attenuation normalization from figure 7and also the comparison of the attenuation and redistribution normalizations from the figure 10.By uniform illumination we get good main lobe directivity but the side lobe reduction is not good. The disadvantage of high side lobe in uniform illumination case is overcome by making use of Taylor attenuation method. Even though we have good reduction in side lobe value we are suffering with loss in main lobe peak value and also results in a broadening of the main Beam. The disadvantage of attenuation method i.e., main lobe loss in directivity pattern is overcome by the Taylor using redistribution normalization .From the figure 11 Even though we avoid the loss of ,main lobe peak in the directivity pattern of eight linear array ,we always suffer with a loss of aperture efficiency is always incurred with both the attenuation and redistribution methods of achieving amplitude tapers when compared to Uniform illumination method.

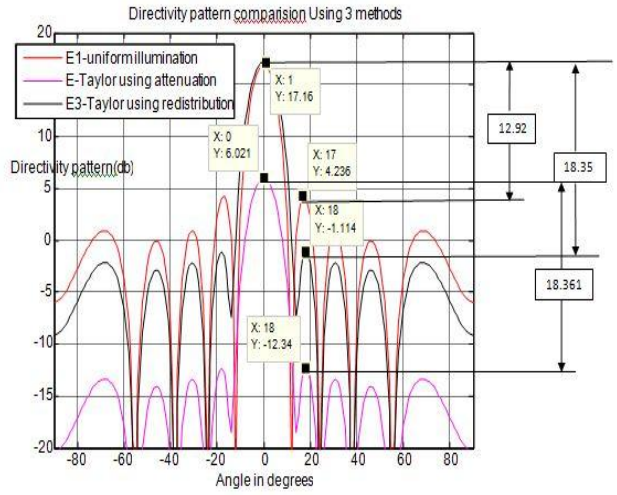


Figure11. Comparison plot of directivity patterns of uniform illumination method and Taylor using attenuation and redistribution normalizations.

VIII.CONCLUSION

This paper presented a discussion of methods used to Suppress the side lobe level of linear array antennas and made the comparisons of source distributions and the directivity patterns of three methods namely uniform illumination method, Taylor line source using attenuation method and redistributed normalization method. Equations to simulate basic array antenna patterns like source distribution and voltage calculations, spacefactor are discussed .The normalizations namely attenuated and redistributed methods were developed and the importance of normalization is discussed. All these are explained by taking example of eight element linear array and the simulations required for this discussion are done using MATLAB Software.

REFERENCES

[1] T.T.Taylor , "One parameter family of line Sources producing modified $\frac{sin(\Pi u)}{\Pi u}$patterns, "*Hughes Aircraft Co.Tech.Mem.324, culver city, calif .,Contract AF 19(604)-262-F-14,September 4,1953*.

[2] Taylor, T. T., "Design of Line-Source Antennas for Narrow Beam width and Low Side Lobes," *IRE Transactions on Antennas and Propagation, January,1955, pp. 16-28.*

[3] Taylor, T. T., "Design of Circular Aperture for Narrow Beam width and Low Side lobes*," IRE Transactions on Antennas and Propagation, January, 1960, pp. 17-22*.

[4] Raffoul, G. W., and J. L Hilburn, "Radiation Efficiency of an X-Band Waveguide Array," *IEEE Transactions on Antennas and Propagation, March 1974, pp. 355-357.*

[5] R.S.Elloit, "Design of Line Source Antennas for Narrow Beam Width and Asymmetric Low Side lobes, "*IEEE Trans. Antennas Propagat.,Vol.AP-23,No.1,PP.100-107,January 1975.*

[6] Hansen, R. C., "Aperture Efficiency of Villeneuve $\bar{n}$ Arrays," *IEEE Transactions on Antennas and Propagation, Vol AP-33, No. 6, June 1985, pp. 666-669.*

[7] ConstantineA.Balanis ,Antenna Theory ,Analysis and Design ,Third Edition Wiley India edition

[8] Glenn D.Hopkins, Justin Ratner Anya Traille, Vic Tripp" Aperture Efficiency of Amplitude Weighting" *IEEE Transactions on Antennas and Propagation, Dec 2006*

[9]T.Isernia,F.J.Aresena O.M.Bucci, M.D'Urso, J.F.Gomez and J.A.Rodrigue," A hybrid approach for the optimal synthesis of pencil beams through array antennas,"*IEEETrans.Antennas propag.,Vol.52,no.11 pp.2912-2918nNov.2004*

[10] N.G.Gomez, J.J.Rodriguez, K.L.Melde and K.M.Mcneill, "Design of low –side lobe linear arrays with high aperture efficiency and interference nulls, "*IEEE Antennas Wireless Propag.lett.,Vol.8,pp. 607-610,2009*

Genetic Operators in TSP

Mr. Akash Kadiyan[1] ,Mr. Sandeep Jaglan[2]
[1]Department of Computer Science, N..C. College of Engineering, Israna
Panipat, Haryana, INDIA.
[1]Department of Computer Science, N.C. College of Engineering, Israna
Panipat, Haryana, INDIA.

ABSTRACT

In this work Traveling salesperson problem is taken as Domain. TSP has long been known to be NP-complete and is a standard example of such problems. Genetic Algorithm (GA) is an approximate algorithm that doesn't always aim to find the shortest tour but to find a reasonably short tour quickly, which is a search procedure inspired by the mechanisms of biological evolution. In genetic algorithms, crossovers are used as a main search operator for TSP. Briefly speaking: the role of crossovers is to generate offspring that are better tours by preserving partial tours from the parents. There were a lot attempts to discover an appropriate crossover operator. This paper presents the strategy which used to find the nearly optimized solution to these type of problems. It is the order crossover operator (OX) which was proposed by Davis, which constructs an offspring by choosing a subsequence of one parent and preserving the relative order of cities of the other parent.

Keywords: **Genetic algorithm, Order Crossover, Travelling salesman problem.**

1. INTRODUCTION

Genetic algorithms are one of the best ways to solve a problem for which little is known. They are a very general algorithm and so will work well in any search space. All you need to know is what you need the solution to be able to do well, and a genetic algorithm will be able to create a high quality solution. Genetic algorithms use the principles of selection and evolution to produce several solutions to a given problem.

1.1 Problem specification

The Traveling Salesman Problem (TSP) is a classic combinatorial optimization problem, which is simple to state but very difficult t o solve. The problem is to find the shortest possible tour through a s e t of N vertices so that each vertex is visited exactly once. This problem is known to be NP-hard, and cannot be solved exactly in polynomial time. Many exact and heuristic algorithms have been developed in t h e field of operations research (OR) to solve this problem. . In the sections that follow, we briefly introduce the OR problem-solving approaches to the TSP. Then, t h genetic algorithms are discussed.

1.2 Exact algorithms

The exact algorithms are designed to find the optimal solution to the TSP, that is, the tour of minimal length. They are computationally expensive because they must (implicitly) consider all solutions in order to identify the optimum. These exact algorithms are typically derived from the integer linear programming (ILP) formulation of the TSP, Where N is the number of vertices, dij is the distance between vertices i and j and the xij's are the decision variables: xij is set to 1 when arc (i,j) is included in the tour, and 0 otherwise. (xij) X denotes the set of sub tour-breaking constraints

Min ij dijxij
Subject to:
j xij = 1 , i=1,..,N
i xij = 1 , j=1,..,N
(xij) X
xij = 0 or 1 ,

That restricts the feasible solutions to those consisting of a single tour. Although the sub tour-breaking constraints can be formulated in many different ways Without the sub tour breaking constraints, the TSP reduces to an assignment problem (AP), and a solution like the one shown in would then b e feasible. Branch and bound algorithms are commonly used to find a n optimal solution to the TSP, and the above AP-relaxation is useful to generate good lower bounds on the optimal value. This is true in particular for asymmetric problems, where dij □ dji for some i,j.

2. METHODOLOGY

2.1 Order Crossover (OX) Davis (85), Oliver et al. (87)

Ordered two-point crossover is used when the problem is of order based, for example in U-shaped assembly line balancing etc. Given two parent chromosomes, two random crossover points are selected partitioning them into a left, middle and right portion. The ordered two-

point crossover behaves in the following way: child 1 inherits its left and right section from parent 1, and its middle section is determined

The Path matrix considered in this is drawn under topic Figures. This crossover operator extends the modified crossover of Davis by allowing two cut points to be randomly chosen on the parent chromosomes. In order to create an offspring, the string between t h e two cut points in the first parent is first copied to the offspring. Then, the remaining positions are filled by considering the sequence of cities in the second parent, starting after the second cut point (when the end of the chromosome is reached, the sequence continues a t position 1).

parent 1 : **1 2 | 5 6 4 | 3 8 7**
parent 2 : 1 4 | 2 3 6 | 5 7 8

---

offspring
(step 1) : - - **5 6 4** - - -
(step 2) : 2 3 **5 6 4** 7 8 1

Figure The order crossover.

Clearly, OX tries to preserve the relative order of the cities in parent 2, rather than their absolute position. In **Figure of PMX** , the offspring does not preserve the position of any city in parent 2. However, city 7 still appears before city 8, and city 2 before city 3 in the resulting offspring.

2.2 Cycle crossover (CX) Oliver et al. (87)

The cycle crossover focuses on subsets of cities that occupy the same subset of positions in both parents. Then, these cities are copied from the first parent to the offspring (at the same positions), and t h e remaining positions are filled with the cities of the second parent. In this way, the position of each city is inherited from one of the two parents. However, many edges can be broken in the process, because the initial subset of cities is not necessarily located at consecutive positions in the parent tours.

In Figure , the subset of cities {3,4,6} occupies the subset of positions {2,4,5} in both parents. Hence, an offspring is created b y filling the positions 2, 4 and 5 with the cities found in parent 1, and by filling the remaining positions with the cities found in parent 2.

parent 1 : **1 3 5 6 4 2 8 7**
parent 2 : 1 4 2 3 6 5 7 8

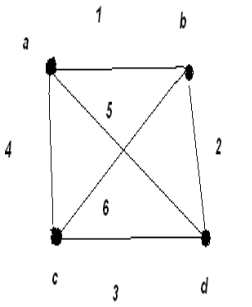
---

offspring : 1 **3** 2 **6 4** 5 7 8

Figure The cycle crossover.

3. RESULTS

3.1 Tables, Figures and Equations



	a	b	c	d
a	0	1	4	5
b	1	0	6	2
c	4	6	0	3
d	5	2	3	0

Figure 1- Path matrix

Table 1- Result of OX and CX

Sample no.	**ox, No. of iteration**	**Shortest path**	**Cyclic crossover, No. of iteration**	**Shortest path**
1	1	86	1	80
2	1	348	3	330
3	1	1727	1	1500
4	1	605	2	590
5	2	2432	2	2388

From the above given table conclusion is clearly drawn out that cyclic crossover is best from the techniques that compared here. In terms of shortest path cyclic crossover(CX) provides improved results.

3.2 EQUATIONS

$F(x) = g(F(x))$ (1)

Where f -objective function , g transforms the value of the objective function to a non-negative number and F-resulting relative fitness.

The most fit individuals and the fitness of the others is determined by the following rules:

- $MIN = 2.0 - MAX$
- $INC = 2.0 \times (MAX - 1.0) / n$
- $LOW = INC / 2.0$ (2)

The fitness of individuals in the population may be calculated directly as,

$$f(x_i) = \frac{2 - MAX + 2\,(MAX - 1)\,x_i - 1}{n - 1} \quad (3)$$

Probability of each chromosomes selection is given by:

$$P_s(i) = f(i) / \sum_{J=1}^{N} f(j) \quad (3)$$

$P_s(i)$ and $f(i)$ are the probability of selection and fitness value

4- DISCUSSION

Genetic algorithms for the TSP-

The description of the genetic algorithm included many genetic terms. In order to better understand how genetic algorithms can be applied to combinatorial optimization problems, the following equivalence will be useful. Combinatorial Optimization Genetic Algorithm, Encoded Solution , Chromosome Solution, Decoded Chromosome Set of Solutions, Population Objective function.

Fitness function-

In a TSP context, each chromosome encodes a solution to the problem (i.e. a tour). The fitness of the chromosome is related to the tour length, which in turn depends on the ordering of the cities. Since the TSP is a minimization problem, the tour lengths must be transformed, so that high fitness values are associated with short tours, and conversely. A well-known approach is to subtract each tour length to the maximum tour length found in the current population. Other approaches are based on the rank of the tours in the population The genetic algorithm searches the space of solutions by combining the best features of two good tours into a single one. Since the fitness is related to the length of the edges included in the tour, it is clear that the edges represent the basic information to be transferred to the offspring. The success or failure of the approaches described in the following sections, can often be explained by their ability or inability to adequately represent and combine the edge information in the offspring. Difficulties quickly arise when the simple "pure" genetic algorithm is applied to a combinatorial optimization problem like the TSP. In particular, the encoding of a solution as a bit string is not convenient. Assuming a TSP of size N, each city would b e coded using 2 log N bits, and the whole chromosome would encode our as a sequence of N* 2 log N bits. Accordingly, most sequences in the search space would not correspond to feasible tours. For example, it would be easy to create a sequence with two occurrences of the same city, using the mutation operator. Moreover, when the number of cities is not a power of two, some bit sequences in the code would not correspond to any city. In the literature, fitness functions with penalty terms, and repair operators to transform infeasible solutions into feasible ones have been proposed to alleviate these problems However, these approaches were designed for very specific application domains, and are not always relevant in a TSP context.

The preferred research avenue for the TSP is to design representational frameworks that are more sophisticated than the bit string, and to develop specialized operators to manipulate these representations and create feasible sequences. For example, applying the crossover operator at position 2 creates two infeasible offspring, as illustrated in Figure

tour (12564387) : 1 2 | 5 6 4 3 8 7
tour (14236578) : 1 4 | 2 3 6 5 7 8

---

offspring 1 : 1 2 2 3 6 5 7 8
offspring 2 : 1 4 5 6 4 3 8 7

Figure Application of the one-point crossover on two parent tours.

None of the two offspring is a permutation of the cities. The TSP, as opposed to most problems tackled by genetic algorithms, is a pure ordering problem. Namely, all chromosomes carry exactly the same values and differ only in the ordering of *these values. Accordingly, specialized permutation* operators must be developed for this problem. In the following sections, we explain how genetic algorithms can be tailored to the TSP. The extensions proposed in the

literature will be classified according to the representational framework used t o encode a TSP tour into a chromosome, and the crossover operators used to manipulate these representations

5. CONCLUSION

Gives the results of experiments comparing the proposed method with the conventional approach. In these experiments, the number of children generated by one crossover is limited because of calculation costs. Therefore, they tried generate better individuals within a limited number of children and again purposed a new crossover method that accelerates the local search efficiency. First, the children generated by the first parents are evaluated for their fitness. Then, some number of top children with an elite rate set beforehand is selected as the next parents.

TSP is optimization problem which is used to find minimum path for salesperson. The Actual use of tsp is routing in network. Minimum path will helps to reduce the overall receiving time and improves system performance. The work proposed here intends to test the performance of different Crossover used in GA and compare the performance for each of them and compare to others. This thesis presents an investigation on different crossover techniques used in GA .

Since there are other methods traditionally adopted to obtain the optimum distance for TSP. This work aims at establishing the superiority of Genetic Algorithms in optimizing TSP. One of the objectives of this research work is to find a way to converge fast. Since precise minimum path remains a great challenge, the objective of this work is to develop some new and practical model with computational intelligence algorithms. As can be seen from the references, many models have been developed for TSP. From the experimental results the conclusion can be drawn that different methods might outperform the others in different situations.

6. REFRENCES

[1]. Potvin ,Jean-Yves(n.d) Genetic Algorithms for the Traveling Salesman Problem: Montréal (Québec)Canada H3C 3J7, Centre de Recherche sur les Transports Université de Montréal

[2]. Qi-yi, Zhang & Shu-chun, Chang (2009) An Improved Crossover Operator of Genetic Algorithm China: Transportation Command Department Automobile Management Institute of PLA

[3]. Rasheed, Khaled(1999)Guided Crossover:A New Operator For Genetic Algorithm Based Optimization,NewBrunswick,NJ08903,USA: Computer Science Department, Rutgers University

[4]. Shang, Yi &Li, Guo-Jie(1991) New Crossover Operators In Genetic Algorithms, *P. R. China: National Research Center for Intelligent Computing Systems (NCIC).*

[5]. Singh, Vijendra & Choudhary, Simran (2009) Genetic Algorithm for Traveling Salesman Problem: Using Modified Partially-Mapped Crossover Operator, sikar, Rajasthan, India: *Department of Computer Science & Engineering, Faculty of Engineering & Technology, Mody Institute of Technology & Science, Lakshmangarh*

[6]. Su, Fanchen et al(2009) New Crossover Operator of Genetic Algorithms for the TSP, P.R. China: Computer School of Wuhan University Wuhan.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-520-526 ISSN: 2249-6645

Security Issues in Wireless Sensor Networks

Priyanka Shrivastava
(Department of Computer Science & Engineering, Shri Ram Institute of Technology, Jabalpur
Rajiv Gandhi Technical University, Bhopal)

ABSTRACT

Wireless Sensor Networks (WSN) is a recent advanced technology of computer networks and electronics. The WSN increasingly becoming more practicable solution to many challenging applications. The sensor networks depend upon the sensed data, which may depend upon the application. One of the major applications of the sensor networks is in military. So security is the greatest concern to deploy sensor network such hostile unattended environments, monitoring real world applications. But the limitations and inherent constraints of the sensor nodes does not support the existing traditional security mechanisms in WSN. Now the present research is mainly concentrated on providing security mechanism in sensor networks. In this context, security aspects of the sensor networks like requirements, classifications, and type of attacks etc., is analyzed in this survey paper.

1. INTRODUCTION

The sensor network is a group of self- organized, low priced sensor nodes and creates network in spontaneous manner. The WSN combines sensing, computation and communication in a single small device, called Sensor Node. The sensor node mainly contains radio, battery, microcontroller and power devices. Another term of sensor node is "mote". The sensors in a node provides the facility to get the data like pressure, temperature, light, motion, sound etc and capable of doing data processing. The main goal of the applications is achieved by the cooperation of all sensor nodes in the network. There are many sensor network applications like such environmental data collection, security monitoring, medical science, military, tracking etc. when sensor networks are randomly deployed in a hostile environment, security becomes extremely important factor. Because sensed data of sensor nodes is prone to different types of malicious attacks before reaching base station. Security mechanisms are needed in communication part of the networks to provide safe data. The security is also important concern to get full advantage of in-network data processing sensor networks. Protecting such a sensed data is complicated task. Even through wireless sensor network is an advanced technology of network, it is extremely different from traditional wireless networks. This is, due to the unique characteristics of sensor nodes in WSN. So existing security mechanisms of traditional wireless networks are not directly applied in WSN. Sensor networks are closely interacting physical environment. So sensor nodes are also deployed in all areas even physical accessible attacks and broadcasting sensed data in network. So these reasons give a scope to new security mechanism rather than applying existing traditional security mechanisms in WSN.

2. SECURITY REQUIREMENTS IN WSN

The objective of confidentiality is required in sensors environment to protect information traveling among the sensor nodes of the network or between the sensors and the base station from disclosure. Authentication in sensor networks is essential for each sensor node and base station to have the ability to verify that the data received was really sent by trusted sender or not. This authentication is needed during the clustering of sensor node in WSN. We can trust the data sent by the nodes in that group after clustering. Integrity controls must be implemented to ensure that information will not be altered in any unexpected way. Many sensor applications such as pollution and healthcare monitoring rely on the integrity of the information to function with accurate outcomes. Secure management is needed at base station, clustered nodes, and protocol layer in WSN. Because security issues like key distribution to sensor nodes in order to establish encryption and routing information need secure management.

3. ATTACKS IN WSN

The basic categories of attacks against privacy in sensor networks are eavesdropping, disruption and hijacking. The eavesdropping is used to know the output of sensor networks by listing transmitted messages of sensor nodes. There are mainly two ways to know about output data by concealing from sensor nodes or sending queries to sensor nodes or root nodes or aggregation points or attacks sensor nodes. The former approach is called passive eavesdropper and later approach is called active eavesdropper. The location of eavesdropper plays major role in getting information. This attack affects the property of confidentially, authentication in WSN. So proper encryption mechanism, message authentication code are needed before broadcasting data. The disruption mainly influences output of the network. The semantic disruption injects messages, corrupts data or changes values in order to render the aggregate data corrupted, useless and incomplete. Physical disruption renders the sensor readings by directly manipulating the environment. The hijacking approach is used to take the control over sensor node in network. The hijacking mechanism gives more power to eavesdropping and disruption by hijacking main sensor nodes. Another major attack in WSN is Denial of Service attacks. Some of the denials of service attack are at routing layer, link layer and transport layer. One of the denials of service attack is jamming networks. That is simply interfaces transmission frequency of WSN. There are mainly two types in jamming. In constant jamming, no messages are able to send or receive by a node in WSN. So this is complete jamming of network. In Intermittent jamming, the nodes are exchange messages with highly risks. Another new attack in WSN is Sybil attack. This Sybil attack is defined as a "malicious

device illegitimately taking on multiple identities". This attack is affecting redundancy mechanism, routing algorithms, resource allocation procedure and data aggregation mechanism. With little effort, an adversary may capture nodes, analyze and replicate them, and surreptitiously insert these replicas at strategic locations within the network. They may allow the adversary to corrupt network data or even disconnect significant parts of the network. This attack can change entire network goal. This attack affects Integrity, confidentiality.

4. SECURITY MECHANISMS

Now days, the researchers are attracted by security concepts of wireless sensor networks. Many researchers have proposed some security mechanisms in wireless sensor networks. In this section, we are dealing briefly on several existing security mechanisms for WSN's. These are:

4.1 "SecFleck: Public key cryptography in wireless senor networks"

Approach is used to provide the message security services as confidentiality, Integrity and Authenticity in WSN at computationally fast and lower energy utilization. To design and implementation of public key system in WSN needs new version hardware and software in mote. This approach is named as secFleck. It uses trusted module platform chip at hardware level and some software primitives. This approach uses RSA algorithm to implement asymmetric public key system. This approach has taken smaller RSA exponent (65537) and key size (2048) to provide security levels. This approach uses new operating system called Flack OS (FOS). FOS is a c-based cooperative multi-threaded operating system with public key cryptography primitives like encryption, decryption, singing, signature verification etc. Even this approach works fine for message security level, the learning new OS functions is length and complicated process. It also needs new hardware to provide message security level.

4.2 "LiSP: A Lightweight security protocol for wireless sensor networks"

LISP aims to provide authentication without retransmission of keys and also provides scalability in computing. It uses symmetric key system approach. It uses temporary keys and master keys. Temporary keys (TK) are used to encrypt and decrypt data packets. The master key (MK) is used to send temporary keys to single node. After network had been deployed, this protocol automatically selects one group of cluster heads as key server. The key server is used to distribute the temporal key, authenticate new nodes and detect nodes that have been compromised. When a key server transmits a packet for the first time it contains the length of the TK buffer, the key refresh rate, and the initial TK. The need for a Message Authentication Code is eliminated because the nodes are able to implicitly authenticate the TK by checking to see if the new TK matches the sequence of the other TK's in the TK buffer. LiSP provides a great deal of protection from compromised nodes and key servers. The keying system with implicit authentication allows the sensor to quickly detect whether or not the key that was sent from the key server is authentic or not. As long as the refresh rate is not very fast the sensors will not run out of battery power at a fast rate. LiSP is very scalable because the key server does most of the calculations and the key server can change depending on whether the key server has been compromised or not. This protocol is used to reduce the retransmission of keys and provides implicit message authentication scheme to reduce the overhead. The keying mechanism depends upon application of wireless sensor networks.

4.3 TinySec: A link layer security architecture for wireless sensor networks"

TinySec is a light weight and link layer security protocol. It provides security services as message Integrity, message authentication and access control at routing level and Reply protection in Adversary. It supports two different security options. They are Authenticated Encryption and Authentication only. In the Authenticated Encryption, the payload is encrypted first and then packet is encrypted using MAC. In Authentication only, the packet is directly encrypted with MAC without encrypting payload. This approach is used Cipher Blocked Chaining to encryption. TinySec is independent of cipher, key scheme, and application. The TinySec packets are more in size then WSN packets, due to this; it needs more computing and processing power.

4.4 "SPINS: Security Protocol for Wireless Sensor Networks"

This protocol is used to provide security services as freshness, Authentication, Confidentiality and Integrity. The two-way authentication, data confidentiality, freshness and integrity are provided with the help of Secure Network Encryption Protocol (SNEP) scheme and Authentication for Broadcast messages is provided with the help of µTELSA (the "micro" version of the Timed, Efficient, Streaming, and Loss-tolerant Authentication Protocol) scheme. A block cipher RC5 algorithm was used by SNEP But it gives chances to eavesdropping to get plain and cipher text in way. Due to semantic security is low in SNEP implementation. The Localized Encryption and Authentication Protocol security mechanism provides confidentiality and authentication mechanisms in sensor networks. This mechanism uses four different keys for each sensor node and controller to maintain master keys. They are individual key, pair-wise key, cluster key and group key. The individual key is unique for each node and used to provide secure communication between node and base station. This key is pre-loaded into each sensor node before deployment. A cluster key is a shared key and is shared by all neighbor nodes in the cluster. It is mainly used for securing broadcast messages in cluster groups because in-network computation is done at the cluster heads in WSN.

The pair-wise shared key used to provide secure communication and authentication between immediate nodes or one hop nodes in WSN. This key is used before transmitting cluster key in cluster group. It is generated when the same key nodes are deployed in a single hop distance. The group is also a shared key. This key is shared by base station and set of nodes for broadcasting encrypted messages. This key used for hop-by-hop translation messages. The nodes are stationary in this approach. This approach needs more resource in-terms of computation power, memory to store keys and processing resources. But

according to sensor network characteristics, this approach is inefficient and power consumable. This approach does not give good results on security damaged sensor applications. This approach should be applied prior to deployment of sensor network application.

In Random key pre-distribution schemes, a centralized key server generates a large key pool at offline. This generation of keys is done in key distribution phase. In key discovery phase, each sensor broadcasts their key identifiers or private shared keys. Then sensor nodes get the information about neighbor and network information after processing shared keys. The communication of data has to be done by shared key authentication. Too many sensor nodes are usually deployed for any sensor applications. Assigning unique keys to sensor node is a cumbersome problem. Even thorough, this mechanism used modified schemes like Purely Random Key Pre-distribution and Structured Key Pool Random Key Pre-distribution are inefficient to assigning keys to nodes in WSN. The attackers make use of advantage of decentralized pool key generation. Public cryptography such as such as Diffie-Hellman key establishment at booting stage in base station, gives single point of failure of sensor network. So to provide efficient security mechanism, decryption should be done at cluster nodes and communicates the nodes or distributes messages in hierarchical manner. This scheme reduces number of keys in network, resource utilization and make utmost impossible to attacker to hijack.

4.5 "Fast Authenticated Key Establishment Protocols for Self-Organizing wireless Sensor Networks"

This protocol has a goal to provide efficient authenticated key transferring mechanism. It uses elliptic Curve Cryptography (ECC) to provide encryption for sensor nodes. Cracking the private key is very difficult even the size of ECC keys length is less. Public keys are used to authenticate keys certificates. So during the process of authenticate keys certificates, this approach is usually finds public keys. These certificates are generated by sensor node and security manager. This work is accomplished by computation server if needed. The main drawback of using this key establishment protocol is that sometimes a computation server may be needed for some of the computations. The amount of packets that are exchanged to authenticate a key seems like lengthy process to authenticate a key. It is difficult to figure out the strength of this protocol. Because this depends upon the keys and they contains random values.

The adversary attack leads to node replication attack with little effort. One approach to detect the replication node in wireless sensor networks is centralized scheme. In the Centralized scheme, all nodes in the network transfers a list of their neighbor's claimed locations to a central base station. Then base station can examine the lists for conflicting location claims. Even though this approach is efficient, the nodes closest to the base station will receive the brunt of the routing load and will become attractive targets for the adversary. This protocol is also delays revocation, since the base station must wait for all of the reports to come in, analyze them for conflicts and then flood revocations throughout the network. Suppose adversary attacks at base station then centralized approach is inefficient and does not do well. At this case, this protocol gives single point of failure. The network lifetime is also decreases due to high traffic at base station surroundings. Even though this approach detects all replicated node in easy way, it requires more storage area in each node and also requires communication messages. Another scheme to overcome the difficulties in centralized scheme is Location Detection scheme. In this scheme, instead of implementing node replication detection scheme at base station, it process at node's neighbor. It uses a voting mechanism; it collects neighbor's opinions on the legitimacy of the node. This approach is unable to detect the clones (i.e. nodes giving support to adversary) in disjoint neighborhood in network. It fails to detect subvert and clone if they are more than two hops away. Due to these drawbacks, this protocol became inefficient to find replication nodes in WSN. One simple approach to detect the distributed replication nodes is Simple Broadcast protocol. In this approach, each node broadcast authenticated messages about their location and also stores the information about neighbor nodes. Even though this approach gives 100% results, it may not works if adversary attacks at key areas or communication paths. This approach costs more in form of communication for large networks. One of the improvements of Simple Broadcast Scheme is Deterministic Multicast Protocol. The main of this approach is to reduce the communication of simple broadcast scheme by sharing the node's location to a subset of deterministically chosen node, called witness node. This subset may be fixed for a particular node. The witness nodes are selected based on function of node ID's and probability. So it uses multicast approach to give judgment over nodes location claim. Due to this, the number of message transfers in the network is decreased. This is also fails if adversary attacks or jams the messages in the network. Because it shares the node's location to a limited subset of deterministically chosen nodes only. This approach is not doing well, if any one of the witness node is caught by adversary.

4.6 "Distributed Detection of node replication attacks in wireless sensor networks"

This protocol deals with detection of node replication attacks due to adversary at protocol level (routing layer). It uses two routing algorithms Randomized Multicast and Line selected Multicast. The adversaries have to be detected as soon as it occurs otherwise replicated nodes are increases in next data gathering cycle. Assume that the adversary cannot readily create new IDs for nodes. In the cloned formation, this assumed to be at least one node as legitimate neighbor to clone. It also assumes the adversary in stealthy manner. Due to this, the detection of adversary is complex. So it uses one protocol that sweeps the network, using SWATT technique to remove compromised node and human interactions. Here it assumes that the adversary can read and write the messages using only nodes under adversary control. [i.e. read and writing messages should do in adversary control parts by adversary.]

This also works in a situation that, the adversary can change the topology of the network by adding replicas.

4.7 Commutative Cipher based En-route Filtering (CCEF)

CCEF exploits a bootstrapping phase to establish trust between individual sensor nodes and the base station. In the operational phase, the base station can initiate a query-response session and install per-session security states in the sensor nodes at any time. The tasked sensor nodes response by generating and endorsing data reports on their sensing results. When the reports are forwarded to the base station, each intermediate node verifies the authenticity of the reports, and filters the fabricated ones. The base station further verifies the reports that it receives, and reacts to the compromised nodes by refreshing the session state.

Commutative Cipher based En-route Filtering scheme (CCEF) defends against event fabrication attacks without symmetric key sharing among sensor nodes. CCEF exploits the typical operational mode of query-response in sensor networks, and installs security states in the nodes in an on-demand manner. Specifically, in CCEF, each node has a unique ID and is preloaded with a unique node key. The base station initiates a query-response session by sending out a query to task specific sensor nodes to report their sensing results. The base station prepares two keys for each session: one session key and one witness key. The session key is securely sent to source node, i.e., the node tasked to generate reports, while the witness key is in plaintext and recorded by all intermediate nodes. A legitimate report is endorsed by a node MAC jointly generated by the detecting nodes using their node keys, and a session MAC generated by the source node using the session key. Through the usage of a commutative cipher, a forwarding node can use the witness key to verify the session MAC, without knowing the session key, and drop the fabricated reports. The base station further verifies the node MAC in the report that it receives, and refreshes the session key upon detection of compromised nodes.

4.8 Interleaved hop-by-hop authentication (IHA)

This technique deals with false data injection attack by enabling the base station to verify the authenticity of a report that it has received as long as the number of compromised sensor nodes does not exceed a certain threshold. Further, it attempts to filter out false data packets injected into the network by compromised nodes before they reach the base station, thus saving the energy for relaying them.

This scheme is particularly useful for large-scale sensor networks where a sensor report needs to be relayed over several hops before it reaches the base station and for applications where the information contained in the sensor reports is not amendable to the statistical techniques used by SIA (e.g., non-numeric data). In this scheme at least t + 1 sensor nodes have to agree upon a report before it is sent to the base station. Further, all the nodes that are involved in relaying the report to the base station authenticate the report in an interleaved, hop-by-hop fashion. Here t is a security threshold based on the security requirements of the application under consideration and the network node density. This scheme guarantees that if no more than t nodes are compromised, the base station will detect any false data packets injected by the compromised sensors. In addition, for a given t, this scheme provides an upper bound B for the number of hops that a false data packet can be forwarded before it is detected and dropped. If every non compromised node on the path between a cluster head and the base station knows the ids of the nodes that are t +1 hops away from it on the path, then B = t; otherwise, without this knowledge, B = (t - 1)(t - 2).

4.9 Localized Encryption and Authentication Protocol (LEAP)

Localized Encryption and Authentication Protocol, a key management protocol for sensor networks is designed to support in-network processing, while at the same time providing security properties similar to those provided by pair wise key sharing schemes. In other words, the keying mechanisms provided by LEAP enable in-network processing, while restricting the security impact of a node compromise to the immediate network neighborhood of the compromised node. LEAP includes support for multiple keying mechanisms. The design of these mechanisms is motivated by the observation that different types of messages exchanged between sensor nodes have different security requirements, and that a single keying mechanism is not suitable for meeting these different security requirements. Specifically, this protocol supports the establishment of four types of keys for each sensor node– an individual key shared with the base station, a pair wise key shared with another sensor node, a cluster key shared with multiple neighboring nodes, and a group key shared by all the nodes in the network. Moreover, the protocol used for establishing these keys for each node is communication and energy-efficient, and minimizes the involvement of the base station.

LEAP also includes an efficient protocol for inter-node traffic authentication based on the use of one-way key chains. A salient feature of the authentication protocol is that it supports source authentication (unlike a protocol where a globally shared key is used for authentication) without preventing passive participation (unlike a protocol where a pair wise shared key is used for authentication).

The packets exchanged by nodes in a sensor network can be classified into several categories based on different criteria, e.g. control packets vs data packets, broadcast packets vs unicast packets, queries or commands vs sensor readings, etc. The security requirements for a packet will typically depend on the category it falls in. Authentication is required for all type of packets, whereas confidentiality may only be required for some types of packets. For example, routing control information usually does not require confidentiality, whereas (aggregated) readings transmitted by a sensor node and the queries sent by the base station may need confidentiality. No single keying mechanism is appropriate for all the secure communication that is needed in sensor networks. As such, LEAP supports the establishment of four types of keys for each sensor node –an individual key shared with the base station, a pair wise key shared with another sensor node, a cluster key shared with multiple neighboring nodes, and a group key that is shared by all the nodes in the network.

Individual Key Every node has a unique key that it shares pair wise with the base station. This key is used for secure communication between a node and the base station. For example, a node can use this key to compute message authentication codes (MACs) for its sensed readings if the readings are to be verified by the base station. A node may also send an alert to the base station if it observes any abnormal or unexpected behavior of a neighboring node. Similarly, the base station can use this key to encrypt any sensitive information, e.g. keying material or special instruction that it sends to an individual node. **Group Key** is a globally shared key that is used by the base station for encrypting messages that are broadcast to the whole group. For example, the base station issues missions, sends queries and interests. Note that from the confidentiality point of view there is no advantage to separately encrypting a broadcast message using the individual key of each node. However, since the group key is shared among all the nodes in the network, an efficient rekeying mechanism is necessary for updating this key after a compromised node is revoked.

Cluster Key is a key shared by a node and all its neighbors, and it is mainly used for securing locally broadcast messages, e.g., routing control information, or securing sensor messages which can benefit from passive participation. Researchers have shown that in-network processing techniques, including data aggregation and passive participation are very important for saving energy consumption in sensor networks. For example, a node that overhears a neighboring sensor node transmitting the same reading as its own current reading can elect to not transmit the same. In responding to aggregation operations such as MAX, a node can also suppress its own reading if its reading is not larger than an overheard one. For passive participation to be feasible, neighboring nodes should be able to decrypt and authenticate some classes of messages, e.g., sensor readings, transmitted by their neighbors. This means that such messages should be encrypted or authenticated by a locally shared key. Therefore, in LEAP each node possesses a unique cluster key that it uses for securing its messages, while its immediate neighbors use the same key for decryption or authentication of its messages.

Pair wise Shared Key Every node shares a pair wise key with each of its immediate neighbors. In LEAP, pair wise keys are used for securing communications that require privacy or source authentication. For example, a node can use its pair wise keys to secure the distribution of its cluster key to its neighbors, or to secure the transmissions of its sensor readings to an aggregation node. Note that the use of pair wise keys precludes passive participation.

4.10 Location-aware end-to-end data security (LED)

LED is an integrated security design providing comprehensive protection over data confidentiality, authenticity, and availability. It overcomes the limitations of the existing hop- by-hop security paradigm and achieves an efficient and effective end-to-end security paradigm in WSNs. It exploits the static and location-aware nature of WSNs, and proposes a novel location-aware security approach through two seamlessly integrated building blocks: a location-aware key management framework and an end-to-end data security mechanism. In this approach, each sensor node is equipped with several types of symmetric secret keys, some of which aim to provide end-to-end data confidentiality, and others aim to provide both end-to-end data authenticity and hop-by-hop authentication. All the keys are computed at each sensor node independently from keying materials preloaded before network deployment and the location information obtained after network deployment, without inducing extra communication overhead for shared key establishment. Location-aware end-to-end data security design (LEDS) then provides a secure and reliable data delivery mechanism, which is highly resilient to even a large number of compromised nodes.

The features of LEDS and the contributions are outlined as follows:

In LEDS, the targeted terrain is virtually divided into multiple cells using a concept called virtual geographic grid. LEDS then efficiently binds the location (cell) information of each sensor into all types of symmetric secret keys owned by that node. By this means, the impact of compromised nodes can be effectively confined to their vicinity, which is a nice property absent in most existing security designs. What the attacker can do is to misbehave only at the locations of compromised nodes, by which they will run a high risk of being detected by legitimate nodes if effective misbehavior detection mechanisms are implemented. Second, LEDS provides end-to-end security guarantee. Every legitimate event report in LEDS is endorsed by multiple sensing nodes and is encrypted with a unique secret key shared between the event sensing nodes and the sink. Furthermore, the authenticity of the corresponding event sensing nodes can be individually verified by the sink. This novel setting successfully eliminates the possibility that the compromise of nodes other than the sensing nodes of an event report may result in security compromise of that event report, which is usually the case in existing security designs.

Third, LEDS possesses efficient en-route false data filtering capability to deal with the infamous bogus data injection attack. As long as there are no more than t compromised nodes in each single area of interest, LEDS guarantees that a bogus data report from that cell can be filtered by legitimate intermediate nodes or the sink deterministically. Last, LEDS provides high level assurance on data availability by counteracting both report disruption and selective forwarding attacks, simultaneously. By taking advantage of the broadcast nature of wireless links, LEDS adopts a one-to-many data forwarding approach, which is fully compatible with the proposed security framework. That is, all reports in LEDS can be authenticated by multiple next-hop nodes independently so that no reports could be dropped by a single node(s). Thus, LEDS is highly robust against selective forwarding attacks as compared to the traditional one-to-one forwarding approach used by existing security designs. In addition, LEDS adopts a (t; T) threshold linear secret sharing scheme (LSSS) so that the sink can recover the original report from any t out of T

legitimate report shares. Not only this approach enhances the event report authenticity by requiring T sensing nodes to collaboratively endorsement the report, but also makes LEDS resilient to the interference from up to T;t compromised nodes in the event area. LEDS is highly resilient to both types of attacks.

4.11 Location-based resilient security (LBRS)

This technique overcome the threshold limitation and achieves graceful performance degradation to an increasing number of compromised nodes. Location-based security approach based on two techniques: location-binding keys and location-based key assignment. In this approach, symmetric secret keys bind to geographic locations, as opposed to sensor nodes, and assign such location-binding keys to sensor nodes based on their deployed locations. a Location-Based Resilient Security (LBRS) solution, demonstrates that such a location-based approach can effectively limit the damage caused by even a large collection of compromised nodes. In LBRS, the terrain is divided into a regular geographic grid, and each cell on the grid is associated with multiple keys. Based on its location, a node stores one key for each of its local neighboring cells and a few randomly chosen remote cells. To detect fabricated reports, it is required that a real event be endorsed through multiple keys bound to the specific location of the event. An attacker that has compromised multiple nodes may obtain keys bound to different cells, but he cannot combine such keys to fabricate any event without being detected. To limit the damage of network re source waste, each node uses its keys of remote cells to verify and drop forged reports passing through it.

Location-based security design is highly resilient to compromised nodes for three reasons. First, it prevents the attacker from arbitrarily abusing a compromised key, be cause a key bound to a geographic location can only be used for purposes related to that particular location (e.g., to endorse events detected there). Second, it constrains the damage when the attacker compromises multiple nodes and accumulates their keys, because a collection of keys bound to different locations cannot be used together for any meaningful purpose. Finally, it limits the keys stored by individual nodes, because each node is assigned only a few keys based on its location. As a result, the security protection offered by our design degrades gracefully, without any threshold break-down, when more and more nodes are compromised.

4.12 Statistical En-route Filtering (SEF)

SEF exploits the sheer scale and dense deployment of large sensor networks. To prevent any single compromised node from breaking down the entire system, SEF carefully limits the amount of security information assigned to any single node, and relies on the collective decisions of multiple sensors for false report detection. When a sensing target (henceforth called “stimulus” or “event”) occurs in the field, multiple surrounding sensors collectively generate a legitimate report that carries multiple message authentication codes (MACs). A report with an inadequate number of MACs will not be delivered. As a sensing report is forwarded towards the sink over multiple hops, each forwarding node verifies the correctness of the MACs carried in the report with certain probability. Once an incorrect MAC is detected, the report is dropped. The probability of detecting incorrect MACs increases with the number of hops the report travels. Depending on the path length, there is a non-zero probability that some reports with incorrect MACs may escape enroute filtering and be delivered to the sink. In any case the sink will further verify the correctness of each MAC carried in each report and reject false ones. This is the first effort that addresses false sensing report detection problems in the presence of compromised sensors. SEF is able to detect and drop 80 to 90% injected reports by a compromised node within 10 forwarding hops, thus reducing energy consumption by 50% or more in many cases. The SEF design seeks to achieve the following goals:

4.12.1 Early detecting and dropping of false data reports

Identifying false reports allows the user to avoid taking responses to fabricated events. Although this can be done either during the data delivery process or at the sink after the data is delivered, early en-route detection of such reports can prevent them from reaching the sink, thus saving energy and bandwidth resources of nodes on data forwarding paths.

4.12.2 Low computation and communication overhead

Given the resource constraints of low-end sensor nodes, SEF strives to scale to large sensor networks and be resilient against node failures. We will show that by using only hash computations which are efficient even on low-end sensor hardware, SEF can detect and en-route drop false reports injected by an attacker who captures up to a threshold number of nodes. SEF consists of three components which work in concert to detect and filter out forged messages: (1) each legitimate report carries multiple MACs (in the form of a Bloom filter) generated by different nodes that detect the same stimulus, (2) intermediate forwarding nodes detect incorrect MACs and filter out false reports en-route, and (3) the sink verifies the correctness of each MAC and eliminates remaining false reports that elude en-route filtering.

In SEF there is a global key pool. However only the sink has the knowledge of the entire pool. Each sensor stores a small number of keys that are drawn in a randomized fashion from the global key pool before deployment. Once a stimulus appears in the field, multiple detecting nodes elect a Center-of-Stimulus (CoS) node that generates the report. Each detecting node produces a keyed MAC for the report using one of its stored keys. The CoS node collects the MACs and attaches them to the report in the form of a Bloom filter. These multiple MACs collectively act as the proof that a report is legitimate. A report with an insufficient number of MACs will not be forwarded. The key assignment procedure should ensure that each node can only generate part of the proof for a legitimate report. Only by the joint efforts of multiple detecting nodes can the complete proof be produced. Therefore to get a forged data report forwarded a compromised node has to forge MACs to assemble a seemingly complete proof. At the same time, the key assignment procedure should also ensure that any two nodes share common keys with a certain probability. When the report with forged MACs is forwarded by

intermediate nodes, probabilistic key sharing allows them to examine the correctness of the MACs probabilistically, thus detecting and dropping false reports en-route. The sink serves as the final goal-keeper for the system. When it receives reports about an event, the sink verifies every MAC carried in the report because it has complete knowledge of the global key pool. False reports with incorrect MACs that sneak through en-route filtering will then be detected.

Currently SEF also does not address the issues of how to identify compromised nodes or revoke compromised keys. For identification, neighbor nodes may overhear the channel to detect unusual activities of compromised nodes such as high traffic volume and notify the sink. After the nodes are identified, the user may deploy new nodes and the sink could flood instructions to revoke compromised keys and propagate new ones.

In summary, SEF is not designed to address all the attacks that a compromised node may launch, such as dropping legitimate reports passing through it, recording and replaying legitimate reports, or injecting false control packets to disrupt other protocols. Existing techniques can be used to address some of these issues points out that one can use multipath forwarding to effectively alleviate dropping of legitimate reports demonstrate that sensors can use a cache to store the signatures of recently forwarded reports, thus preventing identical packets from being forwarded again.

5. Conclusion

Sensor networks serving mission-critical applications are potential targets for malicious attacks. Although a number of recent research efforts have addressed security issues such as node authentication, data secrecy and integrity, they provide no protection against injected false sensing reports once any single node is compromised. These techniques aim at detecting and dropping such false reports injected by compromised nodes. Takes advantage of the large scale and dense deployment of sensor networks. Collaborative filtering of false reports requires that nodes share certain amount of security information. The more security information each forwarding node possesses, the more effective the en-route filtering can be, but also the more secret the attacker can obtain from a compromised node. Further step includes evaluation of the tradeoffs between these two conflict goals, and gaining further insight on how to build a sensor network that can be at once resilient against many compromised nodes as well as effective in detecting false data reports through collaborative filtering.

REFERENCES

[1] M. Anand, Z. Ives, and I. Lee. “Quantifying Eavesdropping Vulnerability in Sensor Networks”, In Proc. of the 2nd International VLDB Workshop on Data Mgnt. for Sensor Networks (DMSN), 2005.

[2] G. Gaubatz, J.-P. Kaps, and B. Sunar. “Public key cryptography in sensor networks”, –Revisited. In Proc. of the 1st ESAS, 2004.

[3] A. Perrig, J. Stankovic, and D. Wagner. “Security in Wireless Sensor Networks”, Communications, ACM, 47(6):53–57, 2004.

[4] David W. Carman, Peter S. Kruus, and Brian J. Matt. Constraints and approaches for distributed sensor network security. NAI Labs Technical Report #00-010, September 2000.

[5] L. Zhou and Z.J. Hass. Securing ad hoc networks. 13(6), November/December 1999.

[6] A. Wood and J. Stankovic, .Denial of Service in Sensor Networks, IEEE Computer, Oct. 2002.

[7] Elaine Shi and A. perrig, .Designing Secure Sensor Networks,. Wireless Communication Magazine, 11(6), December 2004.

[8] J. Jung, T. Park and C. Kim, .A Forwarding Scheme for Reliable and Energy-efficient Data Delivery in Cluster-based Sensor Networks,. IEEE Communication Letters, Vol.9, No.2: 112-114, Feb. 2005.

[9] W. Du, J. Deng, Y. Han, and P. Varshney. A Pairwise Key Pre-distribution Scheme for Wireless Sensor Networks. In Proc. of 10th ACM Conference on Computer and Communications Security (CCS), Washington DC, October 27-31, 2003.

[10] H. Vogt. Integrity Preservation for Communication in Sensor Networks. Technical Report No. 434, ETH Zrich, Institute for Pervasive Computing, February 2004.

[11] D. Malan. Crypto for tiny objects. Technical Report TR-04-04, Harvard University, 2004.

[12] F. Ye, H. Luo, S. Lu, and L. Zhang. Statistical en-route filtering of injected false data in sensor networks. In INFOCOM, 2004.

[13] F. Ye, G. Zhong, S. Lu, and L. Zhang, “GRAdient Broadcast: A Robust Data Delivery Protocol for Large Scale Sensor Networks,” ACM Wireless Networks (WINET), vol. 11, no. 2, March 2005.

[14] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, D. Sivakumar, and Luca Trevisan. Counting distinct elements in a data stream. In Proc. RANDOM 2002, pages 1–10, 2002.

[15] C. Karlof and D. Wagner. Secure Routing in Sensor Networks: Attacks and Countermeasures. In Proc. of First IEEE Workshop on Sensor Network Protocols and Applications, May 2003.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-527-537 ISSN: 2249-6645

Accurate Identification of Performance for Rotor-Bearing Systems Using the Modified Modelling Under Gyroscopic Effect

H.A.AL-Khazali[1] and M.R.Askari[2]

[1] School of Mechanical & Automotive Engineering, Kingston University, London,UK
[2] School of Aerospace & Aircraft Engineering, Kingston University, London,UK

ABSTRACT

The rotor-bearing system of modern rotating machines constitutes a complex dynamic system. The challenging nature of rotordynamic problems have attracted many scientists and engineers whose investigations have contributed to the impressive progress in the study of rotating systems. The purpose of the present paper is to investigate the effects of modal parameters on the noise produced by rotor-bearing systems under gyroscopic effect. To do this, we study reaction force in left and right bearing under gyroscopic effect in rotating machinery with high speed of rotation using modal data. We find modal parameter of modal in experimental part validate with simulation using ANSYS 12., and study effect of mass eccentricity of the rotor on the noise of the bearing are investigated, and the simulation results are presented advanced modelling and simulation techniques; active vibration controls; malfunctions and condition monitoring aspects through the graph of the bending stress with respect time of the bearing for various rotational speeds of the rotor.

Keyword-**Rotor-Bearing, Modelling, Reaction force, Bending stress, Gyroscopic effect.**

I. INTRODUCTION

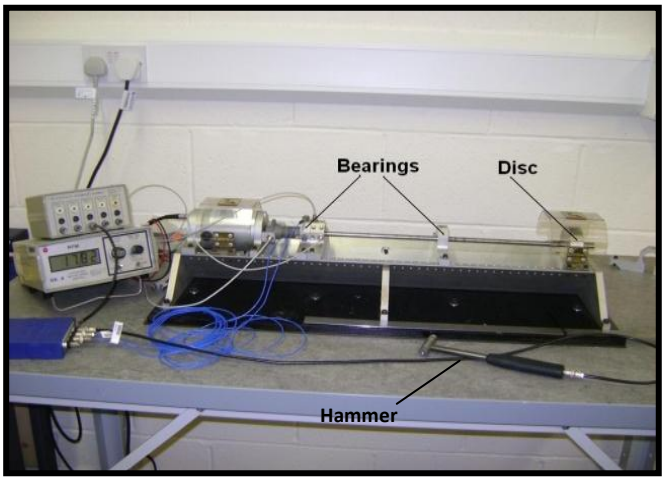
The bearings used for supporting rotating machinery are one of the crucial elements by which the safe operation of the machinery is ensured. In recent years, with continuing demands for increased performance, many rotating industrial machines are now being designed for operation at high speed, a trend which has resulted in increased mechanical vibration and noise problems. Many researchers have studied the vibration characteristics of bearings [1–3], but there is relatively little information regarding their modified modelling under gyroscopic effect; (A gyroscope Fig.(1) is a device that can be used to maintain orientation based on the principles of angular momentum. It is a mechanism by means of which a rotor is journeyed to spin around an axis) [4,5]. However, there have been no studies on the effects of design parameters on the noise of rotor-bearing systems. In practice, it is very important to know how much the bearing noise can be [6]. However, there have been no studies on the effects of design parameters on the noise of rotor-bearing systems. In practice, it is very important to know how much the bearing noise can be reduced by design parameters such as bearing width, radial clearance, oil viscosity, mass eccentricity of the rotor, and so on. In other words, it is very important to know what parameters are dominant on bearing noise. It is also expected that [6,7].

The modal properties of the bearing can provide diagnostic information on abnormal phenomena of the rotor-bearing system. For example, if the frequency characteristics .The purpose of the present paper is to investigate the effects of modal parameters on the noise of rotor bearing systems. With the advancement in high-speed machinery and increases in their power/weight ratio, the determination of the rotor dynamic characteristics through reliable mathematical models gains prime importance. The advancement in modern instrumentation and computational capabilities has helped in implementing simulation techniques of these complex models. Modern machinery is bound to fulfill increasing demands concerning durability as well as safety requirements. On-line condition monitoring strategies are becoming increasingly commonplace in a greater range of systems [8,9].

Rotors are structures with special properties due to their rotation (causing e.g. the gyroscopic effect), due to their bearings (fluid film bearings, magnetic bearings) and in many cases due to surrounding fluids (seal forces). Therefore rotordynamics requires special engineering tools although the structural properties of the rotors and their supports could well be modelled by any general finite element program [4,10].

The recent development of magnetic bearings, which are now more and more introduced in industrial applications of turbomachines, required an extension of existing rotordynamic tools to model the specific characteristics of this bearing type and the controllers [8,9&10].

II. METHODS



Picture.1 Experimental setup for the modal testing.

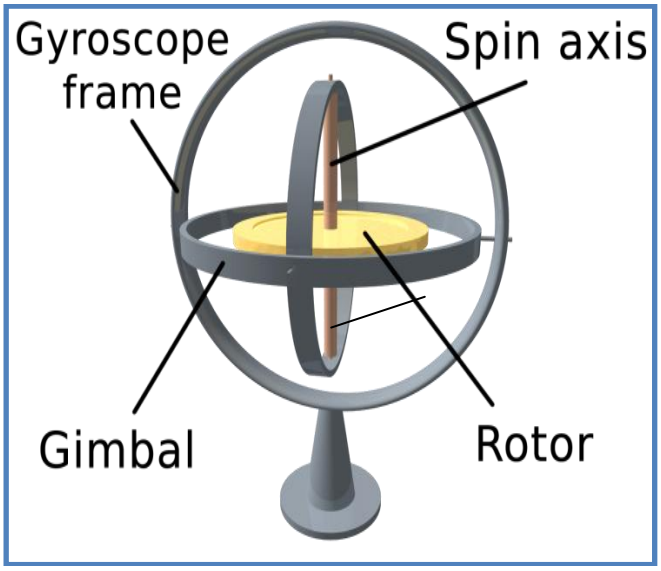


Fig.1 The gyroscopic effect [4,5].

2.1 Equations of motion

The general equations of motion for a multi-degree of freedom vibratory system shown in picture (1), may be written as [10,11]:-

$$[M]\{\ddot{q}(t)\}+[[G+C](\Omega)]\{\dot{q}(t)\}+[[B]+[K](\Omega)]\{q(t)\}=\{F(t)\} \quad .. (1)$$

$$Y_i(x)=\begin{cases}\dfrac{P\,a\,x}{6EIL}(x^2-l^2); & 0\le x\le l\\ \dfrac{P\,(x-l)}{6EIl}[a(3x-l)-(x-l)^2]; & (l\le x\le l+a)\end{cases} \quad ....(2)$$

Table (1) Defintion of parameter for gyroscopic setup.

	Rotor Dia.	**0.01 m**		
	P	0.8 kg	P=M*9.81	0.007848KN
	X	0.24 m		
	a	0.24 m		
I=π*d^4/64	I	4.91E-10	MASS MOMENT OF INERTIA	

Table (2) Calculations natural frequency & stiffness of the system before rotation.

Y deflection=1.18E-03			
ω=89.99623 rad/sec		89.99623	rad/s
ω_n=89.99623 rad/sec			
f_n	14.32334486	Hz	
n	859.4006918	rpm	
ω=(k/M)^0.5		k=M*(ω_n^ 2)	
	K	6479.457131	N/m

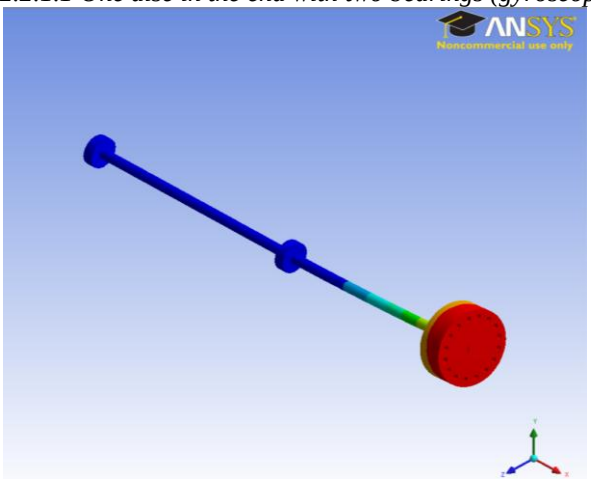
2.2 Imitation model in (ANSYS 12.)

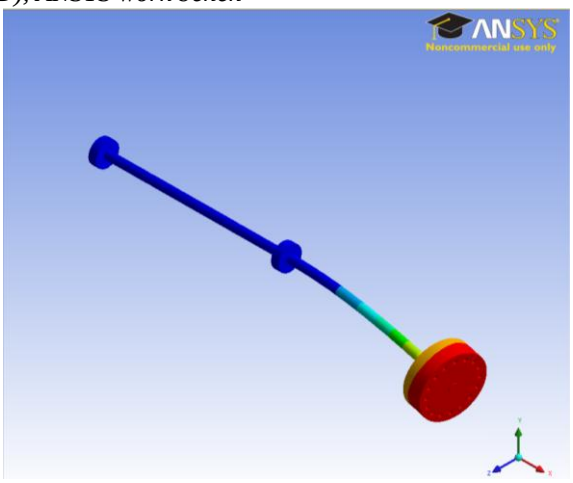
A program has been written in (ANSYS 12),A model of rotor system with an overhung disc with multi degree of freedom (Y and Z directions) has been used to demonstrate the above capability see Fig.(2).Postprocessing commands(/POST1). Applying of gyroscopic effect to rotating structure was carried by using (CORIOLIS) command. This command also applies the rotating damping effect. Another command which was used in input file (SYNCHRO) that Specifies whether the excitation frequency is synchronous or asynchronous with the rotational velocity of a structure in a harmonic analysis; [10,12&13].

Fig. 2 Finite element model (gyroscopic geometry) ANSYS work bench (three dimensions).

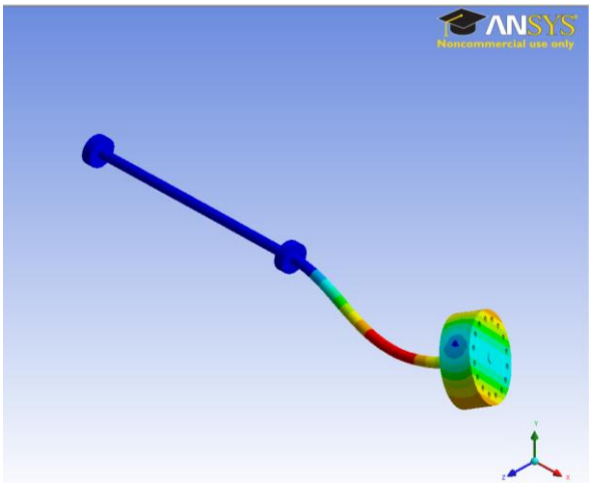
2.2.1 The ANSYS Animation

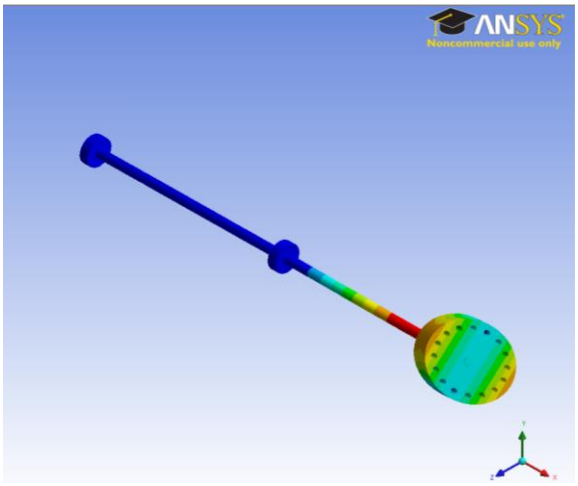
2.2.1.1 One disc in the end with two bearings (gyroscopic effect)(3D), ANSYS work bench



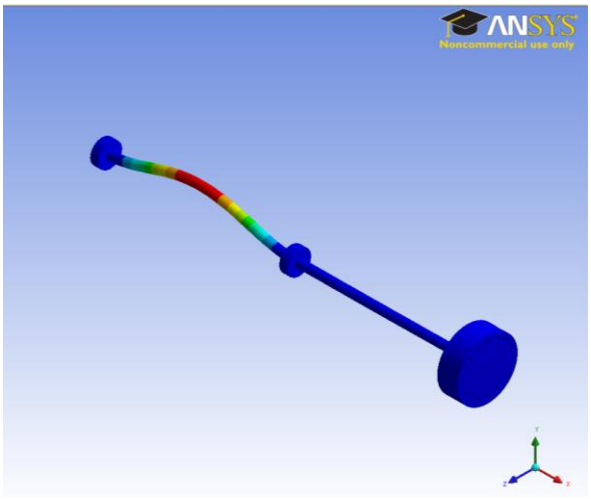


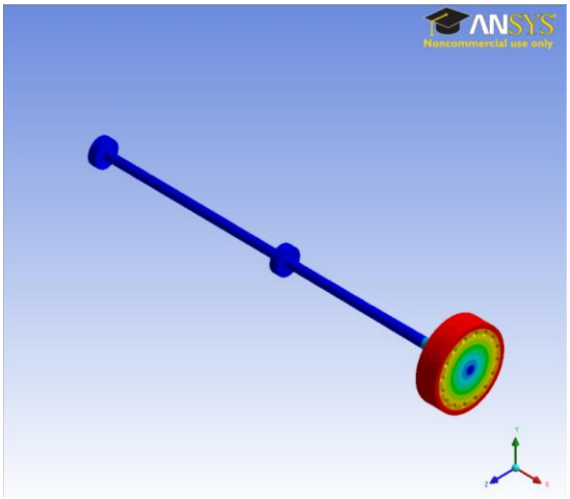
A-First mode shape.Natural frequency 15.47 Hz,(3-D).





B-Second mode shape.Natural frequency 217.01Hz,(3–D).

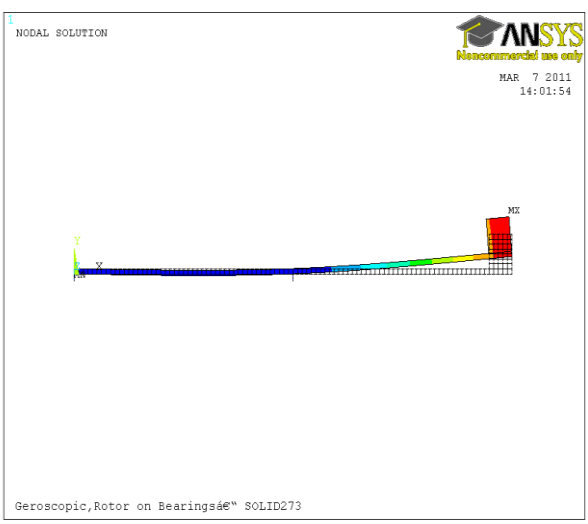


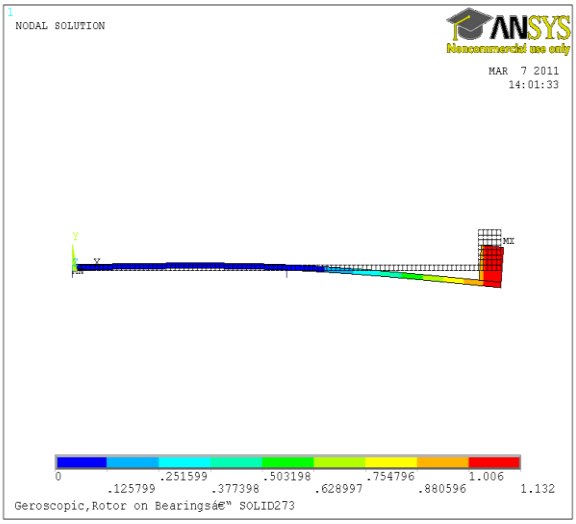


C-Third mode shape.Natural frequency 508.06Hz. D-Fourth mode shape.Natural frequency 626.85Hz.

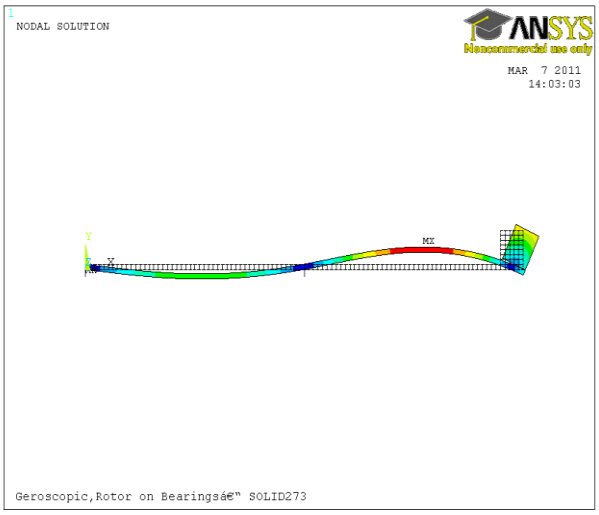
Fig.3 Finite element method simulations (FEM),different mode,ANSYS workbench;

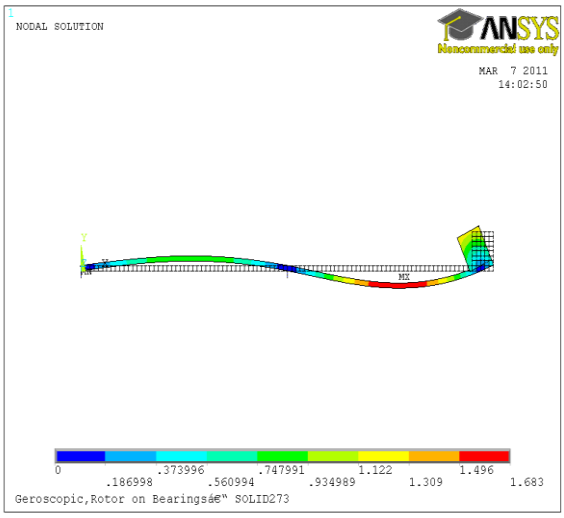
2.2.1.2 One disc in the end with two bearings (Gyroscopic effect),(2D) ANSYS APDEL



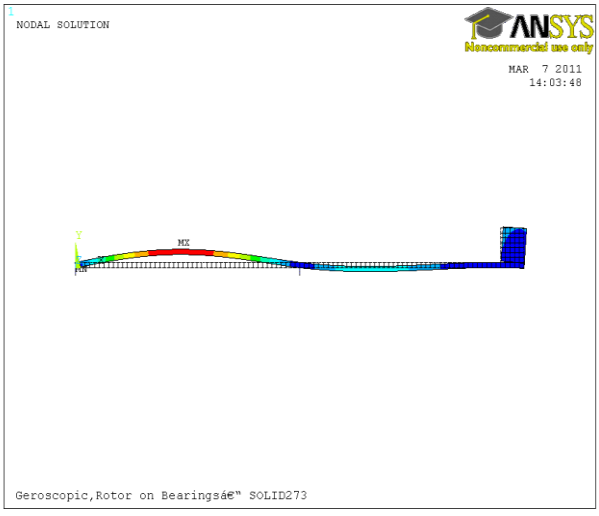


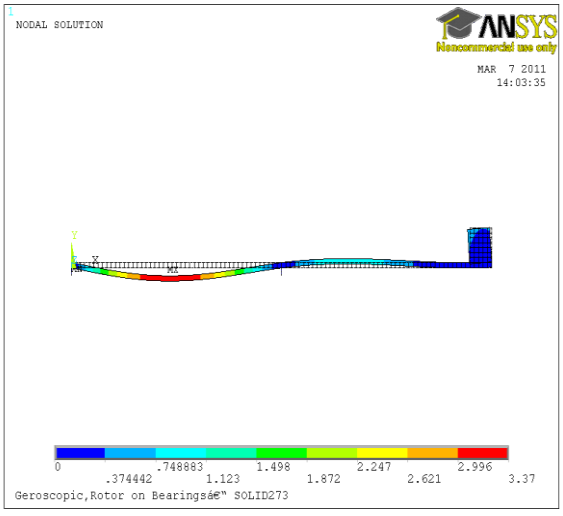
A- First mode shape.Natural frequency 15.703 Hz,(2-D).





B-Second mode shape.Natural frequency 216.8 Hz,(2–D).





C-Third mode shape.Natural frequency 507.39Hz,(2-D).

Fig.4 Finite element method simulations (FEM),different mode, ANSYS APDEL;

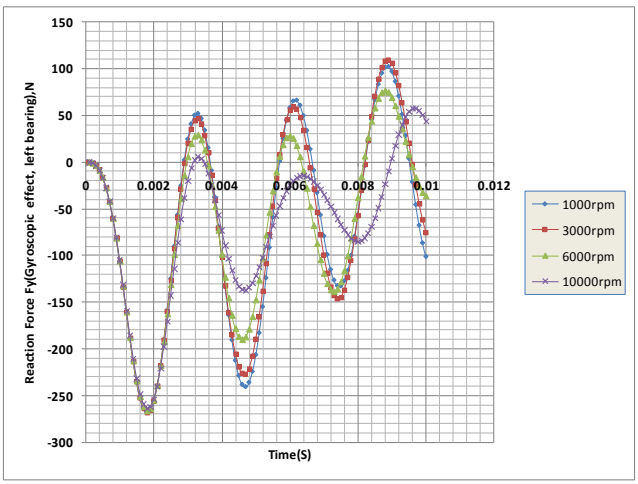
2.3 Test setup

The rotor consisted of a shaft with a nominal diameter of 10 mm, with an overall length of 610 mm.Two journal bearings, RK4 Rotor Kit made by Bentley Nevada (the advanced power systems energy services company), could be used to extract the necessary information for diagnostic of rotating machinery, such as turbines and compressor. The test rotor is shown in picture (1). Basically; Been testing the process will be conducted on the rotary machine as the project is based on rotary dynamics reach practical results for the purpose of subsequently applied machinery rotary by using (Smart office program),and then do the experimental testing using the impact test, installed fix two accelerometer(model 333B32),sensitivity (97.2&98.6) mv/g in Y&Z direction and roving the hammer(model 4.799.375,S.N24492) on each point for the purpose of generating strength of the movement for the vibration body and the creation of vibration for that with, creating a computer when taking reading in public that he was dimensions and introducing it with the data within the program (Smart office)[14&15&16].Configuration for testing on the machines with rotary machine the creation of all necessary equipment for that purpose with the design geometry wizard[17].

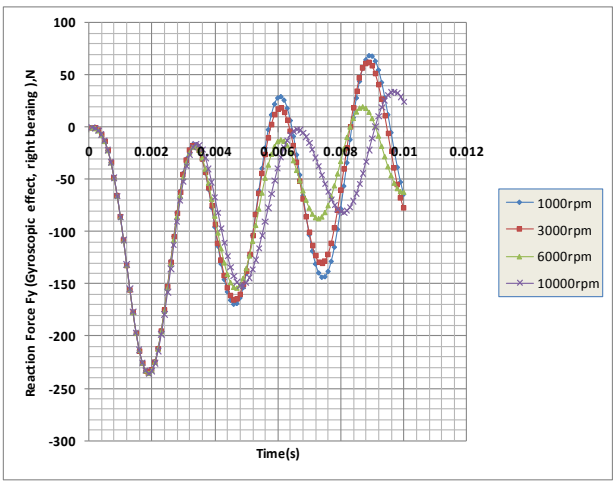
III. RESULTS (TABLES&FIGURES)

3.1 Response forces in the left and right bearings (gyroscopic effect)

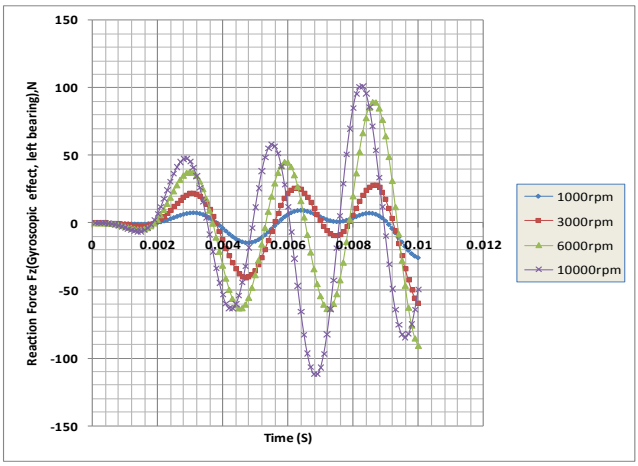
We find the relation between the reaction forces with respect time by using further simulation, can we see from the Fig.(5-A,B,C,D),the performance of reaction forces in the right and left bearings with different speed of rotations:-



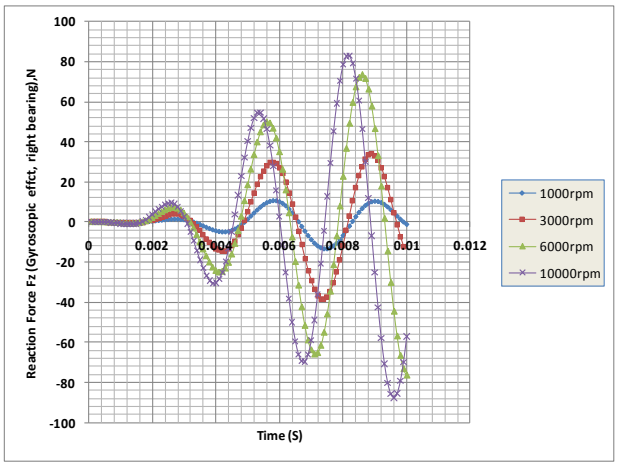
A-Reaction force(Fy) left bearing.



B-Reaction force(Fy) right bearing.



C-Reaction force (Fz) left bearing.



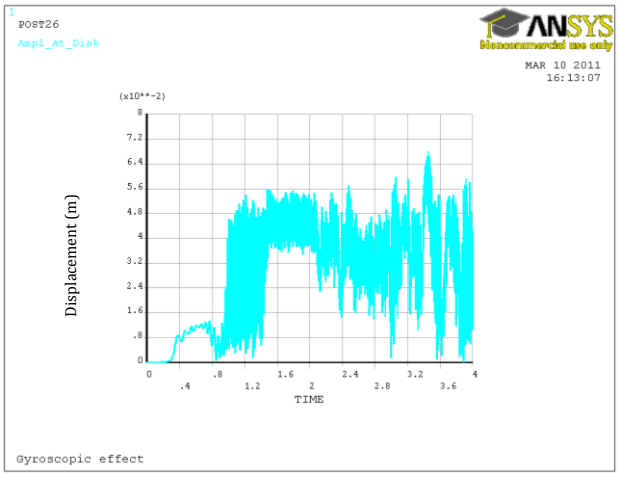
D-Reaction force(Fz) right bearing.

Fig.5 Relation between reaction force bearings versus time at different speed of rotation(gyroscopic effect);

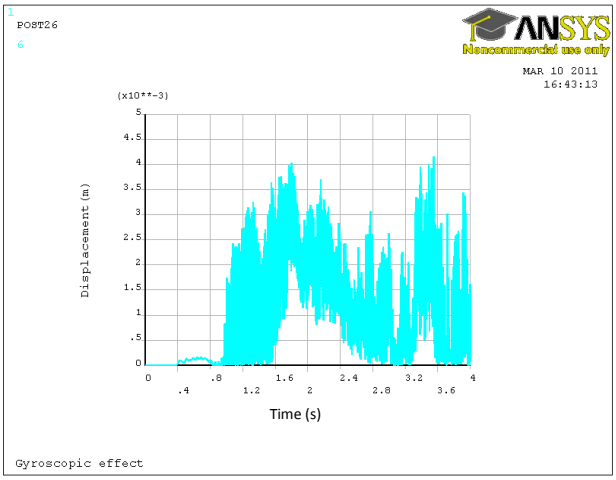
3.2 Unbalance effect

3.2.1 Unbalance with add mass (simulation result)

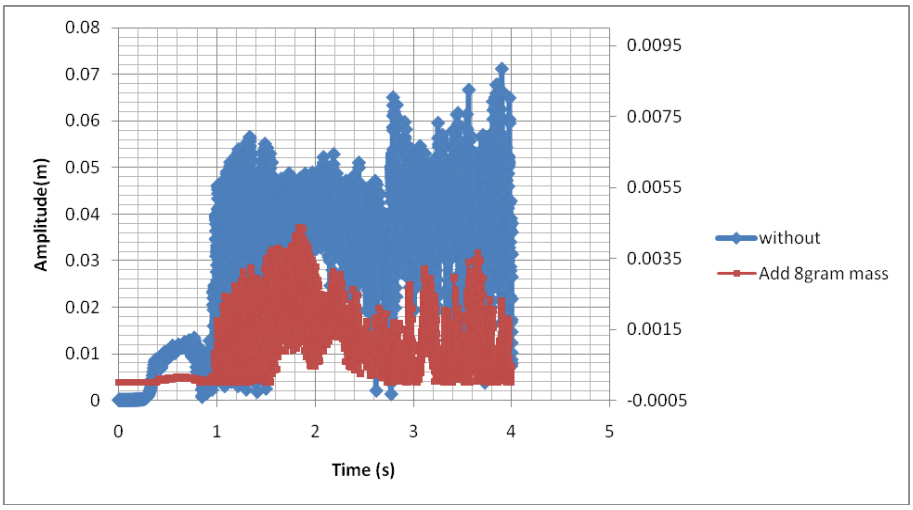
In this set simulation, unbalance loading is applied to the system to be at the optimum phase angles of $Ø$=90° *and* $Ø$=270° respectively .ANSYS simulation of the set shown in Fig.(6).



A- Displacement versus time before add mass.



B-Displacement versus time after add,8 gram mass.



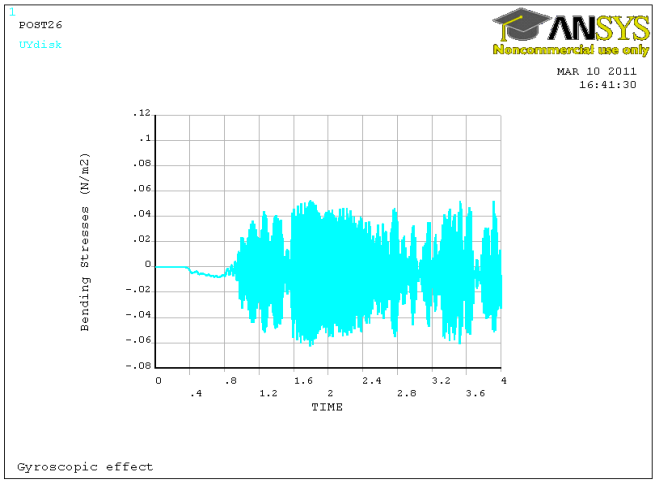
C-Merge comparison.

Fig. 6 The Amplitude versus time,(A-With out load,B-After add 8 gram mass&C-Merge);

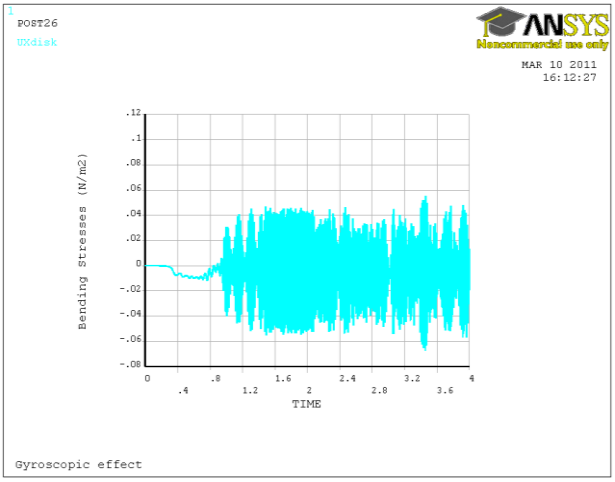
3.2.2 Behaviour of bending stresses with unbalance when add mass

We discover the relation between the bending stress versus time(second),see Fig.(7–A,B),the performance of bending stresses at gyroscopic effect in the middle when add 8 gram mass in the disc at phase angles of $Ø$=90° *and* $Ø$=270° respectively.

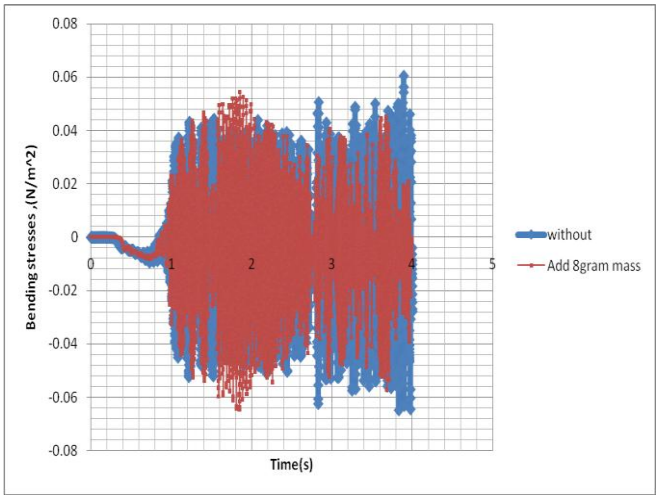
The bending stresse decreases in both direction of motion (Y, Z),see Fig (7-C,D) that mean reduce the reaction force in the bearing to make the bearing long save life.



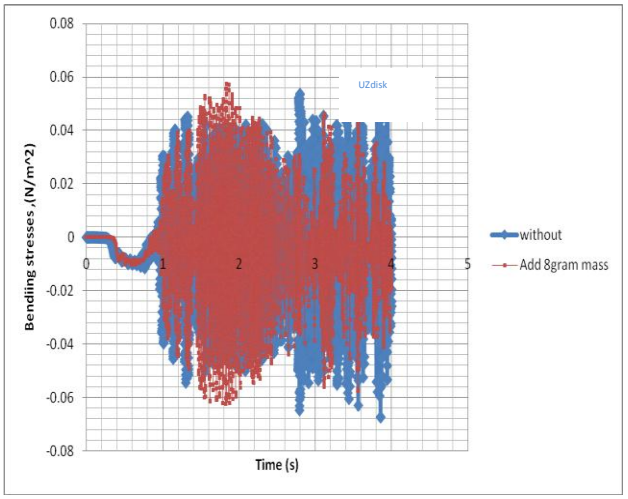
A-Bending stresses in Y direction before add mass.



B-Bending stresses in Y direction after.

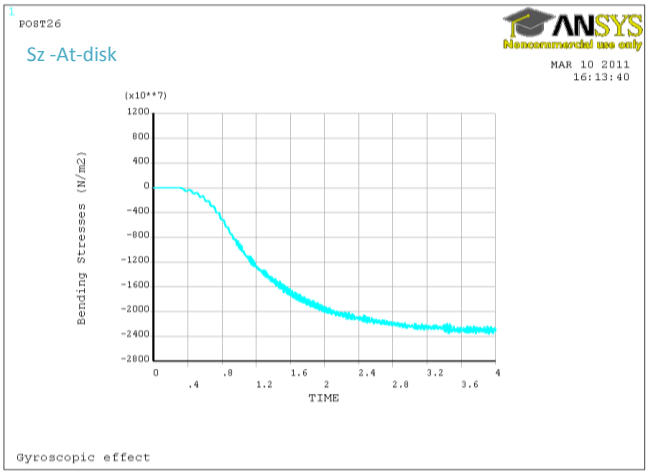


C-Merge in Y direction.

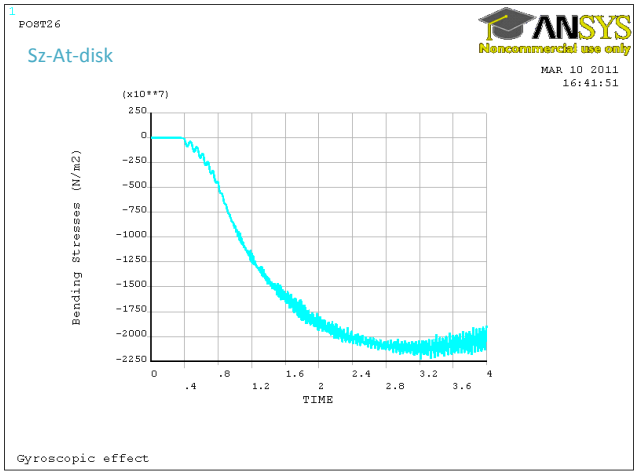


D-Merge in Z direction.

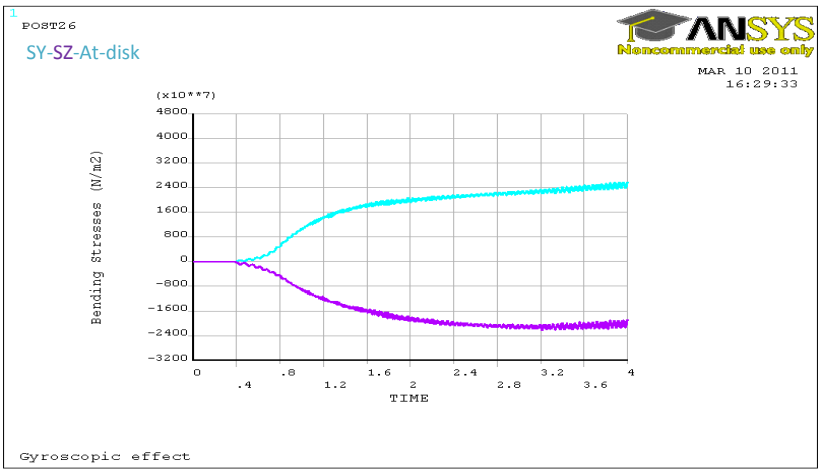
Fig.7 Relation between the bending stresses versus time (sec.);



A-Sz before add the mass .



B-Sz after add 8 gram mass.



C- Sy –Sz at disc.
Fig.8 Bending stresses sample in Y and Z direction (gyroscopic effect);

3.3 Discover damping ratio(ζ) from modal analysis

We discover the damping ratio (ζ) for different mode shape by cur fitting [11,18&19], (multi degree of freedom system) in experimental part, (Table 3) and see Fig.(9).

Table (3)
Natural frequency and damping ratio (ζ) for gyroscopic effect rang (0-500) Hz,(experimental part).

Name	Natural Frequency (Hz)	Damping Ratio(ζ) %	Modal A[kg/s]
Mode1	15.137	75.773	1.387959e-04 +i6.447278e-05
Mode 2	216.51	26.637	0.000103579 +i2.700067e-5

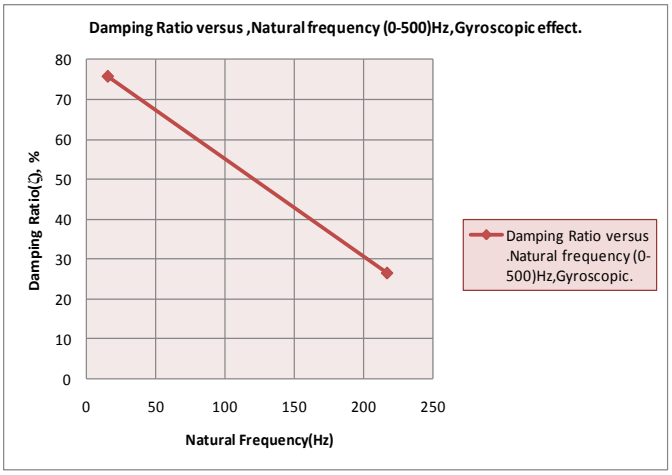


Fig. 9 Damping ratio(ζ) versus natural frequency (0-500)Hz,gyroscopic effect.

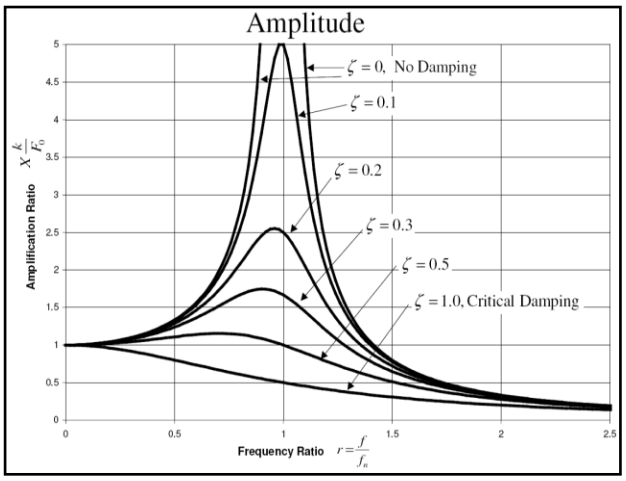


Fig 10.Variation of amplification ratio with r [16,20].

3.4 System identification and vibration monitoring in gyroscopic effect

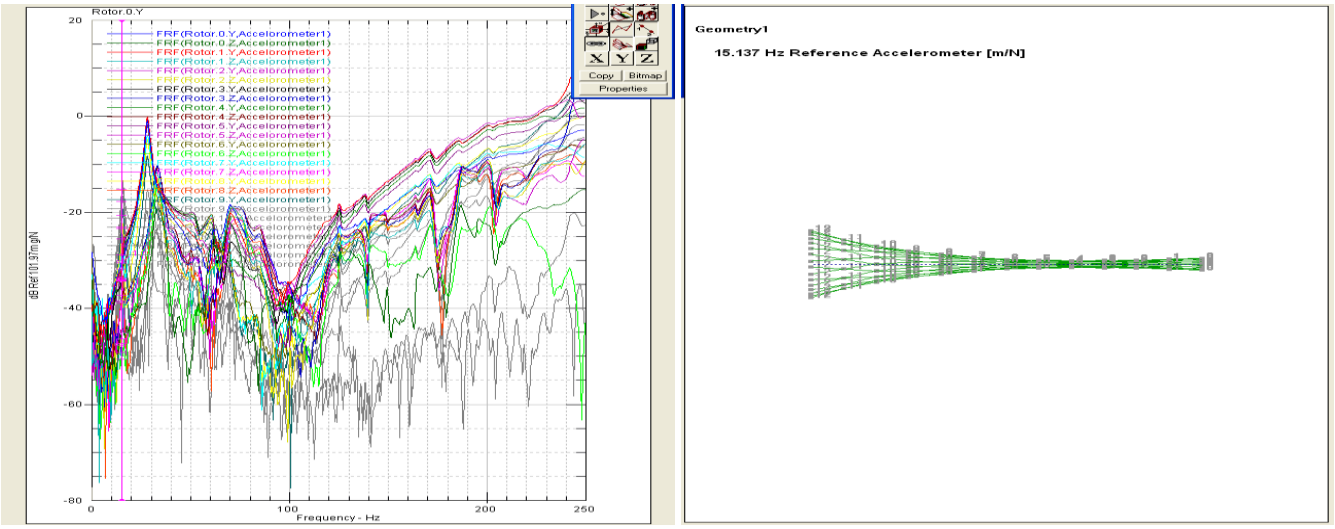


Fig. 11 Gyroscopic effect,(FRF) versus frequency (Hz),(first mode shape).Natural frequency 15.137 Hz.

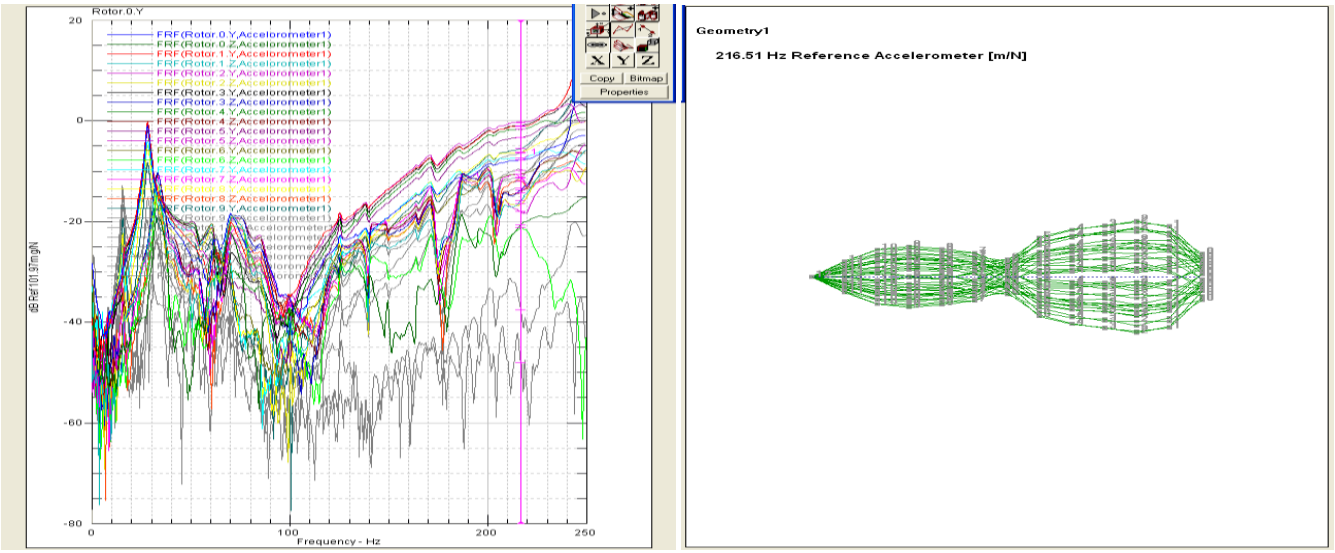


Fig.12 Gyroscopic effect,(FRF) versus Frequency (Hz),(second mode shape).Natural frequency 216.51Hz.

3.5 Contrast measured and predicted natural frequencies for gyroscope

All the result nearby each other between the experimental and simulation (ANSYS) for gyroscope without increasing the speed, see the result in (Table 4) and Fig. (14) for contrast.

Table (4)
Contrast between natural frequency (Hz),outcomes from experiment&ANSYS,(gyroscopic effect) at speed 30 rpm.

Mode Shape	ω_n (ANSYS)Gyroscopic (Hz)	Frequency Gyroscopic Experiment(Hz)	Error %
1	15.703	15.137	1.158007973
2	216.8	216.51	-0.133943005

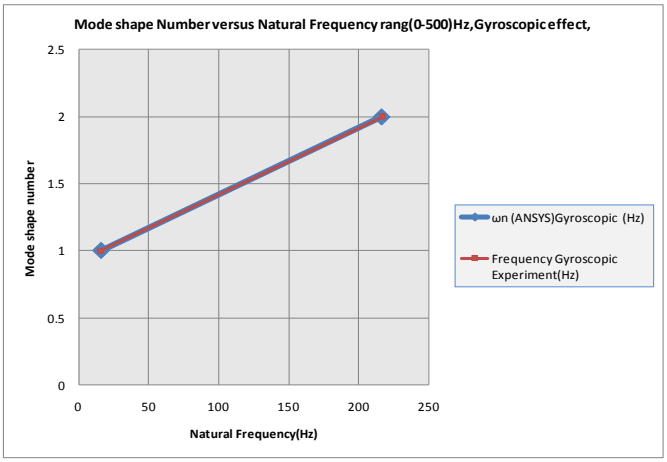


Fig. 13 Mode shape number versus natural frequency experiment and ANSYS,(gyroscopic effect).

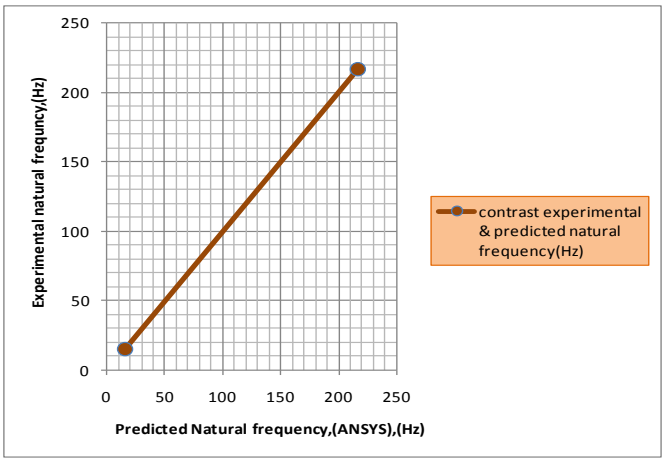


Fig. 14 Natural frequency(experiment versus ANSYS), (gyroscopic effect).

IV. DISCUSSION AND CONCLUSION

In this paper investigate the behaviour of bearing rotor system with gyroscopic effect has been cared out ,a simple mathematical model has been used, however more elaborate models based on a much large degree of freedom may be used based on suppleness or stiffness influence coefficients. The mathematical models may also be used to refine the measured data and help in removal of contaminated data. It is therefore feasible to create a mathematical model as a database for various systems for condition monitoring during their life time of the machines.

For further studies, there is no need to make more experiments about this study while ANSYS gives accurate results. We used (ANSYS) to find the relation between the reaction bearing forces (*N*) with respect time can we see Fig.(5-A,B,C&D).This performance in the right and left bearings with different speed we see when increasing speed of rotation the reaction force increasing for both right and left bearings when increasing the speed of rotation but from the figure above we see the maximum reaction force in Y direction in left and right bearings when the motor run up, after a few second is become decreases. While the reaction forces in Z direction is began increasing slowly in left and right bearings until reach maximum value when the speed is increasing.That mean we must take care to left bearing when run up the motor because this bearing carry maximum reaction force at the began.During study this performance of reaction force in both bearings can aid in the design of low-noise rotor-bearing systems and reduce the reaction force in the bearing to make the bearings long save life by lubracation.In order to investigate the effects of design parameters on the noise of rotor-bearing systems, the effects of radial clearance and width of bearing, lubricant viscosity, for various rotational speeds. It is found that, as a general rule, the noise of the bearing decreases as the lubrication viscosity increases, the width of the bearing increases, and the radial clearance of the bearing decreases.

The locations of the adding balance masses in suppressing the vibration amplitudes are decided to be at the optimum phase angles of $\emptyset=90^{\circ}$ *and* $\emptyset=270^{\circ}$ respectively.It was observed for each of the different eccentricity ratios studies. The critical adding mass ratios can also be predicted through its linear relationship with the eccentricity ratios,The simulation values obtained from the ANSYS see Fig.(6), this results showed that could reduce the vibration by reducing amplitude when add 8 gram mass in the angles show above; As a result, can reduce the vibration more effectively and modified method described in this paper to solve real-world engineering problems.

We discover the relation between the bending stress versus time(second),see Fig. (7) the behavior of bending stresses at gyroscopic effect when added 8 gram mass in the disc the bending stresse decreases in both direction of motion (Y,Z) that mean reduce the reaction force in the bearing to make the bearing long and save life .

From Table (3) detection damping ration (ζ) in experimental part for the first and second mode at speed 30 rpm, and we can see from Fig (9) the decreased the damping ratio caused increased natural frequency until reach maximum amplitude when the system reach resonance $\boldsymbol{\omega} = \boldsymbol{\omega}_{n,}$ when damping ration (ζ) approximately = 0),(free vibration) is clear in Fig.(10)[16,21].

From Table(4),contrast measured and predicted natural frequencies for gyroscopic effect all the outcome nearby each other between the experimental shown in Fig.(11),(12) and model simulation (ANSYS) shown in Fig.(3),(4)for gyroscopic effect without rising the speed,see the result in (Table 4) and is more clear in Fig. (13)&(14) for contrast. Plotting the experimental value against the predicted on for each of the modes included in the contrast shown in Fig.(14).In this way it is possible to see not only the degree of correlation between the two sets of results, but also the nature(and possible case) of any discrepancies which do exist.The points plotted should lie on or close to straight line of slope [17,22].

4.1 Summaries what have learned

A gyroscope is a device that can be used to maintain orientation based on the principles of angular momentum.
As a general rule, the noise of the bearing decreases as the lubrication viscosity increases, the width of the bearing increases, and the radial clearance of the bearing decreases.

ACKNOWLEDGMENTS

The authors are deeply appreciative support derived from the Iraqi Ministry of Higher Education, Iraqi cultural attaché in London and Kingston University London for supporting this research.

REFERENCES

[1] F Choy, M Braun and Y Hu, Nonlinear transient and frequency response analysis of a hydrodynamic journal bearing, *ASME Journal Tribol,114*, 1992,48–54.
[2] D Yoon, O Kwon, M Jung and K Kim, Early detection of damages in journal bearings by AE monitoring. *Journal of Acoustic Amiss,13(1)*, 1995,1–10.
[3] B Rho and K Kim, Acoustical properties of hydrodynamic journal bearings, *Tribol,Int,36(1)*, 2003,61–6.
[4] G Ferraris , V Maisonneuve and M Lalanne, Prediction of the dynamic behaviour of non symmetric coaxial co or counter – rotating rotor, *Journal Sound and Vibration,195(4)*,1996,776-788.
[5] http://www.gyroscopes.org/how.asp
[6] B Rho, D Kim and K Kim, Noise analysis of oil lubricated journal bearings.,*I MechE Part C J Mech Eng Sci,217(3)*, 2003, 65–71.
[7] L.L. Beranek, I.L.Ver,*Noise and vibration control engineering*(John Wiley&Sons,1992).
[8] J .Schmied,. et.al., Application of madyn 2000 to rotordynamic problems in industrial machinery, *Proc. GT2007*, Montreal, Canada, May (2007).
[9] M Spirig, J Schmied, P Jenckel ,U Kanne, Three practical examples of magnetic bearing control design using a modern tool, *ASME Journal of Engineering for Gas Turbines and Power,124*, October 2002,1025-1031.
[10] H. D Nelson, Mc Vaugh, J. M.,The dynamics of rotor-bearing systems using finite element, *ASME Trans.,Journal of Industry,98(2)*,1976,593-600.
[11] D.Formenti, M. H. Richardson, Global curve fitting of frequency response measurements using the rational, *International Modal Analysis Conference*, 1985,390–397.
[12] H D Nelson, A Finite Rotating Shaft Element using Timoshenko Beam Theory, *Trans. ASME, Journal of Mechanical Design,102(4)*,1980,793-803.
[13] ANSYS 12 Help Menu (*can be found with ANSYS 12*).
[14] http://www.mpihome.com/
[15] E. Swanson, C. D. Powll, and S.Weismann, *A practical review of rotating machinery critical speeds and modes* (1991).
[16] S .S .Rao, *Mechanical vibrations* (Prentice-Hall: Inc. Second Edition, Singapore,2005).
[17] D.J. EWINS, *Modal testing: theory and practice* (John Wily: Exeter, England,1995).
[18] J . Rybczynski, Maps of vibrational symptoms of bearing misalignment defects in large power turbo set, *Proceedings of ASME Turbo Expo*, Berlin, Germany, June (2008).
[19] M. Lalanne and G. Ferraris, *Rotor dynamics prediction in engineering* (2nd edition, publishing by John Wiley&Sons Ltd, England,1998).
[20] J. H .Ginsberg, *Mechanical and structural vibratio, theory and applications*(Georgia Institute of Technology, John Wily & Sons: Inc.United States, 2001).
[21] D Childs, *Turbomachinery rotordynamics* (Wiley: Inter Science, 1993).
[22] J. Schmied, F. Betschon, Engineering for rotors supported on magnetic bearings, *Proceedings of the 6th International Symposium on Magnetic Bearings*, August 5-7, Boston,1998.

AUTHORS PROFILE

[1]**Mr.Hisham.A.H.AL-Khazali**, He has PhD Student in Kingston University London/UK. He was born in 28 Aug 1973 Baghdad/Iraq. Received his BSc(Eng) in Mechanical Engineering (1996), University of Technology, Baghdad. MSc in Applied Mechanics, University of Technology, Baghdad (2000).

E-mail, k0903888@kingston.ac.uk

[2]**Dr.Mohamad.R.Askari**, BSc(Eng), MSc, PhD, CEng, MIMechE, MRAeS. He has (Principal Lecturer, Blended Learning Coordinator),Member teaching staff in Kingston University London/UK, His Teaching Area: Applied Mechanics, Aerospace Dynamics, Dynamics and Control, Structural and Flight Dynamics, Engineering Design, Software Engineering to BEng Mechanical and Aerospace second and final years.Year Tutor for BEng Mechanical Engineering Course and School Safety Advisor.

E-mail, M.Askari@Kingston.ac.uk

www.ijmer.com Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645

Survey on Techniques for Plant Leaf Classification

Prof. Meeta Kumar[1], Mrunali Kamble[2], Shubhada Pawar[3], Prajakta Patil[4],Neha Bonde[5]

* (Department of Computer Engineering, MIT College of Engineering/ Pune University, India)
** (Department of Computer Engineering, MIT College of Engineering/ Pune University, India)
*** (Department of Computer Engineering, MIT College of Engineering/ Pune University, India)
**** (Department of Computer Engineering, MIT College of Engineering/ Pune University, India)
***** (Department of Computer Engineering, MIT College of Engineering/ Pune University, India

ABSTRACT

In this paper we present survey on various classification techniques which can be used for plant leaf classification. A classification problem deals with associating a given input pattern with one of the distinct classes. Plant leaf classification is a technique where leaf is classified based on its different morphological features. There are various successful classification techniques like k-Nearest Neighbor Classifier, Probabilistic Neural Network, Genetic Algorithm, Support Vector Machine, and Principal Component Analalysis. Deciding on the method for classification is often a difficult task because the quality of the results can be different for different input data. Plant leaf classifications has wide applications in various fields such as botany, Ayurveda, Agriculture etc. The goal of this survey is to provide an overview of different classification techniques for plant leaf classification.

Keywords - **Leaf classification, image preprocessing, classifier, k-Nearest Neighbor, SVM**

I. INTRODUCTION

Plant recognition or classification has a broad application prospective in agriculture and medicine, and is especially significant to the biology diversity research. Plant leaf classification finds application in botany and in tea, cotton and other industries. Plants are vitally important for environmental protection. However, it is an important and difficult task to recognize plant species on earth. Many of them carry significant information for the development of human society. The urgent situation is that many plants are at the risk of extinction. So it is very necessary to set up a database for plant protection. We believe that the first step is to teach a computer how to classify plants.

Leaf recognition plays an important role in plant classification. Plants are basically identified based on flowers and fruits. However these are three dimensional objects and increases complexity. Plant identification based on flowers and fruits require morphological features such as number of stamens in flower and number of ovaries in fruits. Identifying plants using such keys is a very time consuming task and has been carried out only by trained botanists. However, in addition to this time intensive task, there are several other drawbacks in identifying plants using these features such as the unavailability of required morphological information and use of botanical terms that only experts can understand. However leaves also play an important role in plant identification. Moreover, leaves can be easily found and collected everywhere at all seasons, while flowers can only be obtained at blooming season. Shape of plant leaves is one of the most important features for characterising various plants visually. Plant leaves have two-dimensional nature and thus they are most suitable for machine processing.

Our paper presents survey of different classification techniques. Before classification can be done on basis of leaf some preprocessing is needed. And most important step prior classification is feature extraction. For classification different techniques are available. Some of them are k-Nearest Neighbor Classifier, Neural Network, Genetic Algorithm, Support Vector Machine, and Principal Component Analysis. In section II we will discuss preprocessing to be performed on the acquired image .In section III we have discussed overview of classification techniques and how they can be used for recognition of a species of a plant leaf. Finally in section V we

www.ijmer.com Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645

conclude and discuss the future work that can be done. Table 1 shows comparative study for classification techniques we studied through this survey.

II. LEAF IMAGE ACQUISITION AND PREPROCESSING

First step for plant leaf classification is image acquisition. Image acquisition includes plucking leaf from plant and then, the digital color image of the leaf is taken with a digital camera. After leaf image is obtained some pre-processing is needed. This stage includes grayscale conversion, image segmentation, binary conversion and image smoothing. The aim of image pre-processing is to improve image data so that it can suppress undesired distortions and enhances the image features that are relevant for further processing. Color image of leaf is converted to grayscale image. Variety of changes in atmosphere and season cause the color feature having low reliability. Thus it is better to work with grayscale image. Once image is converted to grayscale it is segmented from its background and then converted to binary. Using one of the edge detectors its contour is detected. Then certain morphological features are extracted from its contour image. This feature vector is then provided to the classifier. Fig. 1 gives block diagram for plant leaf classification process.

III. CLASSIFICATION TECHNIQUES

A classification problem deals with associating a given input pattern with one of the distinct classes. Patterns are specified by a number of features (representing some measurements made on the objects that are being classified) so it is natural to think of them as d-dimensional vectors, where d is the number of different features. This representation gives rise to a concept of feature space. Patterns are points in this d-dimensional space and classes are sub-spaces. A classifier assigns one class to each point of the input space. The problem of classification basically establishes a transformation between the features and the classes. The optimal classifier is the one expected to produce the least number of misclassifications

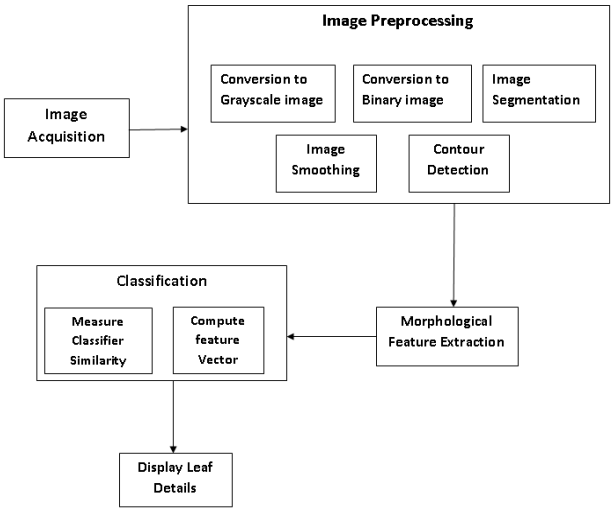


Figure 1 Block diagram for Plant Leaf classification

2.1 k-NEAREST NEIGHBOUR CLASSIFIERS

K Nearest Neighbor classifier calculates the minimum distance of a given point with other points to determine its class. Suppose we have some training objects whose attribute vectors are given and some unknown object w is to be categorized. Now we should decide to which class object w belongs.

Let us take an example. According to the k-NN rule suppose we first select k = 5 neighbors of w. Because three of these five neighbors belong to class 2 and two of them to class 3, the object w should belong to class 2, according to the k-NN rule. It is intuitive that the k-NN rule doesn't take the fact that different neighbors may give different evidences into consideration. Actually, it is reasonable to assume that objects which are close together (according to some appropriate metric) will belong to the same category. According to the k-NN rule suppose we first select k = 5 neighbors of w. Because three of these five neighbors belong to class 2 and two of them to class 3, the object w should belong to class 2, according to the k-NN rule.

www.ijmer.com Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645

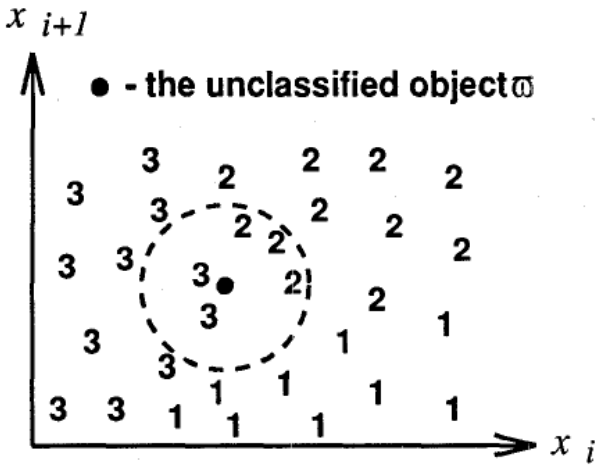


Figure 2 Example for classification using k-NN rule

For plant leaf classification, we first find out feature vector of test sample and then calculate Euclidean distance between test sample and training sample. This way it finds out similarity measures and accordingly finds out class for test sample. The k-nearest neighbor's algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors. k is a positive integer, typically small. If k = 1, then the object is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose k to be an odd number as this avoids tied votes. It is intuitive that the k-NN rule doesn't take the fact that different neighbors may give different evidences into consideration. Actually, it is reasonable to assume that objects which are close together (according to some appropriate metric) will belong to the same category.

2.2 PROBABILISTIC NEURAL NETWORK

Probabilistic neural networks can be used for classification problems. It has parallel distributed processor that has a natural tendency for storing experiential knowledge. PNN is derived from Radial Basis Function (RBF) Network. PNN basically works with 3 layers. First layer is input layer. The input layer accepts an input vector. When an input is presented, first layer computes distances from the input vector to the training input vectors and produces a vector whose elements indicate how close the input is to a training input [3]. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Radial Basis Layer evaluates vector distances between input vector and row weight vectors in weight matrix. These distances are scaled by Radial Basis Function nonlinearly [3]. The last layer i.e. competitive layer in PNN structure produces a classification decision, in which a class with maximum probabilities will be assigned by 1 and other classes will be assigned by 0.A key benefit of neural networks is that a model of the system can be built from the available data. Fig.3 shows architecture of PNN.

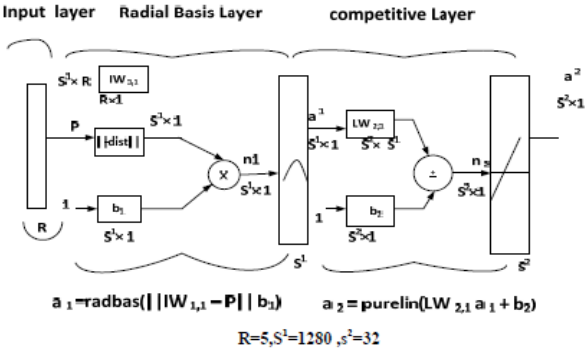


Figure 3 Architecture of PNN

2.3 GENETIC ALGORITHM

Genetic Algorithms are mainly used for feature classification and feature selection. The basic purpose of genetic algorithms (GAs) is optimization .GAs give a heuristic way of searching the input space for optimal x that approximates brute force without enumerating all the elements and therefore bypasses performance issues specific to exhaustive search. Genetic algorithm is used effectively in the evolution to find a near-optimal set of connection weights globally without computing gradient information and without weight connections initialization [1]. Though solution found by genetic algorithms is not always best solution. It finds "good" solution always. Main advantage of GA is that is adaptable and it possess inherent parallelism. Genetic Algorithms handle large, complex, non differentiable and multi model

www.ijmer.com Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645

spaces for image classification and many other real world applications.

2.4 SUPPORT VECTOR MACHINE

Support vector machine (SVM) is a non-linear classifier. The idea behind the method is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification performance. Support Vector Machine is a machine learning tool and has emerged as a powerful technique for learning from data and in particular for solving binary classification problems [3]. The main concepts of SVM are to first transform input data into a higher dimensional space by means of a kernel function and then construct an OSH (Optimal Separating Hyper Plane) between the two classes in the transformed space [3]. For plant leaf classification it will transform feature vector extracted from leaf's contour. SVM finds the OSH by maximizing the margin between the classes. Data vectors nearest to the constructed line in the transformed space are called the support vectors. The SVM estimates a function for classifying data into two classes. Using a nonlinear transformation that depends on a regularization parameter, the input vectors are placed into a high-dimensional feature space, where a linear separation is employed. To construct a nonlinear support vector classifier, the inner product (x, y) is replaced by a kernel function K (x, y), as in (1)

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right) \quad \text{…….. (1)}$$

where f(x) determines the membership of x. We assume normal subjects were labeled as -1 and other subjects as +1.The SVM has two layers [4]. During the learning process, the first layer selects the basis K (xi, x), i=1, 2….N from the given set of kernels, while the second layer constructs a linear function in the space. This is equivalent to finding the optimal hyper plane in the corresponding feature space. The SVM algorithm can construct a variety of learning machines using different kernel functions. Fig 4 shows the linear separating hyper plane where support vector are encircled.

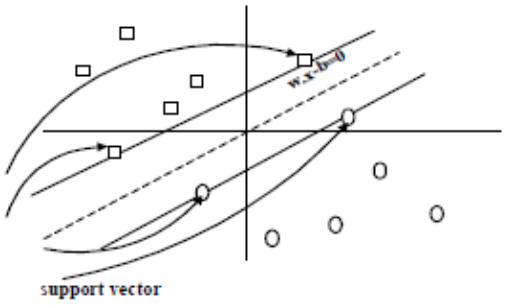


Figure 4 Linear separating hyper planes, the support vectors are circled

Main advantage of SVM is it has a simple geometric interpretation and gives a sparse solution. Unlike neural networks, the computational complexity of SVMs does not depend on the dimensionality of the input space One of the bottlenecks of the SVM is the large number of support vectors used from the training set to perform classification tasks.

2.5 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis is a variable reduction procedure. It is useful when you have obtained data on a number of variables (possibly a large number of variables), and believe that there is some redundancy in those variables. Intuitively, Principal components analysis is a method of extracting information from a higher dimensional data by projecting it to a lower dimension.

Principal component analysis is a basically used because it reduces the dimension of input vector of neural network. This method generates a new set of variables, called principal components. Each principal component is a linear combination of the optimally-weighted observed variables. All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data. Mathematically, PCA transforms the data to a new coordinate system such that the greatest variance by any projection of the

www.ijmer.com Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645

data comes to lie on the first coordinate, the second greatest variance on the second coordinate, and so on [2]. Each coordinate is called a principal component.

Often the variability of the data can be captured by a relatively small number of principal components, and, as a result, PCA can achieve high dimensionality reduction with usually lower noise than the original patterns. The objective of PCA is to perform dimensionality reduction while preserving as much of the randomness in the high-dimensional space as possible But the limitation with PCA is it depends on scaling of variables and it is not always easy to interpret principal components. The main limitation of PCA is that it does not consider class separability since it does not take into account the class label of the feature vector.

IV. CONCLUSION & FUTURE WORK

From study of above classification techniques we come up with following conclusion. The nearest-neighbor method is perhaps the simplest of all algorithms for predicting the class of a test example. An obvious disadvantage of the kNN method is the time complexity of making predictions. Considerable amount of work has been done for recognizing plant species using k Nearest Neighbor technique. Classifying using PNN and SVM can further be explored by researchers, SVM being relatively a new machine learning tool. The most important advantage of PNN is that training is easy and instantaneous.

Additionally, neural networks are tolerant to noisy inputs. But in neural network it's difficult to understand structure of algorithm. SVM was found competitive with the best available machine learning algorithms in classifying high-dimensional data sets. In SVM computational complexity is reduced to quadratic optimization problem and it's easy to control complexity of decision rule and frequency of error. Drawback of SVM is it's difficult to determine optimal parameters when training data is not linearly separable. Also SVM is more complex to understand and implement. Another technique we studying is genetic algorithm. Genetic algorithms are good at refining irrelevant and noisy features selected for classification. But representation of training/output data in genetic programming is complicated. Genetic algorithms provide a comprehensive search methodology for machine learning and optimization.

PCA is used because it has advantage of reduced vector. The main limitation of PCA is that it does not consider class separability since it does not take into account the class label of the feature vector.

Future direction for researchers can be to explore more robust techniques for recognition of plant leaves using a combination of classifying techniques like SVM, kNN, PNN.

Mobile applications for plant leaf classification can be created which can be best learning tool for botany students. Also this application can be used in agricultural field for weed identification which in turn will help for proper determination of pesticides and fertilizers.

www.ijmer.com Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645

COMPARATIVE STUDY

TABLE 1 Comparative Study of classification techniques for plant leaf classification

Classification Techniques	Pros	Cons
1. kNN Classifier	1. Simplest 2.Robust with regard to search space 3.No training is required, confidence level can be obtained	1. Expensive testing of each instance 2. Sensitiveness to noisy or irrelevant inputs 3.Lazy Learning
2. Probabilistic Neural Network	1. Tolerant of noisy inputs 2. Instances can be classified by more than one output 3. Adaptive to changing data	1. Long training time 2. Large complexity of network structure 3. too many attributes can result in over fitting
3.Genetic Algorithm	1. Handle large, complex, non differentiable and multi model spaces 2. Refining irrelevant and noise genes 3. Efficient search method for a complex problem space	1. Computation or development of scoring function is nontrivial 2. Not the most efficient method to find some optima, rather than global 3. Complications involved in the representation of training/output data
4. Support Vector Machine	1. Good generalization capability 2. Sparseness of the solution and the capacity control obtained by optimizing the margin 3. SVMs can be robust, even when the training sample has some bias	1. Slow training 2. Difficult to understand structure of algorithm 3. limitation is speed and size, both in training and testing
5.Principal Component Analysis	1. Used for variable reductions 2. Choose weights depending on the frequency in frequency domain. 3. Extract the maximum information in the data by maximizing the variance of the principal components.	1. Does not perform linear separation of classes 2. Scaling of variables 3. The largest variances do not correspond to the meaningful axes

REFERENCES

Journal Papers:

[1] M.Seetha, I.V.muralikrishna, B.L. Deekshatulu, B.L.malleswari, Nagaratna, P.Hegde a *Artificial neural networks and other methods of image classification*, *Journal of Theoretical and Applied Information Technology*, © 2005 - 2008 JATIT. All rights reserved.

[2] Krishna Singh, Indra Gupta, Sangeeta Gupta, *SVM-BDT PNN and Fourier Moment Technique for classification of Leaf, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 3, No.* 4, December, 2010.

[3] Krishna Singh, Dr. Indra Gupta and Dr Sangeeta Gupta , *Retrieval and classification of leaf shape by support vector machine using binary decision tree, probabilistic neural network and generic Fourier moment technique: a comparative study*, *IADIS International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing 2010*

[4] J.-X. Du, X.-F. Wang, and G.-J. Zhang, *Leaf shape based plant species recognition,Applied Mathematics and Computation, vol. 185*, 2007.

[5] A. Kadir, L. E. Nugroho, A. Susanto, P. Insap Santosa, "*Leaf Classification Using Shape, Color, and Texture Features*", *International Journal of Computer Trends and Technology- July to Aug Issue 2011*

[6] Hongjun Lum, Rudy Setiono, *Effective Data Mining using Neural Network, IEEE transactions on knowledge and data engineering, vol. 8, no. 6*, december 1996

[7] J. M. Zurada, *Introduction to Artificial Neural Networks System*. Jaico Publishing House.

[8] J.-X. Du, X.-F. Wang, and G.-J. Zhang, *Leaf shape based plant species recognition*,Applied Mathematics and Computation, vol. 185, 2007.

[9] D.E Goldberg, *Genetic Algorithms in Search, Optimization* and Machine Learning. Addison-Wesley, New York, 1989.

[10] M. Z. Rashad, B.S.el-Desouky, Manal S .Khawasik, *Plants Images Classification Based on Textural Features using Combined Classifier, International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4, August 2011*

[11] D S Guru, Y. H. Sharath, S. Manjunath, *Texture Features and kNN in classification of Flower Images, IJCA special issue on" Recent Trends in Image Processing and Pattern Recognition" RTIPPR, 2010*

[12] Mahmood R Golzarian, Ross A Frick, *Classification of images of wheat, ryegrass and brome grass species at early growth stages using principal component analysis*, Golzarian and Frick Plant Methods 2011, 7:28 http://www.plantmethods.com/content/7/1/28

Proceedings Papers:

[13] Thair Nu Phyu, Survey of Classification Techniques in Data Mining, *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009*, Mar 18 - 20, 2009, Hong Kong.

[14] S. Wu, F. Bao, E. Xu, Y. Wang, Y. Chang, and Q. Xiang, A leaf recognition algorithm for plant classification using probabilistic neural network, in *Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Technology*, Giza, Dec 2007.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue.2, pp-545-551 ISSN: 2249-6645

PV Cell Based Five Level Inverter Using Multicarrier PWM

K.Surya Suresh[1] and M.Vishnu Prasad [1]
[1] Sri Vasavi Institute of Engineering and Technology, EEE Department, Nandamuru, AP, India

Abstract

This paper investigates the performance of a PV cell connected Multi Level Inverter topology. These MLI's are suitable in high voltage & high power application due to their ability to synthesize waveforms with better harmonic spectrum. This paper presents, a proposed scheme adopting the Multi carrier Pulse width modulation concept. The MCPWM Cascaded Multilevel inverter strategy enhances the fundamental output voltage and reduced Total harmonic distortion. The Multilevel inverter circuit analysis and selection of proper references discussed based on the formulation switching patterns. A Single phase five level cascaded inverter is used to explain the methods. The method can be easily extended to an m-level inverter. The cascaded inverter is subjected to a new modulation scheme, which uses multiple modulating signals with a single carrier. In order to justify the merits of the proposed modulation scheme, harmonic analysis for and measured THD and output voltages are compared and discussed

Key Words: Multilevel inverter, Multicarrier Pulse width modulation, Total harmonic distortion, PV Cell, Switching frequency optimal PWM, Sub harmonic PWM modulation index

I INTRODUCTION

Photovoltaic (PV) power generation is very desirable since it is renewable and does not contribute to pollution or Global climate change. PV is especially attractive for - applications in where sunshine is available for most of the time. This paper presents a PV array connected to Cascaded H-Bridge type multi-level inverter to achieve sinusoidal voltage waveform and output sinusoidal current to the utility grid with a simple and cost effective power electronic solution. The topologies of multilevel inverters are classified in to three types the Flying capacitor inverter, the Diode clamped inverter and the Cascaded bridge inverter [1][2]. The proposed scheme of multilevel inverter is the multi carrier sub-harmonic pulse width modulation (MC-SH PWM) [4][5]. The
MC-SH PWM cascaded multilevel inverter strategy reduced total harmonic [6]

II MATHEMATICAL MODEL OF THE PV ARRAY

2.1 Simplified Equivalent Circuit

A solar cell basically is a p-n semiconductor junction. When exposed to light, a current proportional to solar irradiance is generated. The circuit model of PV cell is illustrated in Fig. 1. Standard simulation tools utilize the approximate diode equivalent circuit shown in Fig. 2 in order to simulate all electric circuits that contain diodes. The model is based on two-segment piecewise linear approximation. The circuit consists of R_{on} in series with voltage source V_{on}

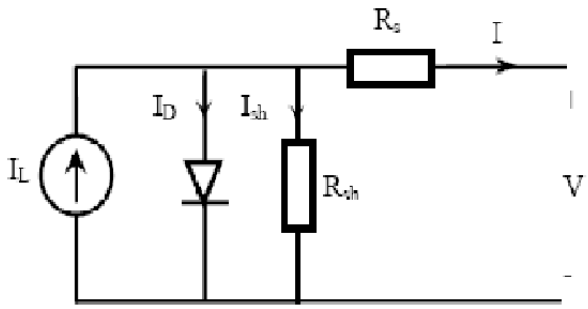


Fig. 1 Circuit model of PV solar cell

2.2. Theoretical Mathematical Model:

The equation [7] that describe I-V characteristics of the solar cell based on simple equivalent circuit shown in Fig. 1, are given below;

$$I_D = I_O \left[e^{\frac{q(V+IRs}{KT}} - 1\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$I = I_L - I_O \left[e^{\frac{q(V+IRs}{KT}} - 1\right] - \frac{V+IRs}{Rsh} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where:

I is the cell current (A).
q is the charge of electron = 1.6x10-19 (coul).
K is the Boltzman constant (j/K).
T is the cell temperature (K).
IL is the light generated current (A).
Io is the diode saturation current.
Rs , Rsh are cell series and shunt resistance (ohms).
V is the cell output voltage (V).

2.3 PV Characteristics:

2.3.1 Current Vs Voltage Characteristics:

Equation (1) was used in computer simulation to obtain the output characteristics of a solar cell, as shown in the figure 2.This curve clearly shows that the output characteristics of a solar cell are non linear and are crucially influenced by solar radiation, temperature and load condition

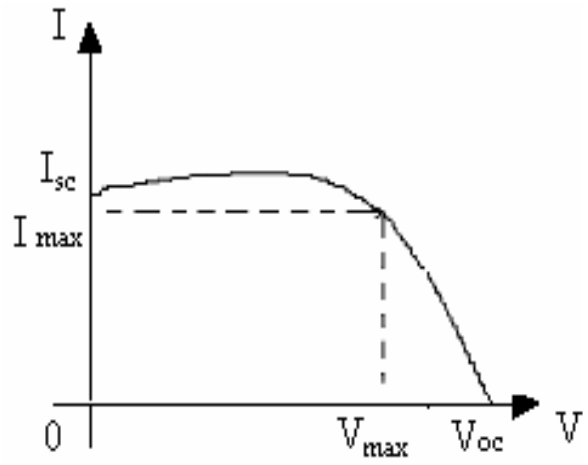


Fig 2: Output Characteristics Of Solar Cell

2.3.2 Power Vs Voltage Characteristics:

Figure 3 shows the typical Power versus Voltage curve of the PV array. In this figure, P is the power extracted from PV array and V is the voltage across the terminals of the PV array [2].

The characteristics have different slopes at various points.When maximum power is extracted from PV array the system is operating at MPP where slope is zero. The PV curve varies according to the current insolation and temperature. When insolation increases, the power available from PV array increases whereas when temperature increases, the power available from PV Array decreases.

2.3.3 Variation in Available Energy due to Sun's Incident Angle:

PV cell output with respect to sun's angle of incidence is approximated by a cosines function at sun angles from 0° to 50° .Beyond the incident angle of 50° the available solar energy falls of rapidly as shown in the figure 4. Therefore it is convenient and sufficient within the normal operating range to model the fluctuations in photocurrent (Iph) verses incident angle is given by Eq(3). [8].

$$I_{ph} = I_{max} \cos\theta \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

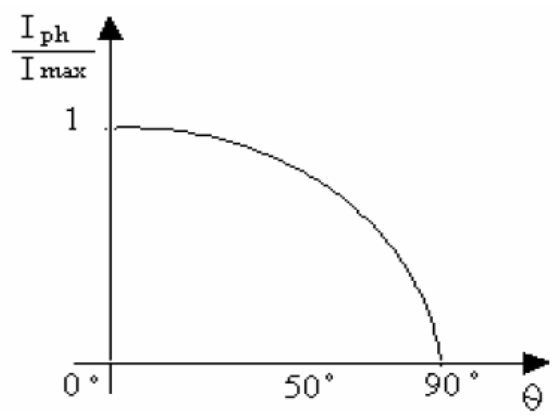


Fig 3: Power Vs Voltage

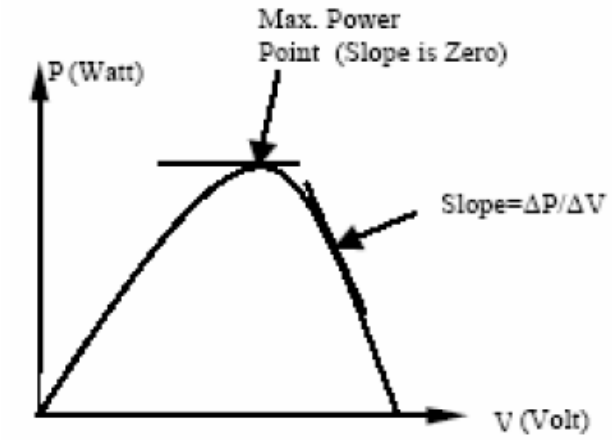


Fig 4: Variation In Available Energy Due Sun's Incident Angle Variation

III CASCADED MULTILEVEL INVERTER

A single-phase structure of an m-level cascaded inverter is illustrated in Figure 5. Each separate dc source (SDCS) is connected to a single-phase full-bridge, or H-bridge, inverter. Each inverter level can generate three different voltage outputs, $+V_{dc}$, 0, and $-V_{dc}$ by connecting the dc source to the ac output by different combinations of the four switches, S_1, S_2, S_3, and S_4. To obtain $+V_{dc}$, switches S_1 and S_4 are turned on, whereas $-V_{dc}$ can be obtained by turning on switches S_2 and S_3. By turning on S_1 and S_2 or S_3 and S_4, the output voltage is 0. The ac outputs of each of the different full-bridge inverter levels are connected in series such that the synthesized voltage waveform is the sum of the inverter outputs. The number of output phase voltage levels m in a cascade inverter is defined by m = 2s+1, where s is the number of separate dc sources. An example phase voltage waveform for an 11-level cascaded H-bridge inverter with 5 SDCSs and 5 full bridges is shown in Figure 6. The phase voltage

$V_{AM} = V_{A1} + V_{A2} + V_{A3} + V_{A4} + V_{A5}$

For a stepped waveform such as the one depicted in Figure 6 with s steps, the Fourier Transform for this waveform follows [9, 13]:

$$V(\omega t) = \left(\frac{4VDC}{\pi}\right) \Sigma [\cos(n\theta_1) + \cos(n\theta_2) +$$

where n = 1, 3, 5, 7, ... (4)

From (4), the magnitudes of the Fourier coefficients when normalized with respect to V_{dc} are as follows

$$H(n) = \frac{4}{\pi n} [\cos(n\theta_1) + \cos(n\theta_2) + \cos(n\theta_3) + \ldots\ldots + \cos(n\theta_s)]$$

where n = 1, 3, 5, 7, ... (5)

The conducting angles, θ_1, θ_2..... θ_s, can be chosen such that the voltage total harmonic distortion is a minimum. Generally, these angles are chosen so that predominant lower frequency harmonics, 5th, 7th, 11th, and 13th, harmonics are eliminated [14].

Multilevel cascaded inverters have been proposed for such applications as static var generation, an interface with renewable energy sources, and for battery-based applications. Cascaded inverters are ideal for connecting renewable energy sources with an ac grid, because of the need for separate dc sources, which is the case in applications such as photovoltaic's or fuel cells

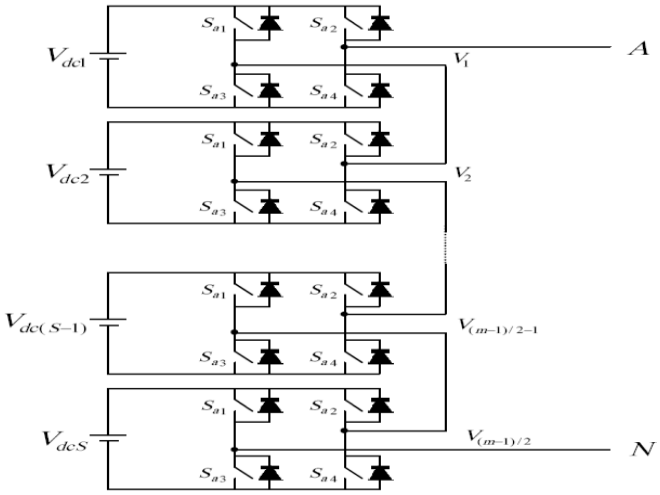


Fig 5 Single-phase structure of a m level multilevel cascaded H bridge inverter

Cascaded inverters have also been proposed for use as the main traction drive in electric vehicles, where several batteries or ultracapacitors are well suited to serve as SDCSs [15, 16]. The cascaded inverter could also serve as a rectifier/charger for the batteries of an electric vehicle while the vehicle was connected to an ac supply.

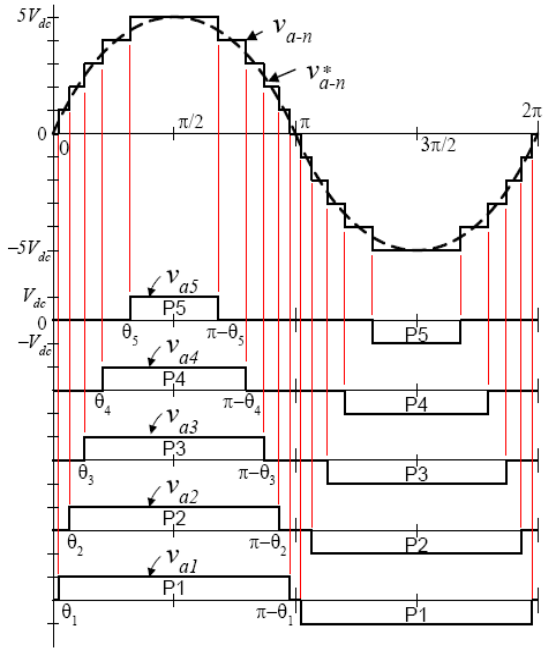


Fig 6. Output phase voltage waveform of an 11-level cascade inverter with 5 separate dc sources

Additionally, the cascade inverter can act as a rectifier in a vehicle that uses regenerative braking

IV MULTICARRIER PULSE WIDTH MODULATION

MULTICARRIER SUB HARMONIC PULSE WIDTH MODULATION (MC-SH PWM)

Fig.7 shows Multicarrier sub harmonic pulse width modulation (MC-SH PWM) modulating signal generation. Fig.4 shows a m-level inverter, m-1 carriers with the same frequency f_c and the same amplitude A_c are disposed such that the bands they occupy are contiguous

The reference wave form has peak to peak amplitude A_m, the frequency f_m, and its zero centered in the middle of the carrier set. The reference is continuously compared with each of the carrier signals. If the reference is greater than s carrier signal, then they active device corresponding to that carrier is switched off.

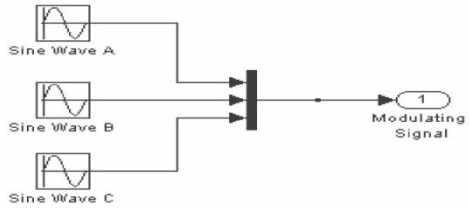


Fig. 7: Multicarrier sub harmonic PWM modulating signal generation

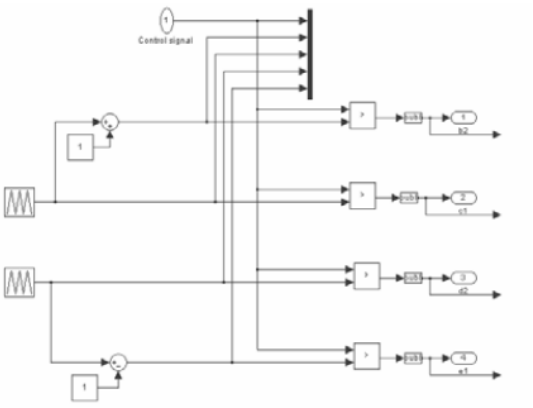


Fig. 8: Multicarrier sub harmonic PWM signal generation

In multilevel inverters, the amplitude modulation index m_a and the frequency ratio m_f are defined as

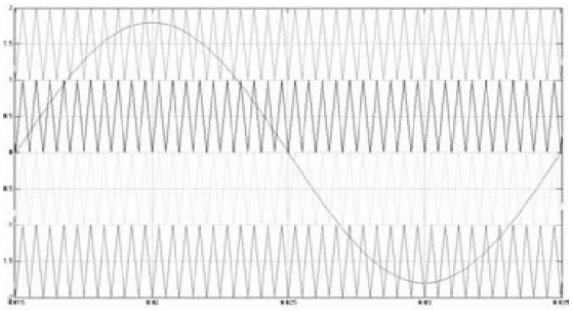


Fig. 9**:** Multicarrier sub harmonic pulse width modulation.

$$m_a = \frac{Am}{(m-1)Ac} \quad \text{..... (6)}$$

$$m_f = fc/fm \quad \text{..... (7)}$$

Using MC-SH PWM THD value can be reduced with reduction in output voltage. In this Paper, When input voltage Vdc = 230V the THD value 8.35% and its output voltage Vac = 9 volts.

V PRAPOSED PROTOTYPE

The Cascaded Multilevel Converters are simply a number of conventional two-level bridges, whose AC terminals are simply connected in series to synthesize the output waveforms. Fig. 10 shows the power circuit for a five-level inverter with two cascaded cells. The Cascaded Multilevel Converters needs several independent DC sources which may be obtained from batteries, fuel cells or solar cells and in this solar cells are used

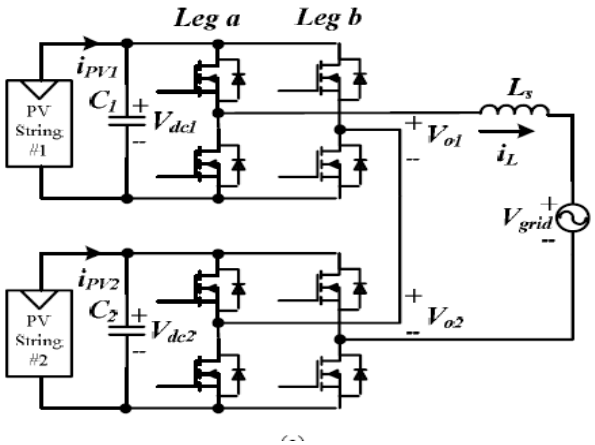


Fig 10 Proposed power circuit for a five-level inverter with two cascaded cells.

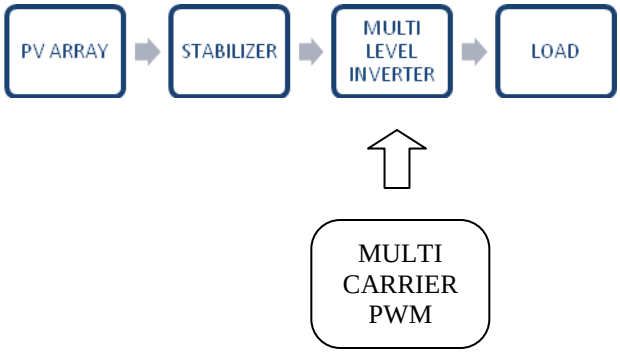
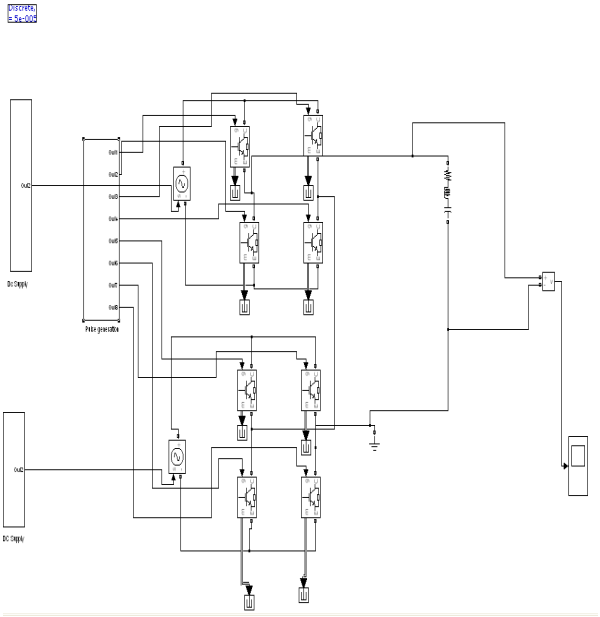


Fig. 11 Block diagram of proposed Multi level inverter

VI SIMULATION RESULTS

Fig 10 shows complete PV generation system based on the proposed multilevel converter has been implemented in a prototype and the proposed block diagram as shown in the figure 11.It is worth mentioning that the output voltage of the PV string arrays should be chosen based on the grid nominal voltage and the minimum desired operating power of each cell. If the power generated by all strings is equal, the output voltage of all cells will be equal. Simulations have been carried out in MATLAB–Simulink to study the performance of the proposed control and modulation scheme. The particular system shown in Fig. 12 is modeled Two PVAs are connected to a passive load through a Five-level cascaded H bridge inverter. Fig 13 shows a PV Array contains six series-



connected 130-V 1000-Wp PV panels.

Fig 12 Five Level Inverter with PV Cell

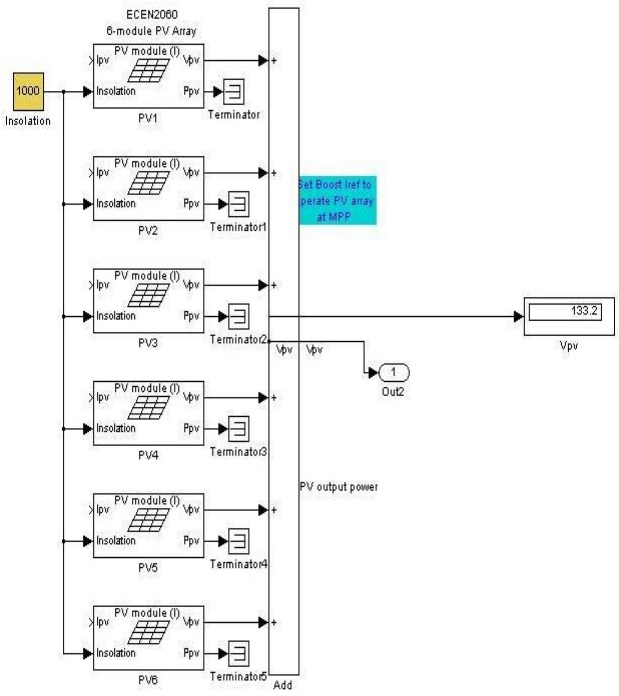


Fig 13 PV Array

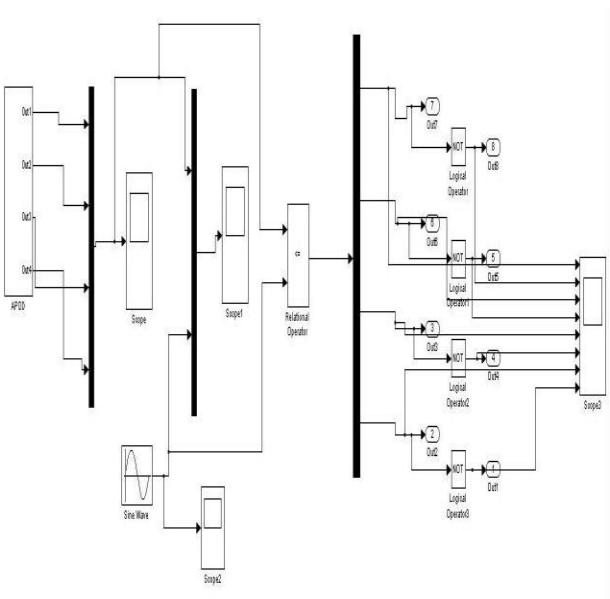


Fig. 14 Pulse Generator with Multi Carrier PWM

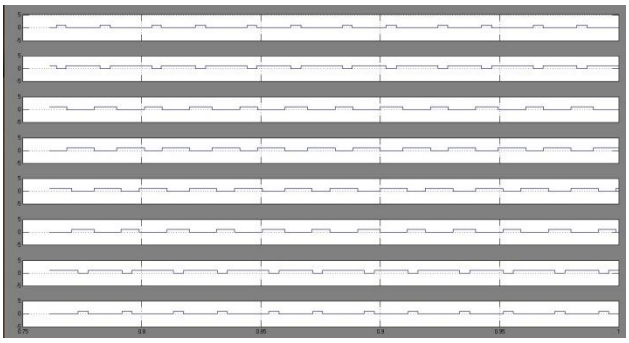


Fig. 15 Genrated Gate pulses

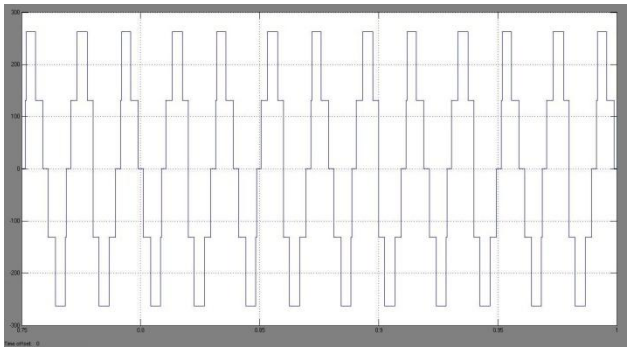


Fig. 16 Output Voltage of Five Level Inverte

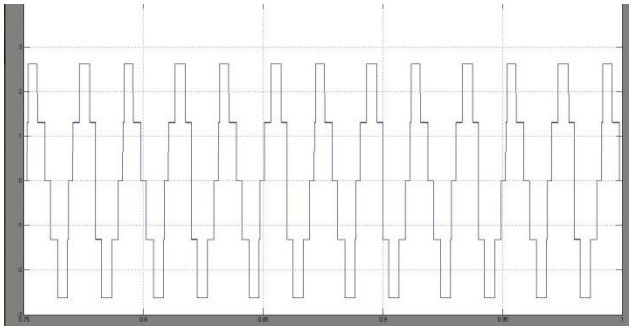
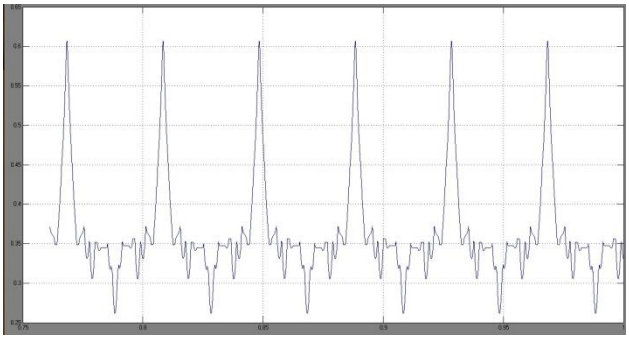


Fig. 17 Output Current of Five Level Inverter



18 THD

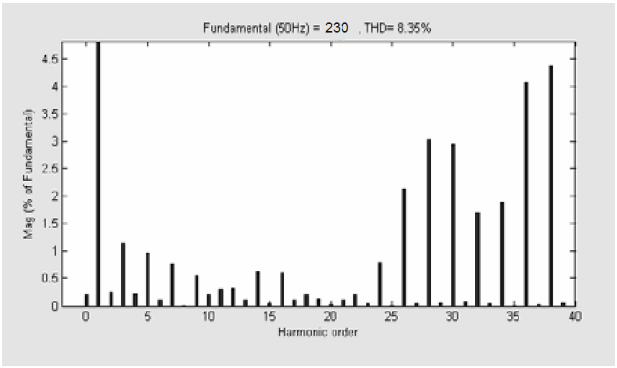


Fig 19 Harmonic spectrum of the simulated output

In this model APOD technique is used. The generated output pulses from the Multi carrier PWM block as shown in the Fig. 14 and those pulses generated are in eight numbers which is required to drive the devices in to ON state with the aid of Muti carrier pulse converter blocks as shown in figure 15. The output voltage and output current are as shown in fig. 16 and fig 17 respectively

The switching patterns adopted are applied at the cascaded multilevel inverter switches to generate five output voltage levels at 0.9 modulation index. The THD of inverter output voltage and Harmonic spectrum of the simulation system is as shown in the fig.18 and Fig 19 respectively, which shows the results are well within the specified limits of IEEE standards. The experimental and simulated results are show satisfactory results in term of total harmonic distortion and output voltage and current waveform shapes To verify the validity of the proposed PV Cell Based Five Level Inverter using multicarrier PWM The results of both output voltage and FFT analysis are verified by simulating the main circuit using MATLAB

CONCLUSION

This paper presented an five-level cascade H-bridge Inverter, which uses MC PWM and PV system with separate solar panels as DC sources to interact with the power grid. The MC PWM strategy reduces the THD and this strategy enhances the fundamental output voltage. When Modulation Index is equal to 0.9 by adopting MC PWM strategy the THD value is 6..35. Those schemes confirmed by simulation results. This proposed prototype can be extended to m-level inverter.

REFERENCES

[1] J.Rodriguez, Jih-sheng Lai, and F Zheng peng, "Multilevel Inverters; A Survey of Topologies, Controls, and Applications", IEEE Trans.Ind.Electron., vol.49 , no4., pp.724-738. Aug.2002.

[2] K.A Corzine, and Y.L Familiant, "A New Cascaded Multi-level H-Bridge Drive", IEEE Trans. Power.Electron., vol.17, no.1, pp.125-131. Jan 2002.

[3] G.Carrara, S.Gardella, M.Marchesoni, R.salutari,and G.sciutto, "A New Multilevel PWM Method; A theoretical analysis", IEEE Trans. Power.Electron., vol.7, no.3, pp.497-505. Jul.1992.

[4] L.M.Tolber, T.G.Habetler, "Noval Multilevel Inverter Carrier based PWM Method", IEEE Ind.Appli., vol.35. pp.1098-1107. Sep/Oct 1999.

[5] B.P.McGrath and Holmes, "Multicarrier PWM strategies for multilevel inverter," IEEE Trans.Ind.Electron., vol.49, no.4, pp.858-867. Aug.2002.

[6] A.M Hava, R.JKerman , and T.A Lipo, "Carrier-based PWM-VSI Overmodulation Strategies: Analysis, Comparison, and Design," IEEE Trans. Power.Electron., vol.13, no.4, pp.674-689. Jul.1998

[7] Infield.D, Kettle Borough.J.G and William

Halcrow, "Modeling and Experimental Study of the Interaction of Multiple PhotoVoltaic Inverters", by the *Centre for Renewable Energy Systems and Technologies, EPSRC*, Loughborough University, pp. 234-239, July, 1998

[8] Philip T.Krein ,Robert S.Balog and Xin Geng,"High-Frequency Link Inverter for fuel cells Based on Multiple Carrier PWM", *IEEE Transactions on Power Electronics*, Vol 19, N0.5, Sep 2004.

[9] L. M. Tolbert, F. Z. Peng, and T. G. Habetler "Multilevel Converters for Large Electric Drives," IEEE Transactions on Industry Applications, vol. 35, no. 1, Jan/Feb. 1999, pp. 36-44.

[10] M. F. Escalante, J. C. Vannier, and A. Arzande "Flying Capacitor Multilevel Inverters and DTC Motor Drive Applications," IEEE Transactions on Industry Electronics, vol. 49, no. 4, Aug. 2002, pp. 809-815.

[11] L. M. Tolbert, F. Z. Peng, "Multilevel Converters as a Utility Interface for Renewable Energy Systems," in Proceedings of 2000 IEEE Power Engineering Society Summer Meeting, pp. 1271-1274.

[12] L. M. Tolbert, F. Z. Peng, T. G. Habetler, "A Multilevel Converter-Based Universal Power Conditioner," IEEE Transactions on Industry Applications, vol. 36, no. 2, Mar./Apr. 2000, pp. 596-603.

[13] L. M. Tolbert, F. Z. Peng, T. G. Habetler, "Multilevel Inverters for Electric Vehicle Applications," IEEE Workshop on Power Electronics in Transportation, Oct 22-23, 1998, Dearborn, Michigan, pp. 1424-1431

[14] R. W. Menzies, Y. Zhuang, "Advanced Static Compensation Using a Multilevel GTO Thyristor Inverter," IEEE Transactions on Power Delivery, vol. 10, no. 2, April 1995, pp. 732-738.

[15] L. M. Tolbert, F. Z. Peng, T. G. Habetler, "Multilevel Inverters for Electric Vehicle Applications," IEEE Workshop on Power Electronics in Transportation, Oct 22-23, 1998, Dearborn, Michigan, pp. 1424-1431.

[16] Leon M. Tolbert, Fang Z. Peng, Tim Cunnyngham, John N. Chiasson, "Charge Balance Control Schemes for Multilevel Converter in Hybrid Electric Vehicles," IEEE Transactions on Industrial Electronics, vol. 49, no. 5, October 2002, pp. 1058-1065

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue2, pp-552-558 ISSN: 2249-6645

Analysis and modelling of work stress in manufacturing industries in Kerala ,India

K. Satheesh Kumar[1], Dr.G.Madhu[2]

[1.]Assistant professor (Sr.Grade), Department of Mechanical Engineering, Federal Institute of Science and Technology, Ernakulam, Kerala, India.

[2].Professor and Head , Safety and Fire Engineering Division , Cochin University of Science and Technology, Ernakulam,Kerala, India

Abstract

This study examines the influence of factors responsible for work stress among the employees in the public sector industries in Kerala, India. The sample size of the subjects selected for the study consists of 75 Engineers, 110 Supervisors and 675 Workers in the selected manufacturing industries in kerala, India. Seven factors were identified with the existing literatures, and in consultation with safety experts for the evaluation of work stress. The instrument developed by using these factors had validity, unidimensionality and reliability. The response rate was 81.3%. It is observed that existence the factors responsible for work stress among all the categories of employees in these industries. It is also noted that relatively low level of control among workers is the main cause of work stress. The factor model and structural equation model proposed are equally good in predicting the work stress in manufacturing industries.

Keywords: Work stress, structural equation model, manufacturing industries, factor model

1. Introduction

Occupational stress is becoming a major problem in both corporate and social sectors .In industrialized countries, there have been quite dramatic changes in the conditions of work, during the last decade due to the economic, social and technical development. As a consequence the people today at work are exposed to high quantitative and qualitative demands at the work place. In multinational companies, lean production, and down sizing has raised stress level of employees [1]. The national institute of occupational safety and health (NIOSH-USA) defines stress as "the harmful physical and emotional responses that occur when the requirements of the job does not match with the capabilities, resources of the workers."

The cost associated with work place stress indicate an international trend among industrialized countries. A recent report says that work related ailments due to work related stress is likely to cost India's exchequer around ₹72000 Crores between 2009-15 [2]. Though India is a fast developing country it is yet to create facilities to mitigate the adverse effects of work stress. The study of work stress in the member states of European Union (EU) points out that an average of 22% of the working Europeans experience work stress [3].

It is noted that work stress occurs among the employees at the context of work and at the content of work [4]. The potential stressors for these hazards in the context of work are organizational culture and function, role in the organization, career development, decision latitude and control, interpersonal relationship at work, work-home interface and change [4,5].

Studies on the employees perceptions and descriptions of their organizations, suggest three distinct aspects of organizational function and culture: organization as a task environment, as a problem solving environment and as a development environment [6,7]. The available evidence suggests that the organization is perceived to be poor in respect to these environments, will likely to be associated with higher stress. It is found that factors like poor communication, poor leadership, and lack of clarity about the organizational objectives and structure of the organization may lead to work stress [8].

Another major source of stress is associated with persons role at work.A great deal of research is done on role ambiguity and role conflict. Role ambiguity is the result of employees uncertainties, lack of information about the job role, expectation and responsibilities [4].It is found that role conflict and role ambiguity are instrumental in developing physiological disorders and says that the above factors can also lead to organizational dysfunction and decreased productivity [4].

Lack of expected career growth is one of main sources of work stress. The factors connected with this are poor promotion polices, job insecurity and poor pay in the organization [4].Earlier studies show that poor promotion prospects and blocked career may lead to work related stress hazard like coronary heart disease (CHD)[9].

Decision latitude and control are important aspects of work stress. These shows the extent which the employees are participating in the decision making process, and also shows the freedom given to the employees for choosing their work. The researchers indicate that individuals with highest income group was a most likely to have low strain due to greater job control [10].

The number of research works points out the need of good relationship with superiors ,support from the superiors and support from the colleagues at work for the elimination of work related stress hazards. It is found that the real source of problems connected with work stress are not located in the work environment, but is person-based, and the most effective way to reduce stress is to change the person based factors. Accordingly a questionnaire has been developed and circulated among the check out assistants in the age groups 18 to 56yrs, who belong to both sex. It is noted that higher level of job demands with lower level of support at work resulted in increased job stress [11].

Many literature points out the work related stress hazards due to work-family conflict .It is found that that work-family conflict is a form of inter role conflict ,in which the role pressures from the work family domains are mutually non compatible in same respect [12].

Change is one of the most commonly found stressor at the context of work[13]. It is observed that changes in the modern work environment as result of technological advances, organizational restructuring and various redesign options can elevate the work stress [4]. Researchers indicate that rapid changes along with poor relationship can lead to one set of work related stress hazards [14].

Like context of work, content of work are also leads to work stress. These factors arise due to improper design of the task ,work load and work pace,and work schedule [4,5].

There are several aspect of job content ,which are found hazardous and these include low value of work ,low use of skills ,repetitive work , uncertainty , lack of opportunity to learn, high attention demand , conflicting demand , insufficient resources [4].The research work shows that ,work related stress hazards arise due to meaning less task and lack of variety etc…It is also noted that most stressful type of work are those which have excessive demand and pressures that do not match with the workers knowledge and abilities [15].

The studies on the effect of work stress among men and women working groups in USA and found that due to high psychological work demands like excessive work load and time pressures leads to work stress and cause depression and anxiety in young working adults[16]. It is noticed that work related stress hazards like depressive disorders and abdominal fat among workers due high work demands [17]. A higher correlation between work stress and Coronary Heart Disease (CHD) was noted by many researchers in their study among male and female employees of different age groups [18].

Two major factors responsible for work stress due to the improper work schedule are shift work and long working hours .The studies conducted in Italy among the shift workers observed that shift work leads to poor sleep and health related problems [19].

Studies conducted among white collar workers in Sweden, points out that work stress is associated with men subjected to long working hours (75 hours/week) and it is shown that this leads to wide range of ill health in men and women[20].

Several models have been proposed to explain the causes of work related stress. Frankenhaeuser have described a model where stress is defined in terms of imbalance between the perceived demands from the environment and individuals perceived resources to meet those demands[21].This imbalance can be caused by quantitative overload (A very high work pace, too much work to do etc…) or qualitative overload (too much responsibility, problems too complex to solve, conflicts etc…)

A well known model describing work stress or strain is the demand control model proposed by karesek and Theorell and developed and expanded by others. According to this model, the combination of high demands and lack of control and influence (low job discretion) over the work situation causes high work strain[22].

Johannas Siergrist proposed a new model for stress at the work called the effort-reward imbalance model. According to this model, lack of adequate reward in response to the individual's achievement efforts is considered to contribute to high stress levels and elevated health risks .Reward could be obtained in terms of economic benefits, such as higher income [23, 24].

Factor analysis is the basic model and has received a lot of attention in the field for many years [25] and is used for the develop the relationship of a set of variables [26, 27].

Structural equation modelling of work stress was done by many researchers earlier [28].In this association between the different variables namely stress, health, work, family and finance are analyzed. The structural equation modelling was done by means of confirmatory factor analysis.

2. Subjects

Total number of subjects selected for this study is 830 and the resulted sample consists of Engineers (75 Nos.),

Supervisors (110 Nos.) and workers (675 Nos.). Participants selected for this study consists of both male and female employees of age between 25 to 55 and had sufficient educational back ground for their job. All employees are permanent and working in shifts in rotation and each shift consists of 8 hour duration per day. However the majority of the employees, in these industries were males and number of woman participants is about 10% of the male participants. All the industries are large scale and profit making for the last five years and located at different districts of Kerala, India. .

3. Methods

From the literature review and with the consultation of safety experts seven factors were identified for the evaluation of work stress in the absence of well defined factors for the evaluation of work stress in Kerala ,INDIA. They are demand, control, manger support, peer support relationship, role and change. The final draft of the questionnaire had 35 items with seven subscales .All the questions were likert type with five fixed alternatives(always, often, sometimes ,rarely, never). In addition to this 10 demographic questions are also included in the questionnaire. This questionnaire was refined and validated further by means of confirmatory factor analysis (CFA)[29,30].This resulted in removal of five items from the questionnaire. The number of retained items in the questionnaire were demand (7 items), control (4 items),manager support (4 items),peer support (4 items),relationship (4 items), role (5 items) and change (2 items). The values of Comparative Fit Index (CFI), Tucker Lewis Index (TLI),and Cronbach alpha shows that the refined scale has good validity and unidimensionality in addition to reliability [31-33]. The analysis was performed by using the software AMOS-7 [34].The filled up schedules are then carefully edited for completeness, consistency and accuracy . The overall response rate was 81.3%.

On the basis of data so collected, the influence of factors on works stress analysis is performed using one-way ANOVA . A Factor modelling of work stress was done by means of Alpha factor analysis and Structural equation modelling of work stress was done further to find the association of factors responsible for work stress in manufacturing industries.

4. Results

4.1. Correlation Matrix

A correlation analysis between the variables /factors so identified was performed and the result of the analysis is given in the Table-1.It is noted that all the correlations were positive, but no significant correlation was found between the variable/factors(<0.5). Therefore the variable selected for the study can be treated as indep endent variables for the purpose of research. The correlation analysis were carried out by means of SPSS-15.

Table – 1 Correlation between the factors

Variables/ Factors	Demand	Control	Manager support	Peer support	Relationship	Role	Change
Demand	1	0.354	0.249	0.240	0.310	0.214	0.196
Control	0.354	1	0.279	0.227	0.310	0.168	0.251
Manager support	0.249	0.279	1	0.426	0.319	0.313	0.357
Peer support	0.240	0.227	0.426	1	0.498	0.313	0.461
Relationship	0.310	0.310	0.319	0.498	1	0.440	0.474
Role	0.214	0.168	0.313	0.313	0.440	1	0.353
Change	0.196	0.251	0.357	0.461	0.474	0.353	1

4.2 .Influence of factors on different categories of employees

The influence of these factors are analyzed among different categories of employees by means of one-way ANOVA .The result of the test is given in the Table -2 .The test is conducted for 0.5 level significance.

Table-2. Mean score of factors

Variables/Factors		Designation			F-value	P-value
		Engineer	Supervisor	Worker		
Demand	Mean	25.72	26.08	25.61	0.603	0.548
	S.D	3.78	4.36	4.11		
Control	Mean	15.09	13.70	12.32	16.644	< 0.001
	S.D	2.96	3.85	4.35		
Manager support	Mean	14.64	15.19	13.94	5.953	0.003
	S.D	2.96	3.85	4.35		
Peer support	Mean	15.94	16.10	15.34	3.748	0.024
	S.D	1.87	2.98	3.12		
Relation-ship	Mean	16.28	17.01	16.41	2.035	0.131
	S.D	1.98	2.82	3.04		
Role	Mean	22.79	22.93	22.56	1.187	0.306
	S.D	2.03	2.14	2.56		
Change	Mean	6.69	6.77	7.02	1.313	0.270
	S.D	1.73	2.34	2.04		

The mean score of the factors /variables points out that existence of factors responsible for work stress among all the categories of the employees in these industries.

It is noted that , significant difference in the factors, control, manager support, and peer support ($p<0.05$) among different categories of employees To identify which among the categories has significant difference , Tukey's multiple comparison test for each of the factors and the results are given in the Table -3

Table -3. Significant difference between different categories of employees

Factors/Variables	Difference between different designation levels
Control	Engineer and worker Supervisor and worker
Manager support	Supervisor and worker
Peer support	Supervisor and worker

The post- hoc analysis, reveals that considerable difference in the mean score of the factor "control" exists between engineers and worker.. Further a noted difference is observed for this factor between supervisor and worker .While analyzing the variables manger support and peer support considerable difference is observed only between supervisors and workers

4.3. Modelling of work stress

Modelling of work stress was done by earlier by several researchers [35,36], and this will help to analyze the work stress under the influence of different factors. Accordingly two different type of modelling for work stress carried out for this study are by means of Factor modelling and Structural equation modelling .

4.3.1 Factor modelling of work stress

Factor modelling of work stress was carried out by means of seven factors by Alpha method of factor analysis [37-39]. This yielded two factor structure for work stress as shown below (Table-4).It is noted that for each of the factors some variables had a higher factor loading ($\geq$0.4).For Factor -1,the variables manger support, peer support, relationship, role ,change had a high loading . The variables demand and control had a high loading on Factor -2. It is noted that factors /predictor variables namely demand and control are person based and mean while the other factors are team based and this made us to name the two factors as stress-personnel (Stress-P) and stress-team (Stress-T).

Table -4. Factor Matrix

Variables	Factor	
	1	2
Demand	0.167	0.968
Control	0.328	0.501
Manager support	0.748	0.178
Peer support	0.473	0.089
Relationship	0.689	0.304
Role	0.435	0.217
Change	0.654	0.238

Hence the above factors can be modeled as

Stress-P = 0.968 De +0.501 Cl and

Stress-T = 0.748 Ms + 0.473 Ps + 0.689 Re + 0.435Rl + 0.654 Ch

Where De, Cl, Ms, Ps, Re, Rl, Ch represents the variable demand, control , manager support, peer support, relationship, role and change and the above two models can be effectively used for the evaluation of work stress

4.3.2 Structural equation modelling of work stress.

Structural equation modelling of work stress was done by using the seven factors .This yielded two components for the work stress namely stress-personnel (stress-P),and stress-team (stress-T). The structural equation model was developed by using confirmatory factor analysis [40].This is shown in the Fig- 1. The rectangle represents observed factor /variables ,which are demand ,control, manager support ,peer support ,relationship ,role and change .Ovals are drawn on the diagram to represent work stress, which has been shown as two types stress-personnel (Stress-P) and stress-team (Stress-T).

The variable error is enclosed in a circle. the double headed arrows in the path diagram connect the variables ,which are correlated to each other[34]. The standardized regression weights are shown over the arrows. The squared multiple correlation of each observed variables /factors are represented over each of the respective rectangles.

5. Discussion

The main aim of the study is to develop and analyze the factors responsible for work stress among the employees in the public sector manufacturing industries in Kerala ,India. Accordingly seven factors were developed and the validity,and unidimensionality of the questionnaire was analyzed by means of CFA and the overall reliability of the questionnaire was found satisfactory (>0.70). . Interestingly it is found that the factors responsible for work stress is prominent in different categories of employees namely engineers, supervisors and workers these industries. It is also noted that lack of control among lower categories of employees particularly among workers compared to other categories of employees. The results of many earlier research supports the finding[9,41].

In factor modelling , alpha method of factor analysis was used to develop the model .This yielded two factor structure of work stress namely stress-personnel(Stress-P) and stress-team(Stress-T). This model can be effectively used for predicting the work stress in manufacturing industries .

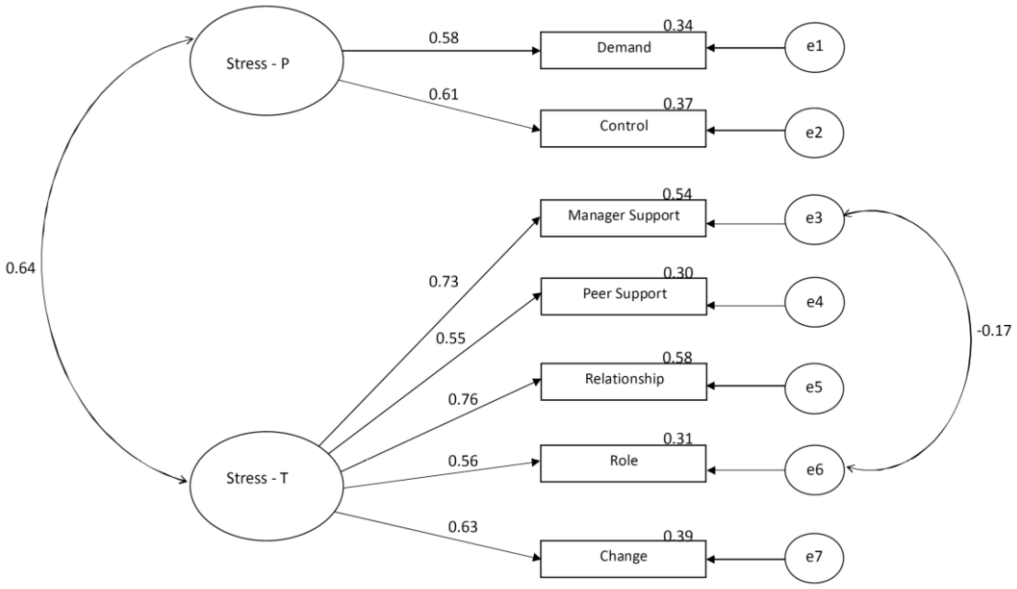


Figure -1 Structural Equation Modelling of Work Stress

Table-5. Model Fit Indices

$\chi 2$	Normed $\chi 2$	df	P-Value	GFI	AGFI	NFI	TLI	CFI	RMR	RMSEA	SRMR
18.336	1.528	12	0.106	0.994	0.986	0.986	0.991	0.995	0.202	0.025	0.0012

Structural equation modelling of work stress by using the seven variables was developed .The following goodness of fit indices are used to access the degree of fit, between the model and sample. Normed χ^2 (acceptable between 1 and 3) [42]. Normed fit index (NFI >0.90 excellent) [42,43].Tucker Lewis Index (TLI >0.90 acceptable, > 0.95 excellent) [44]. Comparative Fit Index(CFI >0.90 acceptable,>0.95excellent)[45,46].Root mean square error of approximation (RMSEA<0.08 acceptable,<0.05 excellent) [47]. Standard root mean square residual (SRMR <0.05 excellent) [48]. The values obtained for NFI,CFI, and TLI , Normed χ^2 ,RMSEA, SRMR values are well with in the acceptable limit ,hence structural equations model is found good in representing the work stress. The modelling was done by using AMOS-7 [34]

Initially an input structural equation model was developed by using the seven variables and it is noted that goodness of fit indices were not with in the acceptable limit for this model .Hence this model was modified further and the modified version is given in the Fig 1. It is noted that the goodness of fit indices for this modified model is well with in the acceptable limit (See Table- 5) and this model can be used to predict work stress among the employees in manufacturing industries.

Like any other research, the study also not free from limitations. The present study is limited only to public sector industries in Kerala, India, where majority of employees are males. Therefore it would be inappropriate to draw conclusions about male and female workers based on this result. The conclusion is drawn based on the data obtained by means of self reported measures. A comparative study was not carried out because of lack of literature or study of work stress in the context of Indian public sector industries.

6. Conclusion

Consistent with the literature, the results indicate that existence of factors responsible for work stress among all the categories of the employees working in public sector industries in Kerala, India and the instrument developed for the evaluation of work stress by using the variables / factors ,namely demand ,control, manager support, peer support, relationship, role and change had validity, unidimensionality and reliability and the instrument can be effectively used for the evaluation of work stress in different type of industries in addition to manufacturing industries . Low level of job control was noticed among lower designation level particularly among workers than engineers and supervisors. The factor model and structural equation model proposed are equally good in representing work stress in the manufacturing industries.

Acknowledgement

Discussions with Mr. Jacob Devassy, President, International Efficiency Institute, Kochi – 27, India (Subsidiary of International Safety Institute Incorporated, Toranto, Canada) is greatfully acknowledged.

References

1. International Labour Organization (ILO) report on work stress, 2005.
2. TheEconomicTimesdated10May2009, http:// eonomictimes.indiatimes .org.
3. European Agency for Safety and Health at Work (EASHAW), European risk observation report, 2005.
4. T. Cox, A. Griffiths ,E. R. Gonzalez, *Research on work related stress.*(European agency for safety and health at Work, Official publication of European communities, Luxemburg ,2000).
5. C. J. Mackay, R. Cousins, P.J. Kelly, S. Lee, R. H. McCaig , Management standards and work related stress in UK policy back ground and science, *Work and Stress*, 18(2), 2004,91-112.
6. T. Cox, I .Howarth, Organizational heal th culture and helping, Work *and Stress*, 4, 1990, 107-110.
7. T.Cox, M. Leiter, The health of the health care organizations, *Work and Stress*, 6,1992,219-227.
8. S. Lekha, A. Griffiths, T. Cox, Work organi zation and work stress, Protecting Workers Health Series No.3, 2003,1-32.
9. H. Bosma, R. Peter, J. Siegrist, M. Marmot, Two alternative job stress models and risk of coronary heart disease, *American Journal of Public Health*, 88(1),1998,68-74.
10. J.Park, Work stress and job performance. *Perspective – Statistics Canada .Catalogue No.75-001XE* , 2007,5-17
11. C. Ben , FIT work demand and work supports, 2007 ,1-3.
12. H. Yang, P. L. Schnall, M. Jauregui, T.C, Su , D. Backer ,Work hours and self reported hypertension among working people in California, .Hypertension,48, 2006,744-750
13. K. Launis, J. Pihlaja, Changes in production concepts emphasize problems in work-related well being, *Safety science*, 45,2007, 603-619.
14. J. Shigemi, Y. Mino, T. Tsuda, A. Babazono, M. Aoyama, The relationship between job stress and mental health at work, *Industrial Health* ,35 , 1997,29-35.
15. World health organization (WHO), Occupational Stress at Work Place, WHO *publication*, 2007, 1-12.
16. M. Melchior, A. Capsi, B.J. Miline, A. Danese, R. Poulton, T.E. Moffit, . Work stress precipitates depression and anxiety in young working women and men, *Psychological Medicine*, 37(8), 2007, 119-1129.
17. L. Levi, European commission guidance on work stress: From word to action, TUTB *News Letter*, 2000, 1-6.
18. T.Chandola,A.Britton,E,Brunner,H.Hemingway, M.Mallic, Meerakumari.,E.Badrick, M. Kivimaki, M. Marmot, Work stress and coronary heart disease: What are the mechanisms?, *European Heart Journal Advance Access*, 2008,1-9.
19. P. M. Conway, P. Companini, S. Sartori, R. Dotty, G. Costa, Main and interactive effects of shift work ,Age and work stress on health initialization sample health care workers, *Applied Ergonomics*,2008, 39(5),630-639.
20. G. Krantz, L. Berntsson, U. Lundberg, Total work load work stress and perceived symptoms in Swedish male and female white collar employees ,*European Journal of Public Health* ,15(2), 2005,209-214.
21. M. Frankenhaesuer, A Psychobiological frame work for research on human stress and coping, in M.H.Appley , R.Thrumbell,(Ed.) , *Dynamics of Stress :Physiological ,Psychological, and Social Perspectives*(Plenum ,New York , 1986)101-116.
22. R. Karasek, T. Theorell, *Healthy work, stress productivity and the reconstruction of working life*.(USA Basic books,1990).
23. J. Siegrist, D. Starke, T. Chandola, I. Godwin, M. Marmot, I. Weid hammer, R. Peter, Measurement of Effort Reward Imbalance at Work: European Comparisons. *Social Science and Medicine*, 58 (8), 2004, 1483-1499.

24. J. Siegrist, Adverse health effects of high effort low reward conditions at work, *Journal of Occupational Health Psychology*, 1, 1996,27-43.
25. S. Y. Lee, *Structural Equation Modeling: A Bayesian Approach*. (Wiley series, 2007).
26. C. Spearman, General Intelligence object tively determined and measured, *American Journal of Psychology*, 15, 1904, 201-293.
27. L.L. Thurstone, A multiple group method of factoring the correlation matrix, *Psychometrika*,10, 1944,73-88.
28. Y. H. Chan, Biostatics 308. Structural equation modeling, *Singapore Medical Journal ,46(2),2005,675-679* .
29. R. Kendell, A. Jablensky, Distinguishing between validity and utility of psychiatric diagnoses .*American Journal of Psychiatry*, 160(1), 2003, 4-12.
30. R. G. Natemeyer, W.O. Bearden, S. Sharma, *Scaling Procedures –Issues and Applications*, (NewDelhi ,Sage,2003).
31. L.J.Cronbach, P. E.Meehl, Construct validity in psychological tests. *Psychological Bulletin*, 52, 1955, 281-302.
32. R. F. Devellis, *Scale development theory and applications*, *Applied social research methods* (Sage ,New Delhi ,2003).
33. S. L Ahire, D.Y.Golhar, M. A .Waller, Development and validation of TQM implementation construct, *Decision Science*, 27(1), 1996, 23-56.
34. J. L.Arbuckle,AMOS 7.0 (AMOS Users Guide .Chickago,IL: Small waters corporation 2006).
35. H.L Hsieh, L.C. Huang, K.J. Su, Work stress and job performance in Hi-Tech industry: A closer view for vocational education, *World Transactions on Engineering and Technology*. 3(1),2004, 147-150.
36. S. Palmer, C. Cooper, K. Thomas, A Model of Work Stress, Counseling at work, *HSE Publication, 2004*, 1-4.
37. A. B. Costello, J. W. Osborne, Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis, *Practical Assessment Research and Evaluation* ,10(7),2005,1-9
38. H.F. Kaiser, J. Coffery, Alpha factor analysis, *Psychomertica* .30,1965,1-14.
39. M. Mc Dermeit, R. Funk, M. Foss, M. Dennis, Exploratory Factor Analysis With Alpha Method and Varimax Rotation, *LI Analysis Training Series* ,Chestnut Health Systems ,Bloomongton, 2000,1-8.
40. D. Harrington, *Confirmatory Factor Analysis* (New York, Oxford University Press, 2009).
41. R. Karasek, Demand-Control/support model, A social, emotional and physiological approach to stress risk and active behavior development, in J.M. Stellman, (Ed.),*Encylopedia of Occupational Health and Safety*,2,(Geneva,ILO,1998) 62-69.
42. J. F.HairJr, R. E.Anderson, R. L.Tatham, W. C.Black, *Multivariate Data* A*nalysis* (Englewood Cliffs, NewJersey, PrenticeHall, 1988).
43. B. M. Byrne, *Structural Equation Modelling with AMOS:Basic concepts, Applications and Programming*(New Jersey, Laurence Erlbaum Associates,2001)
44. L. R. Tucker, C. Lewis, A reliability coefficient for maximum likelihood factor analysis, *Psychometrica*,38, 1973,1-10
45. P. M.Bentler , Comparative fit index in structural models ,*Psychological Bulletin*, 107, 1990,238-246.
46. P. M.Bentler, D. G. Bonnet, Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 88,1980, 588-606.
47. M. W.Brown, R. Cudeck, Alternative ways of assessing model fit ,in K. A.Boolen, J. S. Lang(Ed), *Testing Structural Equation Models*, (Newbury park,CA,Sage 1998)136-162.
48. L.T. Hu, P.M. Bentler, Evaluating Model fit, in R.H .Hoyle(Ed.),*Structural Equation Modeling Concepts Issues and Applications* (London: Thousand Oaks sage publications,1995) 76-99.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue2, pp-559-563 ISSN: 2249-6645

Biodegradable Thermal Insulation for Ice-Coolers

Krishpersad Manohar

(Department of Mechanical and Manufacturing Engineering, The University of the West Indies, St. Augustine, Trinidad and Tobago, West Indies)

ABSTRACT

The widespread use of non-biodegradable low temperature thermal insulation is impacting negatively on the environment with respect to disposal. In this study the effectiveness of using biodegradable locally available coconut fiber for insulation in ice-coolers was investigated. A comparative method was used to determine the effectiveness of a zinc-coated metal sheet ice-cooler with coconut fiber insulation in relation to standard commercially available Rubbermaid and polystyrene ice-cooler. The density of the coconut fiber insulation was varied from 50kg/m^3 to 95kg/m^3 for ice-coolers with wall thickness 38mm, 51mm, and 64mm, respectively. For each density the melt rate was measured and compared with the Rubbermaid and polystyrene coolers. The laboratory built ice-coolers were approximately the same volume and similar in shape to the commercial ice-coolers. From the melt rate experimental results of the three laboratory built coconut fiber insulated ice coolers the 64mm thick 95kg/m^3 density ice cooler performed the best. The 51mm and 64mm thick ice coolers performed consistently better than the Rubbermaid cooler. The 51mm and 64mm thick ice coolers performance were comparable to that of the polystyrene ice cooler.

Keywords – **Biodegradable insulation, coconut fiber, ice-cooler, thermal insulation**

1. INTRODUCTION

Daily degradation of our delicate environment is of concern to everyone. The advent of technology has brought with it glamorous new findings, luxurious facilities and exotic life stiles. However, the negative impact of technological advancement is causing serious and sometimes irreparable damage to the environment. In the field of low temperature insulation technology low-cost foam and polystyrene have been used extensively. Continuous research have perfected the manufacture and production of these materials. One can design these materials for specialized applications. Foam (rigid or flexible) is the most widely used material for low temperature insulation [1]. The wide ranging application of foam insulation covers use in clothing, air conditioning systems, commercial and residential buildings, automobile passenger compartments, refrigerators and ice coolers. The negative effects of wide spread use of non-biodegradable thermal insulation has caught up with modern society. In developing countries disposal of these materials is becoming a bigger problem daily. Discarded polystyrene does not biodegrade for hundreds of years and is resistant to photolysis [2]. Because of this stability, very little of the waste discarded in today's modern, highly engineered landfills biodegrades. Because degradation of materials creates potentially harmful liquid and gaseous by-products that could contaminate groundwater and air, today's landfills are designed to minimize contact with air and water required for degradation, thereby practically eliminating the degradation of waste [3]. Land-fills are being packed to capacity and rampant improper disposal is reeking havoc on the delicate environment.

There is an urgent need for more environmentally friendly biodegradable low temperature thermal insulation. Coconut fiber was always known for its high resilience in moist environments [4]. Before the advent of foam, coconut fiber was widely used in mattresses and cushion seats. Due to a lack of thermo-physical property data and research the possibility of using coconut fiber as an effective low temperature insulator was not explored [5, 6]. With the widespread availability of cheap foam the use of coconut fiber today is almost non-existent.

In this study the thermal insulating property of bio-degradable natural unprocessed coconut fiber was investigated. The effective insulating property was determined by a comparative method. The coconut fiber was tested in three laboratory built ice coolers and compared with two commercially available ice coolers. The commercially available coolers used were the Rubbermaid ice cooler (foam insulation) and the polystyrene ice cooler (polystyrene insulation).

2. METHODOLOGY

To investigate the thermal insulating property of coconut fiber three ice coolers were constructed from zinc-coated metal sheet. The inner shape and dimensions of the coolers were the same as the inner shape and dimensions of the Rubbermaid and polystyrene coolers. To facilitate the testing of different thicknesses of coconut fiber insulation the outer housing was constructed to accommodate 38mm, 51mm and 64mm thick insulation.

Fig. 1 shows a cross sectional view of the laboratory built ice cooler with the inner dimensions. Respective covers for the ice coolers were also constructed from zinc-coated metal sheet to accommodate 38mm, 51mm and 64mm thick insulation. The covers were designed with a 12.5mm protrusion into the ice-cooler compartment. The protrusion together with a thin rubber gasket ensured an air-tight seal at the covers (Fig. 1).

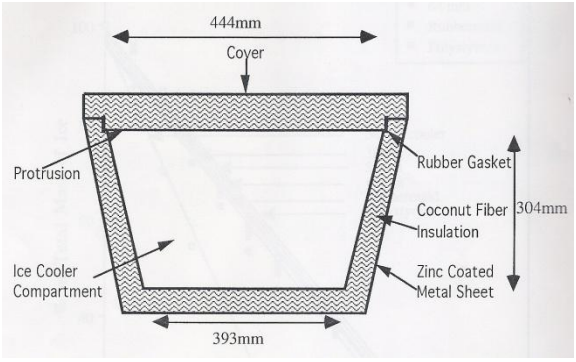


Figure 1: Cross section of laboratory built ice cooler.

To investigate the insulating properties of coconut fiber the cavity between the inner and outer housing of the ice coolers were filled with coconut fiber packed at various densities. The densities tested were 50kg/m^3, 65kg/m^3, 80kg/m^3, and 95kg/m^3. This density range was chosen since it is within the optimum density range for coconut fiber insulation [7].

To measure the effective insulating properties of the coconut fiber ice coolers the melt rate of ice was monitored. Identical specimens of ice in both shape and size were used in each ice cooler (Fig. 2).

Figure 2: Ice specimen used to monitor melt rate.

For each test density experiments were conducted with the 38mm, 51mm, and 64mm thick coconut fiber insulated ice coolers together with the Rubbermaid ice cooler (Fig. 3) and the polystyrene ice cooler (Fig. 4).

Figure 3: Rubbermaid ice cooler.

Figure 4: Polystyrene ice cooler.

One ice specimen was also placed in an open container on the test bench and was also monitored for each test run. Under these conditions for each test run all the specimens were subjected to the same environmental conditions.

At pre-set intervals the ice containers from each cooler was removed and the water (melted ice) drained and weighed. The ice was replaced in their respective containers and checked at the next pre-set time. The melt rate of the ice in each ice cooler and the specimen without an ice cooler was monitored over a twenty hour period. This test procedure was repeated for each test density of coconut fiber insulation.

3. INSTRUMENTATION

The scale used for the weight measurement of the ice samples was a Scientech Series 5000 electronic balance. The measurement range of this instrument was 0 to 5 kg with an accuracy of $\pm$ 0.005g and a resolution of 0.01g.

4. EXPERIMENTAL RESULTS

The melt rate of ice in each ice cooler was measured by subtracting the weight of the drained water from the mass of ice each time. One set of results for the test density 50kg/m^3 coconut fiber insulation ice coolers is given in Table 1. The percentage of the total mass of ice remaining was calculated from the measured results.

Table 2 shows the calculated results for the 50kg/m^3 coconut fiber insulation ice cooler tests. The calculated test data was used to plot respective graphs of percentage of total mass of ice versus time for each ice cooler at the various test densities. A typical set of graphs of the 50 kg/m^3 coconut fiber ice cooler test results from Table 2 is shown in Fig.5. The computer generated best fit line using the method of least squares and the corresponding linear equation was obtained for each test and the gradient of the best fit line provided an indication of the melt rate. Similar tables and graphs were obtained for each of the respective test with 65 kg/m^3, 80 kg/m^3 and 95 kg/m^3 density coconut fiber insulation ice coolers and the respective melt rate obtained from the best fit line. Table 3 gives a summary of the melt rate results for the various test densities.

Table 1: Mass of Ice Remaining in Ice Cooler with Time
Coconut Fiber Insulation at 50 kg/m^3

		Mass of ice remaining with time (measured values)					
Test No.	Time Elapsed (hours)	No ice cooler (g)	38 mm ice cooler (g)	51mm ice cooler (g)	64 mm ice cooler (g)	Rubbermaid ice cooler (g)	Polystyrene ice cooler (g)
1	0.0	1396.6	1377.6	1365.6	1383.9	1362.0	1378.5
2	1.5	1255.2	1342.8	1325.8	1342.0	1294.4	1338.1
3	3.0	987.1	1179.9	1171.2	1165.6	1120.9	1197.9
4	4.5	775.4	1043.9	1041.2	1034.1	987.5	1070.6
5	6.0	592.7	918.7	924.6	911.6	865.9	951.3
6	7.5	437.9	802.4	812.3	796.1	744.6	841.3
7	9.0	254.2	701.2	711.8	693.8	637.4	734.8
8	12.0	0.0	532.0	511.2	535.5	477.5	569.5
9	17.0	0.0	217.3	259.8	238.5	184.4	268.4
10	20.0	0.0	92.0	115.0	116.3	77.3	134.9

Table 2: Mass of Ice Remaining in Ice Cooler with Time
Coconut Fiber Insulation at 50 kg/m^3

		Percentage of total mass of ice remaining with time (calculated values)					
Test No.	Time Elapsed (hours)	No ice cooler (g)	38 mm ice cooler (g)	51mm ice cooler (g)	64 mm ice cooler (g)	Rubbermaid ice cooler (g)	Polystyrene ice cooler (g)
1	0.0	100.00	100.00	100.00	100.00	100.00	100.00
2	1.5	89.51	87.41	97.01	96.09	94.91	96.99
3	3.0	69.89	85.26	85.37	83.81	81.81	86.55
4	4.5	54.32	75.12	75.54	74.85	71.75	77.08
5	6.0	40.89	65.78	66.92	64.96	62.58	68.20
6	7.5	29.51	57.11	58.39	56.40	53.42	60.00
7	9.0	16.00	49.56	50.81	48.81	45.34	52.07
8	12.0	0.00	36.94	38.72	37.07	36.00	37.77
9	17.0	0.00	13.48	16.80	15.03	11.16	17.36
10	20.0	0.00	4.13	5.90	5.96	3.09	7.41

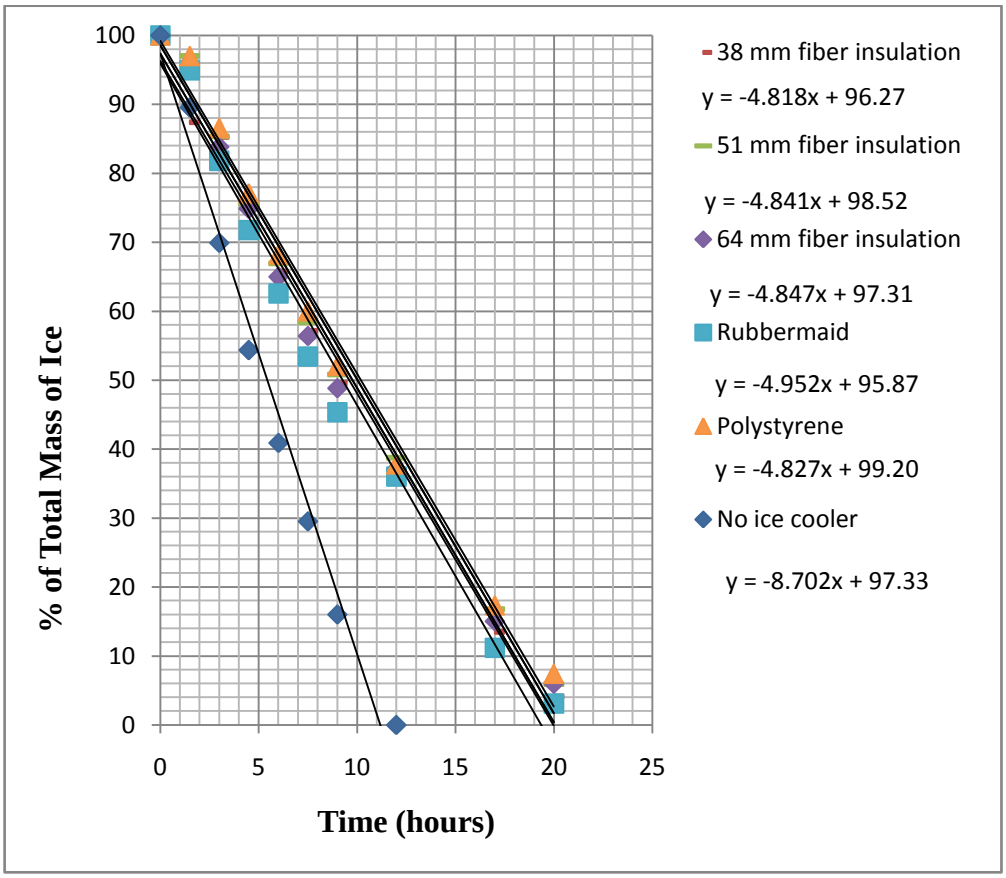


Figure 5: Graph of melt rate with time for test with coconut fiber insulation at 50 kg/m^3 density.

Table 3: Melt Rate of Ice Experimental Results

Ice Cooler Type	**Melt Rate of Ice (% of total mass per hour)**			
	Coconut Fiber Insulated Ice Coolers			
	50 kg/m^3	**65 kg/m^3**	**80 kg/m^3**	**95 kg/m^3**
Coconut fiber 38 mm thick insulation	4.818	4.892	4.818	4.705
Coconut fiber 51 mm thick insulation	4.841	4.866	4.731	4.509
Coconut fiber 64 mm thick insulation	4.847	4.673	4.808	4.480
No ice cooler	8.702	8.752	8.800	8.791
	Reference ice coolers under similar test conditions			
Rubbermaid ice cooler	4.952	4.955	4.995	4.814
Polystyrene ice cooler	4.827	4.827	4.777	4.797

5. DISCUSSION AND CONCLUSIONS

Using a comparative method to investigate the thermal insulating property of coconut fiber in ice coolers eliminated experimental errors due to uncontrollable environmental conditions. Monitoring the melt rate of the ice coolers for each density test run simultaneously provided data which can be directly compared.

Test results of percentage total mass of ice versus time (typical graph shown in Fig. 5) showed an initial slower melt rate with time over the first two hours with a gradual increase. Over the next ten hours the melt rate remained fairly constant. For the last eight hours the melt rate showed a small (< 2%) variation.

For the first two hours the slow melt rate can be attributed to thermal system being in a transition state. The constant melt rate over the next ten hours indicated that equilibrium conditions were established during this period. The small decrease in melt rate during the last eight hours was due to the falling ambient temperature during the night.

From the melt rate of the three laboratory built coconut fiber insulated ice coolers the 64mm thick 95kg/m^3 density ice cooler performed the best. The 51mm and 64mm thick ice coolers performed consistently better than the Rubbermaid cooler. The 51mm and 64mm thick ice coolers performance were comparable to that of the polystyrene ice cooler.

Comparing the melt rate of ice under similar conditions without any ice cooler showed close to 50% reduction in the melt rate. The experimental data showed that coconut fiber insulated ice coolers can be as effective as the commercially available ice coolers.

REFERENCES

[1] L. Glicksmann, M. Schuetz, and M. Sinofsky, Radiation heat transfer in foam insulation, *International Journal of Heat and Mass transfer*, 30(1), 1986, 187-197.

[2] A. Bandyopadhyay, and G. Chandra Basak, Studies on photocatalytic degradation of polystyrene, *Materials Science and Technology* 23(3), 2007, 307–317.

[3] W. Rathje and C. Murphy, *Rubbish! The Archeology of Garbage* (University of Arizona Press, 1989).

[4] G. S. Kochhar and K. Manohar, Effect of moisture on thermal conductivity of fibrous biological insulating materials, *Proc. ASHRAE Thermal Performance of the Exterior Envelope of Buildings VI*, Clearwater Beach, Florida, USA, 1995, 33-40.

[5] G. S. Kochhar and K. Manohar, Thermal conductivity of blanket coconut fiber, *West Indian Journal of Engineering*, 14(1), 1989, 81-85.

[6] K. Manohar, *Heat transfer mechanisms in biological fibrous materials*, doctoral diss., The University of the West Indies, St. Augustine, Trinidad and Tobago, West Indies, 1998.

[7] K. Manohar and G. S. Kochhar, Experimental investigation of the influence of air conduction on heat transfer across fibrous materials, *Journal of Mechanical Engineering Research*, 3(9), 2011, 319-324.

International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.1, Issue2, pp-564-569 ISSN: 2249-6645

Comparison between Treated and Untreated water so as to study water treatment plant of Ahmadpur dist. Latur,

Sayyed Hussain[a], Vinod Mane[b], Surendra Takde[a], Arif Pathan[c], Mazahar Farooqui[c,d].

a- Sir Sayyed College, Aurangabad(MS), India

b- Mahatma Gandhi Mahavidyala, Ahmadpur, Dist Latur

c- Post graduate and research centre, Maulana Azad College, Aurangabad (MS) India.

d- Dr Rafiq Zakaria College for Women, Aurangabad (MS) India.

Abstract:

In the present work we are reported the Physico chemical properties like pH, conductivity, Turbidity, TDS, DO, fluoride, chloride, Sodium, Sulphate , etc. and the values are compared for treated and untreated water samples. The samples were collected from treatment plant of Ahmadpur, Dist Latur. The values changes apparently after the treatment of water.

Keywords: Ahmadpur, Water treatment plant, treated and untreated water, physico chemical properties.

Introduction

Water is the unique component of nature has played the crucial role in the evolution of life from molecules to Water pollution may generally divided into three categories i.e. ground water pollution, surface water pollution, and sea water pollution. Surface water means generally water from rivers lake, ponds etc. Surface water comes in direct contact with the atmosphere, Seasonal streams, rivulets and surface drain so there occurs continues exchange of dissolved and atmospheric gases while the wastes are added through water conveyance. Recently US Department of Health Education and welfare (HEW) has classified surface water pollutants in to different categories i.e. sewage and waste, Industrial effluence, particulate and atmospheric gases, Infectious agents, minerals and chemical compounds, Dissolved toxic pollutants and chemical compounds, dissolved toxic pollutants and surface run off thermal pollutants. , Radioactive nuclides, organic chemical toxic.

In polluted surface water the ions like Na^{+}, K^{+}, mg^{++}, $So4^{--}$, H2Po4 interact forming a variety of complexes, there by deteriorating quality of the Precipitation the of surface water. Chemical processes like ion exchange, chelation, precipitation, coagulation, aggregation, oxidation, reduction and dissolution are operating simultaneously making the surface water extremely a complex system. The Physico – Chemical characteristics of water have direct impact on human beings. Hence the work was planed to investigate or assess the existing quality of take water (untreated) and municipal water (treated) which is supplied to urban area. The work is get

distributed in two parts i.e. in 1st part Physico-chemical analysis of take water observed and in 2nd part Physico-chemical parameters of treated water was analyzed which is purified by municipal corporation.

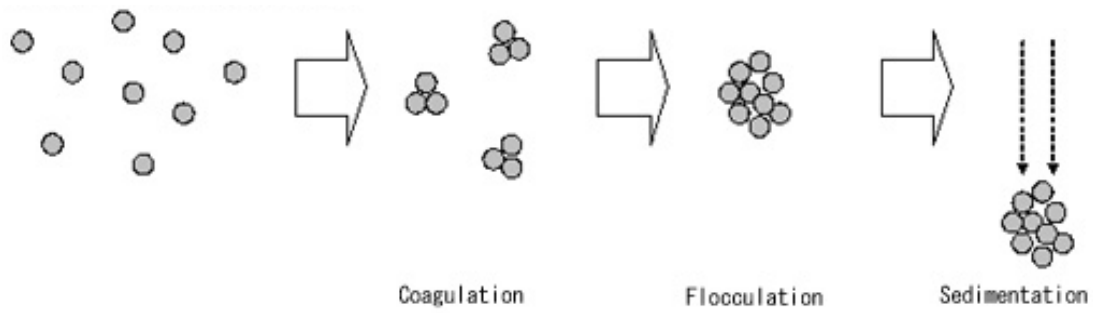
Materials and Methods

Here selection of sites was essential because, in the present investigation we have to analyze such water samples which are from lake and which must be treated by water supply department before supplying to urban population. Hence we select Limboti dam water for present investigation. Form Limboti dam, water is supplied to Ahmedpur city Limboti dam water was treated by Municipal Corporation of Ahmedpur. Municipal Corporation of Ahmedpur has a separate water purification plant, in around two acres of land.

Both water samples ware collected in the month of Jully-2010 and taken in pre cleaned polythene bottles. The collected samples were analyzed for measure physical and chemical water quality parameters like PH, TDS, T. Hardness, CA^{+2},$So4^{--}$,Cl^{-} and fluorides. The analyses were carried out as per methods described by APHA (1998) and NEERI (2007)

Municipal Water Treatment Plant of Ahmedpur city

Many water treatment plants use a combination of coagulation sedimentation, filtration and disinfection to provide clean, safe drinking Water to the public. Worldwide, a combination of coagulation, sedimentation and filtration is the most widely applied water treatment technology, and has been used since the early 20^{th} century.



Process of Coagulation, Flocculation and Sedimentation

The Municipal Water Treatment Plant of Ahmedpur city is based on the same combination i.e. coagulation, sedimentation, filtration and disinfection. Here Alum and Bleaching powder is used as Coagulant and Disinfectant respectively. A large sedimentation tank {fig. No 1} is made and sand filtration {fig.No.2} is used for filtration process.

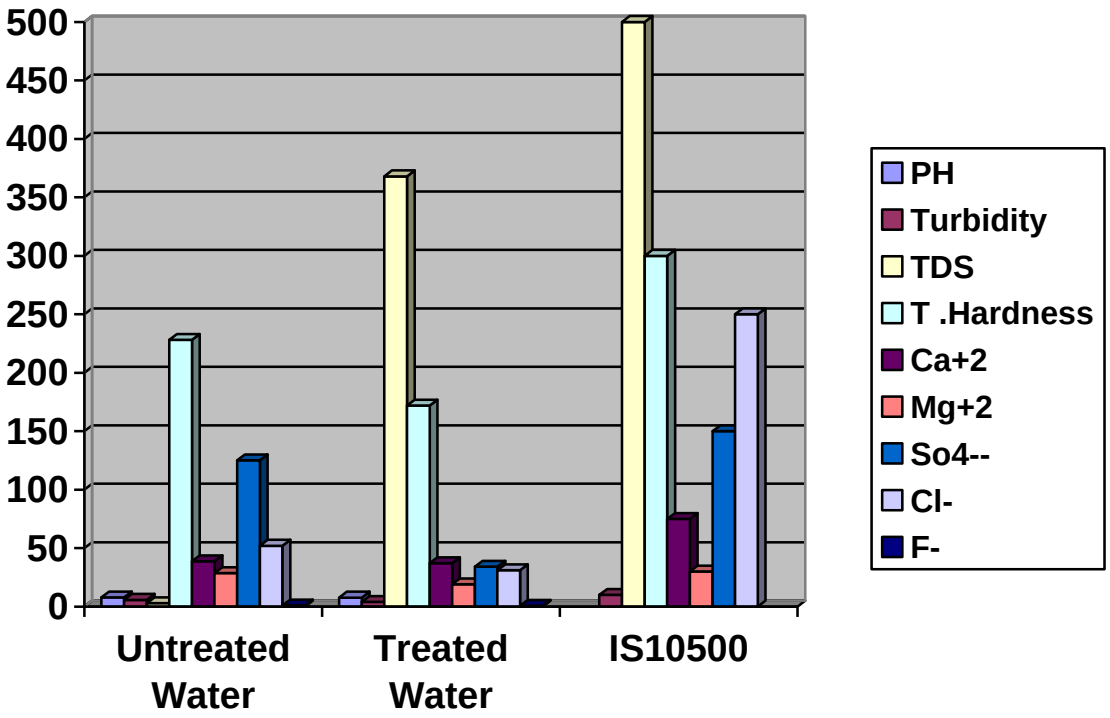
Fig: No.1 Sedimentation Tank

Fig No.2 sand Filtration

Parameters	Limboti Dam (s1) Water	Muncipal (s2) Water	IS10500
PH	7.8	7.6	6.5 to 8.5
Turbidity	5.7	4	10
TDS	228	168	500
T .Hardness	228	172	300
Ca^{+2}	38.7	37	75
Mg^{+2}	28.52	19	30

So_4^{--}	125	34	150
Cl^-	52	31	250
F^-	0.85	0.60	0.6 to 1.2

All parameters are in mg/L except pH and Turbidity, Turbidity in NTU.



Result and Discussion:

Results of Physico-chemical Characteristics of untreated and treated water were recorded and tabulated in table No. 1.

pH:

pH is the measurement of potential activity of hydrogen ions in the sample. pH was positively correlated with electrical conductance and total conductivity [7]. In the present study pH founds 7.8 in lake water sample and 7.6 in treated water sample.

Turbidity:

Turbidity of water observed was 5.7 NTU in Lake Water and 4 NTU in treated water. The difference in turbidities of untreated water and treated water shows the municipal water treatment plant works better in case of lowering turbidity

Total Dissolved solids:

TDS test measures the amount or particles that are dissolved in water [jay1]. The Indian standard (IS10500) is 500 mg /L. In the present investigation, we found 228 mg/L of TDS in lake water where as

168mg/L in treated water sample. S.A. Manjare et *al* [5] observed a range of 100 to 455 mg /L of TDS in Laxmiwadi water sample of Kolhapur.

Total Hardness:-

Prescribed limitation for Hardness and drinking water is 300 mg /L. In our investigation we observed 228 mg /L and 172 mg/ L of Hardness. T Nirmala [8] recommends to not to use water of station S_2 and S_3 of Theni district of Tamilnadu because they Found High concentration than permissible limit i. e. 700 and 1000 mg/L respectively.

Calcium and Magnetism

We found 38.7 and 37 mg/ L of calcium where as 28.52 and 19 mg/ L magnesium in lake and treated water respectively. R. Thangdurai [7] records a range of 11.80 to 20.15 mg/L of calcium in Lake Water sample of Samutharam Lake of Tiruvannamalai district Tamilnadu.

Sulfates:-

In the present investigation, it is observed that the sulfate concentration is much lower as compared to IS 10500. We noted 128 mg/ L in lake water and 34 Mg/ L in treated water. Raval and Malik also [6] also noted sulfate values within the range prescribed by are: 10500 in total 32 located water Samples around Surat city.

Chlorides:-

52 mg/L of chlorides found in untreated water sample where as 31 mg/L of chlorides found in treated water sample. M.Sangeeta et *al* reported 13.4 mg/L, 59.9 and 16.6 mg/L chloride in Vallur Village water sample.

Fluorides:

Intake of excess fluorides causes dental, skeletal and non skeletal fluorosis. Fluorosis has been considered as one of the incurable disease. Hence for prevention is the only solution for the disease [3]. In the present Study the fluoride Concentration was found 0.85 mg/L in untreated water and 0.60 mg/L in treated water Sample which is within the permissible limit.

Conclusion:-

The values of Physico chemical parameter of Dam Water are ranges some what parallel to the standards recommended by ISI. Hence it needs primary or some treatment before supplying to urban Area. The values municipal water Samples shows that the plant or the treatment given to the Dam water is proper and the treated water is fit for drinking.

References

1. **APHA (1998) -** Standard methods or examination of water and waste water treatment 20th Edn. N.W. Washington D.C.
2. **D.P.Gupta, Sunita, and J.P.Sharma-** Phy-Chem. Ana. Of Grd. Water 0f selected area of Kaithal city**.** Researcher 1 (2), 2009**.**
3. **Hem T.K.-Marine pollution in India, an engineering Problem Current Science (1992)**
4. **NEERI (2007) -** Guidance Manual for Drinking water Quality Monitoring and assessment (first Ed.) Pub. By: National Environmental, Engineering Research Institute, Nehru Marg, and Nagpur – 440020
5. **S.A.Manjare, S.A. Vahalankar and D.V. Muley (2010) -** Analysis of water quality using Physico – chemical parameters of Tamdalge Tank in Kolhapur District, Maharashtra Int. J. Adv. Biot. and Res. Vol. 1(2): pp 115 – 119.
6. **Raval and Malik (2010) -** Physico – chemical Characteristic of Ground water (drinking) in and around Surat city (India). J. Environ. Sci and Engg. 52 (4): 343 – 348.
7. **R. Thangdurai, K. Sivkumar and T.Ravimycin-** Monthly Var.in Samutharam Lake of Tiruvannamalai district Tamilnadu, .Asian J of Envir. Science June (2010) vol.5 No. `:19-22
8. **T. Nirmala (2010) -** Water quality assessment in Theni district Tamil Nadu, India J. Aqua. Biol. 25(1): 66 – 68.
9. **WHO (1988) -** International standards for drinking water quality vol. I, "Recommendations" World Health organization, Geneva, 130 P.

International Journal of Modern Engineering Research (IJMER)

www.ijmer.com Vol.1, Issue2, pp-570-579 ISSN: 2249-6645

Inhibition of Corrosion of Carbon steel by Thiourea – Zn^{2+} System in Natural Sea Water

Manivannan M[1*], Rajendran S[2, 3] and A. Krishnaveni[4]

[1]Department of Chemistry, Chettinad College of Engineering and Technology, Karur 639 114, Tamil Nadu, INIDA
[2]Corrosion Research Centre, PG and Research Department of Chemistry, GTN Arts College, Dindigul 624 005, Tamil Nadu, INDIA
[3]Department of Chemistry, RVS School of Engineering and Technology, Dindigul 624 005, Tamil Nadu, INDIA
[4]Department of Chemistry, Yadava College, Madurai 625 014, Tamil Nadu, INDIA

ABSTRACT

The inhibition efficiency (IE) of Thiourea (TU) in controlling corrosion of carbon steel in sea water in the absence and presence of Zn^{2+} has been evaluated by weight loss method. The formulation consisting of 200 ppm TU and 50 ppm Zn^{2+} has 95% IE. A synergistic effect exists between TU and Zn^{2+}. Synergism has been confirmed by Synergism parameter and F-Test. Influence of immersion period and pH on the IE of the inhibitor system has been investigated. AC Impedance study reveals that a protective film formed on the metal surface. The nature of the protective film has been characterized by scanning electron microscopy (SEM) and atomic force microscopy (AFM).

Keywords – **AFM, Carbon steel, F-Test, Sea water, TU**

1. INTRODUCTION

Corrosion can be defined as the deterioration of material by reaction to its environment. The corrosion occurs because of the natural tendency for most metals to return to their natural state. It cannot be avoided, but it can be controlled and prevented using the suitable preventive measures such as alloying, cathodic protection, anodic protection, protective coating and application of inhibitiors, etc. Among all these techniques inhibitors reduce the aggressiveness of the corrosive environment and forming a protective layer on the metal surface thereby the metal and alloys are prevented from corrosion. Especially inhibitors find application in nuclear power plant and engine cooling systems [1,2]. The organic compounds containing hetero atoms like oxygen, nitrogen, phosphorus and sulphur, etc have been employed as corrosion inhibitors to protect the metals from corrosion [3-7]. The corrosion inhibition of thiourea (TU) and its derivatives have been studied in various aqueous environments [8-12]. The presence of one sulphur and two nitrogen atoms containing lone pair of electrons in thiourea molecule makes it's as a very good corrosion inhibitor [13-15].

The present work is undertaken:

(i) to evaluate the inhibition efficiency (IE) of TU in controlling corrosion of carbon steel in sea water which is collected from Bay of Bengal at Marina Beach which is located at Chennai, Tamil Nadu, India (Table 1).

(i) to examine the influence of immersion period (IP) and pH on the IE of the TU – Zn^{2+} system.

(ii) to study the synergism using synergism parameters and analysis of variance.

(iii) to understand the mechanistic aspects of corrosion inhibition and formation of protective film on the metal surface by AC impedance spectra.

(iv) to analyze the protective film formed on the metals surface by scanning electron microscopy (SEM) and atomic force microscopy (AFM).

(v) to propose a suitable mechanism for corrosion inhibition process.

2. EXPERIMENTAL

2.1. Preparation of the specimens

Carbon steel specimens (0.026% S, 0.06% P, 0.4% Mn, 0.1% C and rest iron) of the dimensions 1.0 X 4.0 X 0.2 cm were polished to a mirror finish and

degreased with trichloroethylene and used for the weight loss method and surface examination studies.

2.2. Weight loss method

Carbon steel specimens in triplicate were immersed in 100 mL of the sea water containing various concentrations of the inhibitor in the presence and absence of Zn^{2+} for 1, 3, 5 and 7 days. The corrosion product cleaned with Clark's solution [16]. The parameter of the sea water is given in Table 1. The weights of the specimens before and after immersion were determined using an analytical balance, Shimadzu AY62 model. Then the Inhibition efficiency (IE) was calculated using the equation (1).

$$IE = 100\ [1 - (W_2 / W_1)]\ \% \qquad (1)$$

Where W_1 and W_2 are corrosion rate in the absence and presence of the inhibitor respectively. The corrosion rate (CR) was calculated using the equation (2).

$$CR = \frac{87.6\ W}{DAT}\ mm/y \qquad (2)$$

Table 1. Physico-Chemical Parameters of Sea Water

Parameters	Value
pH	7.66
Conductivity	44200 µmhos/cm
Chloride	16050 ppm
Sulphate	2616 ppm
TDS	30940 ppm
Total hardness	2800 ppm
Calcium	120 ppm
Sodium	6300 ppm
Magnesium	600 ppm
Potassium	400 ppm

Where W = weight loss in mg, D = density of carbon steel, 7.87 g/cm^3, A = surface area of the specimen (10 cm^2) and T = immersion period in hrs.

2.3 Synergism parameter

Synergism parameters are indications of synergistic effect existing between the inhibitors. S_I value is found to be greater than one suggesting that the existence of synergistic effect between the inhibitors [17-20]. The S_I value can be calculated using the formula (3).

$$S_I = \frac{1 - \theta_{1+2}}{1 - \theta'_{1+2}} \qquad (3)$$

Where $\theta_{1+2} = (\theta_1 + \theta_2) - (\theta_1\ \theta_2)$, $\theta = IE/100$, θ_1 = Surface coverage of inhibitor TU, θ_2 = Surface coverage of inhibitor Zn^{2+} and θ'_{1+2} = Combined inhibition efficiency of inhibitor TU and Zn^{2+}.

2.4 Analysis of Variance (F – Test)

F – Test was carried out to investigate whether synergistic effect existing between inhibitor systems is statistically significant [21,22]. If F – value is above 5.32 for 1, 8 degrees of freedom, it was proved to be at statistically significant. If it is below the value of 5.32 for 1, 8 degrees of freedom, it was statistically insignificant at 0.05 level of significance confirmed.

2.5 AC Impedance Spectra

AC Impedance study was carried out in Electrochemical Impedance Analyzer model CHI 660A using a three electrode cell assembly. The working electrode was used as a rectangular specimen of carbon steel with one face of the electrode of constant 1 cm^2 area exposed. A saturated calomel electrode (SCE) was used as reference electrode. A rectangular platinum foil was used as the counter electrode. AC impedance spectra were recorded after doing iR compensation. The real part (Z') and imaginary part (Z'') of the cell impedance were measured in ohms for various frequencies. The corrosion parameters such as charge transfer resistance (R_t) and double layer capacitance (C_{dl}) values were calculated. During the AC impedance spectra, the scan rate (V/s) was 0.005; Hold time at Ef (s) was zero and quiet time (s) was 2.

2.6 Surface Characterization Study

The carbon steel specimens were immersed in various test solutions for a period of one day. After one day the specimens were taken out and dried. The nature of the film formed on the metal surface was analyzed by surface characterization studies such as scanning electron microscopy (SEM) and atomic force microscopy (AFM).

2.6.1 Scanning Electron Microscopy (SEM)

The carbon steel specimens immersed in various test solutions for one day were taken out, rinsed with double distilled water, dried and subjected to the surface examination. The surface morphology measurements of the carbon steel surface were carried out by scanning electron microscopy (SEM) using HITACHI S-3000H SEM.

2.6.2 Atomic Force Microscopy (AFM)

The carbon steel specimens immersed in various test solutions for one day were taken out, rinsed with double distilled water, dried and subjected to the surface examination. The surface morphology measurements of the carbon steel surface were carried out by atomic force microscopy (AFM) using SPM Veeco diInnova connected with the software version V7.00 and the scan rate of 0.7Hz.

3. RESULTS AND DISCUSSION

3.1 Weight loss study

3.1.1. Influence of Immersion period on the IE of TU-Zn^{2+} system

The influence of immersion period on the IE of TU (200 ppm) – Zn^{2+} (50 ppm) system is given in Table 2. It is found that as the immersion period increases, the inhibition efficiency decreases. This is due to the fact that as the immersion period increases the protective film formed on the metal surface is unable to withstand the continuous attack of corrosive ions such as chloride (16050 ppm) present in sea water. There is a competition between two processes, namely, formation of $FeCl_2$ (and also $FeCl_3$) and iron – TU complex on the anodic sites of the metal surface.

It appears that the formation of iron chlorides is more favoured than the formation of iron - TU complex. Moreover, the iron - TU complex film formed on the metal surface is converted into iron chlorides which go into solution and hence, the IE decreases as the immersion period increases [23-24].

Table 2. Influence of immersion period on the inhibition efficiency of TU (200 ppm) and Zn^{2+} (50 ppm) system

System	Immersion Period (Days)			
	1	3	5	7
Sea water CR (mm/y)	0.1030	0.1124	0.1197	0.1247
Sea water + TU (200 ppm) + Zn^{2+} (50 ppm)	0.0030	0.0056	0.0083	0.0099
IE%	97	95	93	92

3.1.2. Influence of pH on the IE of TU-Zn^{2+} system

The influence of pH on inhibition efficiency of TU (200 ppm) – Zn^{2+} (50 ppm) system is given in Table 3. It is found that at (pH 8) the IE is 96 percent. When acid (dil H_2SO_4) is added to attain pH 6, the IE decreases to 94 percent. When NaOH solution is added to the boost the pH, (pH 10) IE decreases to 95 percent.

Table 3. Influence of pH on the inhibition efficiency of TU (200 ppm) and Zn^{2+} (50 ppm) system

Immersion Period: 3 days

System	pH			
	6	8	10	12
Sea water CR (mm/y)	0.1065	0.1035	0.1104	0.0978
Sea water + TU (200 ppm) + Zn^{2+} (50 ppm)	0.0063	0.0041	0.0055	0.0029
IE%	94	96	95	97

When NaOH is added further (pH 12) IE increases to 97 percent. It is found that at pH 8 the IE was 96 percent, when acid was added (pH 6), IE decreased. The protective film was brokened by H^+ ions of the acid. When NaOH was added further (pH 12) $Zn(OH)_2$ was solubilized as sodium zincate Na_2ZnO_2. Now TU was transported towards the metal surface. Hence IE increased [23-25].

3.2 Synergism Parameters (S_I)

The values of synergism parameters are given in Table 4 and 5. Here the values of S_I are greater than one, suggesting a synergistic effect. S_I approaches 1 when no interaction exists between the inhibitor compounds. When $S_I > 1$, this points to synergistic effects. In the case of $S_I < 1$, the negative interaction of inhibitors prevails (i.e. corrosion rate increases).

Table 4. Synergism parameters (SI) for carbon steel immersed in sea water in the absence and presence of inhibitor

Inhibitor system: TU + Zn^{2+} **IP:** 3 days

TU ppm	θ_1	θ_2 (Zn^{2+} = 25 ppm)	θ_{1+2}	θ'_{1+2}	S_I
50	0.10	0.12	0.20	0.30	1.13
100	0.16	0.12	0.18	0.45	1.47
150	0.34	0.12	0.41	0.64	1.61
200	0.50	0.12	0.56	0.80	2.20
250	0.62	0.12	0.66	0.75	1.33

From Table 4 and 5, it can be seen that the values of S_I are greater than unity, suggesting that the phenomenon of synergism existing between TU and Zn^{2+}. Also the synergism parameter (S_I) for the formulation consisting of 200 ppm of TU and 50 ppm of Zn^{2+} is 8.3, which is greater than one. Thus, the enhancement of the inhibition efficiency caused by the addition of Zn^{2+} to TU is only due to the synergistic effect.

Table 5. Synergism parameters (SI) for carbon steel immersed in sea water in the absence and presence of inhibitor

Inhibitor system: TU + Zn^{2+} **IP:** 3 days

TU ppm	θ_1	θ_2 (Zn^{2+} = 50 ppm)	θ_{1+2}	θ'_{1+2}	S_I
50	0.10	0.17	0.25	0.42	1.28
100	0.16	0.17	0.30	0.65	1.99
150	0.34	0.17	0.45	0.80	2.73
200	0.50	0.17	0.58	0.95	8.30
250	0.62	0.17	0.68	0.85	2.10

3.3 Analysis of Variance (ANOVA)

To investigate whether, the influence of Zn^{2+} on the inhibition efficiencies of TU is statistically significant, F – test was carried out. The results are given in Table 6. The results of Analysis of Variance (ANOVA) shows the influence of 25 ppm and 50 ppm of Zn^{2+} on the inhibition efficiencies of 50 ppm, 100 ppm, 150 ppm, 200 ppm and 250 ppm of TU. The obtained F – value 3.13 for 25 ppm of Zn^{2+}, is not statistically significant, since it is less than the critical F – value 5.32 for 1, 8 degrees of freedom at 0.05 level of significance.

Table 6. Distribution of F - value between the IE of various concentrations of TU-Zn^{2+} system.

Zn^{2+} (ppm)	SV	SS	DF	MS	F	LS
25	Between	288	1	288	3.13	$p < 0.05$
	Within	740	8	92		
50	Between	761	1	761	8.36	$p > 0.05$
	Within	728	8	91		

SV – Sources of Variance; SS – Sum of Squares DF – Degrees of Freedom; MS – Mean Square; LS – Level of Significance of F.

Therefore, it is concluded that the influence of 25 ppm of Zn^{2+} on the inhibition efficiencies of various concentrations of TU is not statistically significant. The obtained F – value 8.36 for 50 ppm of Zn^{2+}, is statistically significant, since it is greater than the critical F – value 5.32 for 1, 8 degrees of freedom at 0.05 level of significance. Therefore, it is concluded that the influence of 50 ppm of Zn^{2+} on the inhibition efficiencies of various concentration of TU is statistically significant.

3.4 Analysis of AC Impedance spectra

The AC impedance spectra of carbon steel immersed in sea water in the absence and presence inhibitors are shown in Fig. 1 to 3. The AC impedance parameters such as charge transfer resistance (R_t), double layer capacitance (C_{dl}) and impedance value [log (z/ohm)] are given in Table7. When carbon steel is immersed in sea water the corrosion potential is -731 mV vs saturated calomel electrode (SCE). The R_t value is 83.63 Ω cm^2 and C_{dl} value is 7.7500 x 10^{-6} µF/cm^2. When TU and Zn^{2+} are added to sea water, R_t value increases from 83.63 Ω cm^2 To 99.32 Ω cm^2. The C_{dl} value decreases from 7.7500 x 10^{-6} µF/cm^2 to 6.5273 x 10^{-6} µF/cm^2. This confirms that the formation of protective film on the metal surface. This accounts for the very high IE of TU – Zn^{2+} system. This is further supported by the increase in impedance value [log (z/ohm)] from 2.011 to 2.050 [26-29].

Table 7. Impedance parameters for corrosion of carbon steel immersed in sea water in the absence and presence of inhibitors obtained by AC impedance spectra.

System	R_t Ω cm^2	C_{dl} µF/cm^2	log (z/ohm)
Sea water	83.63	7.7500 x 10^{-6}	2.011
Sea water + TU (200 ppm) + Zn^{2+} (50 ppm)	99.32	6.5273 x 10^{-6}	2.050

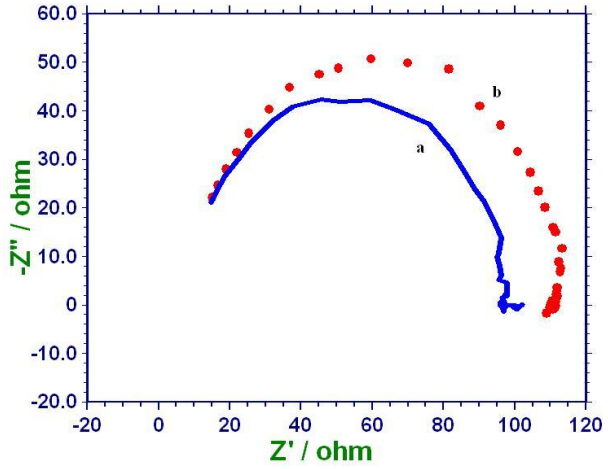


Fig. 1. AC Impedance spectra of carbon steel immersed in various test solutions (Nyquist plots).

(a) Sea water
(b) Sea water + TU (200 ppm) + Zn^{2+} (50 ppm)

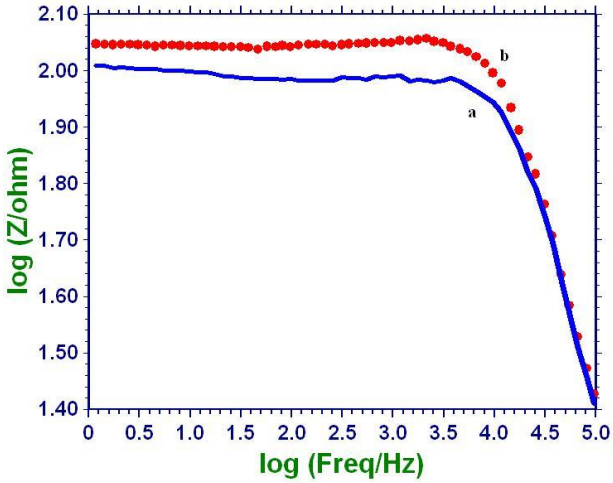


Fig. 2. AC Impedance spectra of carbon steel immersed in various test solutions (Frequency Bode plots).

(a) Sea water
(b) Sea water + TU (200 ppm) + Zn^{2+} (50 ppm)

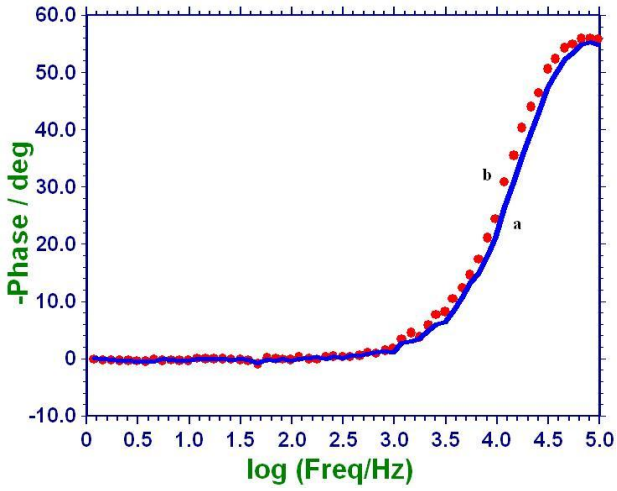


Fig. 3. AC Impedance spectra of carbon steel immersed in various test solutions (Phase Bode plots)

(a) Sea water
(b) Sea water + TU (200 ppm) + Zn^{2+} (50 ppm)

3.5 Scanning Electron Microscopy (SEM)

SEM provides a pictorial representation of the metal surface. To understand the nature of the surface film in the absence and presence of inhibitors and the extent of corrosion of carbon steel, the SEM micrographs of the surface are examined. The SEM micrograph (X 1000) of a polished carbon steel surface (control) in Fig. 4. (a) shows the smooth surface of the metal. This shows the absence of any corrosion products or inhibitor complex formed on the metal surface. The SEM micrograph (X 1000) of carbon steel specimen immersed in the sea water for one day in the absence and presence of inhibitor system is shown in Fig. 4. (b) and (c) respectively.

The SEM micrograph of carbon steel surface immersed in sea water in Fig. 4. (b) shows the roughness of the metal surface which indicates the corrosion of carbon steel in sea water. The Fig.4. (c) indicates that in the presence of 200 ppm TU and 50 ppm Zn^{2+} mixture in sea water, the surface coverage increases which in turn results in the formation of insoluble complex on the metal surface. In the presence of TU and Zn^{2+}, the surface is covered by a thin layer of inhibitors which effectively control the dissolution of carbon steel [12, 30-32].

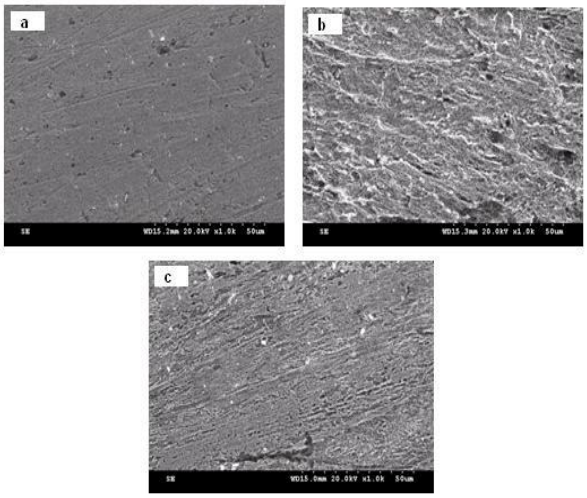


Fig. 4. SEM micrographs of carbon steel surface

(a) Polished carbon steel (control)
(b) Carbon steel immersed in sea water
(c) Carbon steel immersed in sea water containing TU (200 ppm) and Zn^{2+} (50 ppm)

3.6 Atomic Force Microscopy (AFM)

Atomic force microscopy is a powerful technique for the gathering of roughness statistics from a variety of surfaces [33,34]. AFM is becoming an accepted method of roughness investigation [35-39]. All AFM images were obtained using SPM Veeco diInnova AFM instrument operating in contact mode in air. The scan size of all the AFM images is 4.91 µm x 4.91 µm areas at a scan rate of 0.7 Hz. The two dimensional (2D), three dimensional (3D) AFM morphologies and the AFM cross-sectional profile for polished carbon steel surface (reference sample), carbon steel surface immersed in sea water (blank) and carbon steel surface immersed in sea water containing the formulation of 200 ppm of TU and 50 ppm of Zn^{2+} are shown in Fig. 5. (a, d, g), (b, e, h), (c, f, i) respectively.

3.7.1 Root-mean-square roughness, average roughness and peak-to-valley height

AFM images analysis was performed to obtain the average roughness, R_a (the average deviation of all points roughness profile from a mean line over the evaluation length), root-mean-square roughness, R_q (the average of the measured height deviations taken within the evaluation length and measured from the mean line) and the maximum peak-to-valley (P-V)

height values (largest single peak-to-valley height in five adjoining sampling heights) [40]. R_q is much more sensitive than R_a to large and small height deviations from the mean [41]. The summary of the average roughness (R_a), rms roughness (R_{RMS}) and maximum peak-to-valley height (P-V) value for carbon steel surface immersed in various test solutions are given in Table 8.

Table 8. AFM data for carbon steel surface immersed in inhibited and uninhibited environments

Samples	(R_a) Average Roughness (nm)	(R_q) RMS Roughness (nm)	Maximum peak-to-valley (P-V) height (nm)
Polished carbon steel (control)	5.6241	8.1069	44.40
Carbon steel immersed in sea water	32.9000	40.2000	140.60
Carbon steel immersed in sea water containing TU (200 ppm) and Zn^{2+} (50 ppm)	9.0758	12.0118	67.29

The value of R_a, R_q and P-V for the polished carbon steel surface (reference sample) are 5.6241 nm, 8.1069 nm and 44.40 nm respectively, which shows a more homogeneous surface, with some places in where the height is lower than the average depth [42]. The Fig. 5 (a, d, g) displays the uncorroded metal surface. The slight roughness observed on the polished carbon steel surface is due to atmospheric corrosion. The average roughness, root-mean-square roughness and P-V height values for the carbon steel surface immersed in sea water are 32.9000 nm, 40.2000 nm and 140.60 nm respectively. These data suggests that carbon steel surface immersed in sea water has a greater surface roughness than the polished metal surface, which shows that unprotected carbon steel surface is rougher and was due to the corrosion of the carbon steel in sea water. The Fig. 5 (b, e, h) displays corroded metal surface with few pits.

The presence of 200 ppm of Thiourea and 50 ppm of Zn^{2+} in sea water reduced the R_{RMS} by a factor of 3.34 (12.0118 nm) from 40.2000 nm and the average roughness is significantly reduced to 9.0758 nm when compared with 32.9000 nm of carbon steel surface immersed in sea water. The maximum peak-to-valley height also was reduced to 67.29 nm. These parameters confirm that the surface appears smoother. The smoothness of the surface is due to the formation of a compact protective film of Fe^{2+} - TU complex and $Zn(OH)_2$ on the metal surface thereby inhibiting the corrosion of carbon steel. Also the above parameter observed are somewhat greater than the AFM data of polished metal surface which confirms the formation of the film on the metal surface, which is protective in nature.

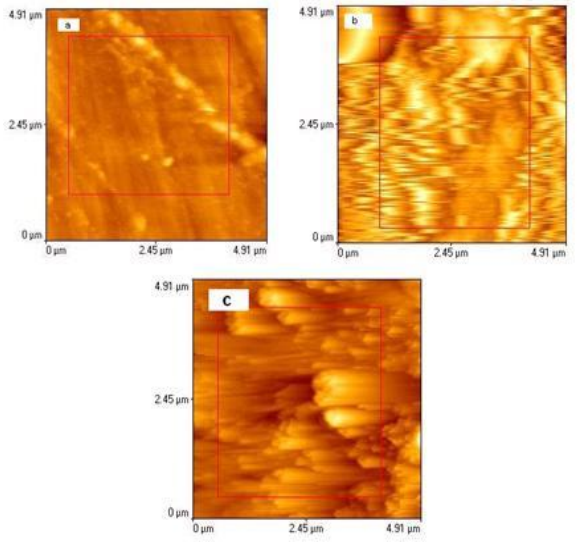


Fig. 5. 2D AFM images of carbon steel surface

(a) Polished carbon steel (control)
(b) Carbon steel immersed in sea water (blank)
(c) Carbon steel immersed in sea water containing TU (200 ppm) and Zn^{2+} (50 ppm)

www.ijmer.com Vol.1, Issue2, pp-570-579 ISSN: 2249-6645

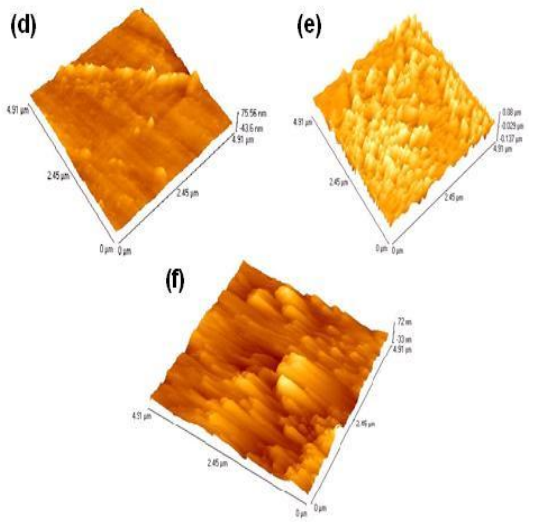


Fig. 5. 3D AFM images of carbon steel surface

(d) Polished carbon steel (control)
(e) Carbon steel immersed in sea water (blank)
(f) Carbon steel immersed in sea water containing TU (200 ppm) and Zn^{2+} (50 ppm)

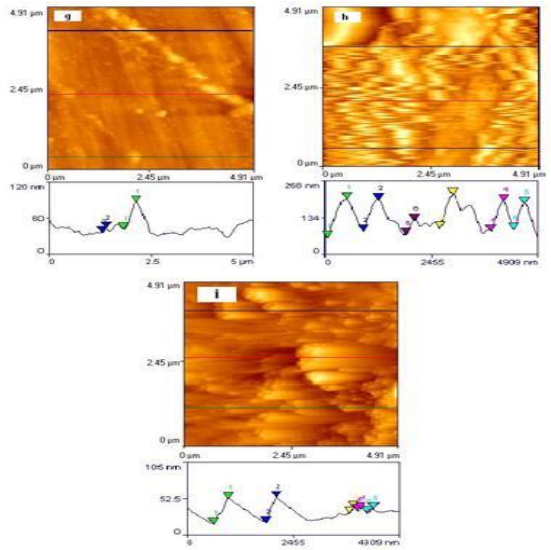


Fig. 5. The cross-sectional profiles which are corresponding to as Shown broken lines (black colour) in AFM images of carbon steel surface.

(g) Polished carbon steel (control)
(h) Carbon steel immersed in sea water (blank)
(i) Carbon steel immersed in sea water containing TU (200 ppm) and Zn^{2+} (50 ppm)

4. MECHANISM OF CORROSION INHIBITION

Based on the above studies the following mechanism can be proposed for corrosion inhibition process.

(i) Zn^{2+} - TU + $Fe^{2+} \rightarrow Fe^{2+}$ - TU + Zn^{2+}

(ii) $Zn^{2+} + 2\ OH^- \rightarrow Zn(OH)_2 \downarrow$

(iii) Protective film consists of Fe^{2+} - TU complex and $Zn(OH)_2$

(iv) At anode: $Fe \rightarrow Fe^{2+} + 2\ e^-$
At cathode: $H_2O + ½\ O_2 + 2\ e^- \rightarrow 2\ OH^-$
$Fe^{2+} + TU \rightarrow Fe^{2+}$ - TU complex
$Zn^{2+} + 2\ OH^- \rightarrow Zn(OH)_2 \downarrow$

(v) It accounts for the synergism of TU-Zn^{2+}.

5. CONCLUSION

The present study leads to the following conclusions:

(i) The formulation consisting of 200 ppm of TU and 50 pm of Zn^{2+} offers 95% IE to carbon steel immersed in sea water.
(ii) Synergistic effect exists between TU and Zn^{2+}.
(iii) AC impedance spectra reveal that the formation of protective film on the metal surface.
(iv) SEM and AFM studies confirm the formation of protective film on the metal surface and hence the corrosion process is inhibited.

ACKNOWLEDGEMENTS

The authors are thankful to their respective managements and Defence Research and Development Organization (DRDO), New Delhi, India.

REFERENCES

[1] Mars G. Fontana, *Corrosion Engineering* (Tata McGraw Hill Education Private Limited, New Delhi, 2005).
[2] Pierre R. Roberge, *Handbook of Corrosion Engineering* (Mc-Graw Hill, 2000)
[3] J.G.N. Thomas, Some New Fundamental Aspects in Corrosion Inhibition: *Proc. 5th European symposium on corrosion inhibitors*, Ferrara, Italy, University of Ferrara, 1981, 453.

[4] B.D. Donnelly, T.C. Downie, R. Grzeskowaik, H.R. Hamburg and D. Short, *Corrosion Science, 38,* 1997, 109.
[5] A.B. Tadros and Y. Abdel-Naby, *Journal of Electroanalytical Chemistry, 224,* 1988, 433.
[6] N.C. Subramanyam, B.S. Shesadri and S.M. Mayanna, Thiourea and substituted thioureas as corrosion inhibitors for aluminium in sodium nitrite solution, *Corrosion Science, 34(4),* 1993, 563-571.
[7] W. Kautex, *Corrosion Science, 28,* 1988, 173.
[8] J.C. Lin, S.L. Chang and S.L. Lee, Corrosion inhibition of steel by thiourea and cations under incomplete cathodic protection in a 3.5% NaCl solution and sea water, *Journal of Applied Electrochemistry, 29 (8),* 1999, 911-918.
[9] S.S. El-Egamy, Corrosion and corrosion inhibition of Cu-20% Fe alloy in sodium chloride solution, *Corrosion Science, 50 (4),* 2008, 928-937.
[10] S.M.A. Hosseini and S. Salari, Corrosion inhibition of stainless steel 301 by 1-methyl-3-pyridine-2-Y1-thiourea in acidic media, *Indian Journal of Chemical Technology, 16,* 2009, 480-485.
[11] S. Divakara Shetty, Prakash Shetty and H.V. Sudhaker Nayak, Inhibition of corrosion of mild steel in hydrochloric acid by N-cyclohexyl-N-phenyl thiourea, *Indian Journal of Chemical Technology, 12,* 2005, 462-465.
[12] P.K. Gogoi and B. Barhai, Corrosion Inhibition of Carbon Steel in Open Recirculating Cooling Water System of Petroleum Refinery by Thiourea and Imidazole in Presence of Zinc (II) Sulphate, *International Journal of Chemistry, 2 (2),* 2010, 138-143.
[13] M. Manivannan and S. Rajendran, Thiourea – Zn^{2+} system as corrosion inhibitor for Carbon Steel in Marine media, *Journal of Chemical, Biological and Physical sciences, 1 (2),* 2011, 241-249.
[14] M.N. Desai, G.H. Thanki and M.H. Gandhi, Thiourea and its derivatives as corrosion inhibitors, *Anticorrosion,15 (7),* 1968, 12-16.
[15] M.A. Quarishi, D. Jamal and R.N. Singh, Inhibition of mild steel corrosion in the presence of fatty acid thiosemicarbazides, *Corrosion, 58 (3),* 2002, 201.
[16] G. Wranglen, *Introduction to Corrosion and Protection of Metals* (Chapman and Hall, London, 1985) 236.
[17] Benita Sherine, A. Jamal Abdul Nasser and S. Rajendran, Inhibitive action of hydroquinone – Zn^{2+} system in controlling the corrosion of carbon steel in well water, *International Journal of Engineering Science and Technology, 2 (4),* 2010, 341-357.
[18] S. Rajendran, S. Shanmuga priya, T. Rajalakshmi and A. John Amalraj, Corrosion inhibition by an aqueous extract of Rhizome powder, *Corrosion, 61,* 2005, 685-692.
[19] S. Agnesia Kanimozhi and S. Rajendran, Inhibitive properties of sodium tungstate – Zn2+ system and its synergism with HEDP, *International Journal of Electrochemical Science, 4,* 2009, 353-368.
[20] K. Anuradha, R. Vimala, B. Narayanasamy, J. Arockia Selvi and S. Raji, Corrosion inhibition of carbon steel in low chloride media by an aqueous extract of Hibiscus Rosasinensis Linn, *Chemical Engineering Communication, 195,* 2008, 352-366.
[21] S. Rajendran, A. Raji, J. Arockia Selvi, A. Rosaly and Thangasamy, Evaluation of Gender Bias in use of Modular instruction and Concepts of Organic chemistry Nomenclature, *Journals of Material Education, 29,* 2007, 245-258.
[22] S. Rajendran, A. Raji, J. Arockia Selvi, A. Rosaly and Thangasamy, Parents Education and Achievement Scores in Chemistry, *EDUTRACKS, 6,* 2007, 30-33.
[23] S. Agnesia Kanimozhi and S. Rajendran, Realization of synergism in sodium Tungstate – Zn^{2+} - N-(Phosphenomethyl) iminodiacetic acid system in well water, *The Open Corrosion Journal, 2,* 2009, 166-174.
[24] T. Umamathi, J. Arockia Selvi, S. Agnesia Kanimozhi, Susai Rajendran and A. John Amalraj, Effect of Na_3PO_4 on the corrosion inhibition efficiency of EDTA – Zn^{2+} system for carbon steel in aqueous solution, *Indian Journal of Chemical Technology, 15,* 2008, 560-565.
[25] R. Kalaivani, B. Narayanasamy, J.Arockia Selvi, A. John Amalraj, J. Jeyasundari and S. Rajendran, Corrosion inhibition by Prussian blue, *Portugaliae Electrochimica Acta, 27 (2),* 2009, 177-187.
[26] S. Shanthi, J. Arockia Selvi, S. Agnesia Kanimozhi, S. Rajendran, A. John Amalraj, B. Narayanasamy and N. Vijaya, Corrosion behaviour of carbon steel in the presence of iodide ion, *Journal of Electrochemical Society of India, 56 ½ ,* 2007, 48-51.
[27] X. Joseph Raj and N. Rajendran, Corrosion inhibition effect of substituted thiadiazoles on brass, *International Journal of Electrochemical Science, 6,* 2011, 348-366.
[28] Ahmed Y. Musa, Abdul Amir H. Kadhum, Mohd Sobri Takriff, Abdul Razak Daud and Siti Kartom Kamarudin, *Modern Applied Science, 3 (4),* 2009, 90.
[29] Susai Rajendran, M. Manivannan, J. Wilson Sahayaraj, J. Arockia Selvi, J. Sathiyabama, A. John Amalraj and N. Palaniswamy, Corrosion Behaviour

of Aluminium in Methyl orange solution at pH 11, *Transactions of the SAEST, 41,* 2006, 63-67.

[30] Xu Yang, Fu Sheng Pan and Ding Fei Zhang, *Materials Science forum, 610-613,* 2009, 920-926.

[31] A. Hamdy, A.B. Farag, A.A. EL-Bassoussi, B.A. Salah and O.M. Ibrahim, Electrochemical behaviour of brass alloy in different saline media: Effect of benzotriazole, *World Applied Sciences Journal, 8 (5),* 2010, 565-571.

[32] Weihua Li, Lichao Hu, Shengtao Zhang and Baorong Hou, Effects of two fungicides on the corrosion resistrance of copper in 3.5% NaCl solution under various conditions, *Corrosion Science, 53 (2),* 2011, 735-745.

[33] S. Rajendran, B.V. Appa Rao and N. Palaniswamy, *Proc. 2nd Arabian Corrosion Conference,* Kuwait, 1996, 483.

[34] J. Arockia Selvi, Susai Rajendran and J. Jeyasundari, Analysis of nanofilms by atomic force microscopy, *Zastita Materijala, 50 (2),* 2009, 91-98.

[35] R. Vera, R. Schrebler, P. Cury, R. Rio and H. Romero, Corrosion protection of carbon steel and copper by polyaniline and poly (ortho-methoxyaniline) films in sodium chloride medium: Electrochemical and morphological study, *Journal of Applied Electrochemistry, 37 (4),* 2007, 519-525.

[36] Ph Dumas, B. Bouffakhreddine, C. Amra, O. Vatel, E. Andre, R. Galindo and F. Salvan, Quantitative Microroughness Analysis down to the nanometer scale, *Europhysics Letters, 22 (9),* 1993, 717-722.

[37] J.M. Bennett, J. Jahanmir, J.C. Podlesny, T.L. Baiter and D.T. Hobbs, Scaning Force Microscopy as a tool for Studying Optical Surfaces, *Applied Optics, 34 (1),* 1995, 213-230.

[38] A. Duaparre, N. Kaiser, H. Truckenbrodt, M.R. Berger and A. Kohler, Microtopography Investigation of Optical surface and Thin films by Light scattering, Optical profilometry and Atomic Force Microscopy, International Symposium on Optics Imaging and Instrumentation, *Proc. SPIE-1995,* San Diego, CA, 1993, 181-192.

[39] A. Duaparre, N. Kaiser and S. Jakobs, Morphology Investigation by Atomic Force Microscopy of Thin films and substances for Excimer laser mirrors, Annual Symposium on Optical Materials for High Power Lasers, *Proc. SPIE-2114,* Boulder, CO, 1993, 394.

[40] C. Amra, C. Deumie, D. Torricini, P. Roche, R. Galindo, P. Dumas and F. Salvan, Overlapping of Roughness spectra measured in microscopic (optical) and microscopic (AFM) bandwidths, International Symposium on Optical Interference Coatings, *Proc. SPIE-2253,* 1994, 614-630.

[41] T.R. Thomas, *Rough Surface* (Longman, New York, 1982).

[42] K.J. Stout, P.J. Sullivan and P.A. Mc Keown, The use of 3D Topographic analysis to determine the Microgeometric Transfer characteristics of Textured sheet surfaces through rolles, *Annals CRIP, 41,* 1992, 621.