International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.3, Issue.1, Jan-Feb. 2013 pp-547-550 ISSN: 2249-6645

A DATA MINING ANALYSIS & APPROACH WITH INTRUSIONDETECTION / PREVENTION FROM REAL

Meenakshi.RM,[1] Mr.E.Saravanan[2]
[1]*Final Year Student, M.Tech CSE Department, Dr.M.G.R.Educational and Research Institute University, Tamil Nadu, India*
[2]*Assistant Professor, Dr.M.G.R.Educational and Research Institute University, Tamil Nadu, India*

***Abstract*:** *We propose a mechanism for false positive/negative assessment with multiple IDSs/IPSs to collect FP and FN cases from real-world traffic and statistically analyze these cases. False positives and false negatives happen to every intrusion detection and intrusion prevention system. IDSs/IPSs can identify a normal activity as malicious one, causing a false positive (FP) or malicious traffic as normal, causing a false negative (FN) .To create a pool of traffic traces causing possible FPs and FNs to IDSs using Attack Session Extraction (ASE).Statistically analyze the packet by preprocessing based on protocol for each layer .Based on thatBinary classifiers are generated for each class of event using relevant features for the class using classification algorithm .Binary classifiers are derived from the training sample by considering all classes other than the current class .Analyzing the KDD data set for pattern matching and the effect of combining different classifiers can be explained with the theory of bias-variance decomposition using multi boosting.*

Index Terms:*False Positive, False Negative, Intrusion Detection/Prevention, Packet Trace, Network Measurement, Peer to Peer Traffic Analysis.*

I. Introduction

Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDPS for other purposes, such as identifying problems with security policies, documenting existing threats, and deterring individuals from violating security policies. Data mining based intrusion detection algorithms aim to solve the problems of analyzing the huge volumes of audit data and realizing performance optimization of detection rules. An IDS/IPS monitors the activities of a given environment and decides whether these activities are malicious or normal based on system integrity, confidentiality and the availability of information resources. As soon as a malicious or an intrusive event is detected, the IDS produces a relative alert and passes it to the network administrator promptly while the IPS not only executes what the IDS does but also blocks network traffic from the suspected malicious source. FPs and FNs cause several problems. For example, FNs generate unauthorized or abnormal activities on the Internet or in computer systems. On the other hand, a lot of FPs may easily conceal real attacks and thus overwhelm the security operator. When real attacks occur true positives (real alerts) are deeply buried within FPs, so it is easy for the security operator to miss them.

The existing work monitor and analyze system calls, application logs, file-system modifications using communication protocol between a connected device .Drawback is the hackers recover the embedding data in original image because the data placed in particular bit position.

II. Overview of the Approach

The aim of the proposed system is to monitor and analyze real network traffic in a computer network like a network sniffer and collects network logs. Then the collected network logs are analyzed for rule violations by using data mining algorithms. When any rule violation is detected,the overall approach of the proposed system is clearly portrait in the flow chart diagram “Fig 1”. The steps involved in this system are

1. Receive the network packet and extract the attributes by protocols HTTP,FTP,SMTP etc;
2. Transmit raw packet to the network and gather statistical information on the realnetwork.
3. Binary classifiers are generated for each class of event using relevant features for the class and classification Algorithm.
4. Binary classifiers are derived from the training sample by considering all classes other than the current class.
5. The main purpose is to select different features for different classes by applying the information gain or gain ratio in order to identify relevant features for each binary classifier.
6. The effect of combining different classifiers can be explained with the theory of bias-variance decomposition.

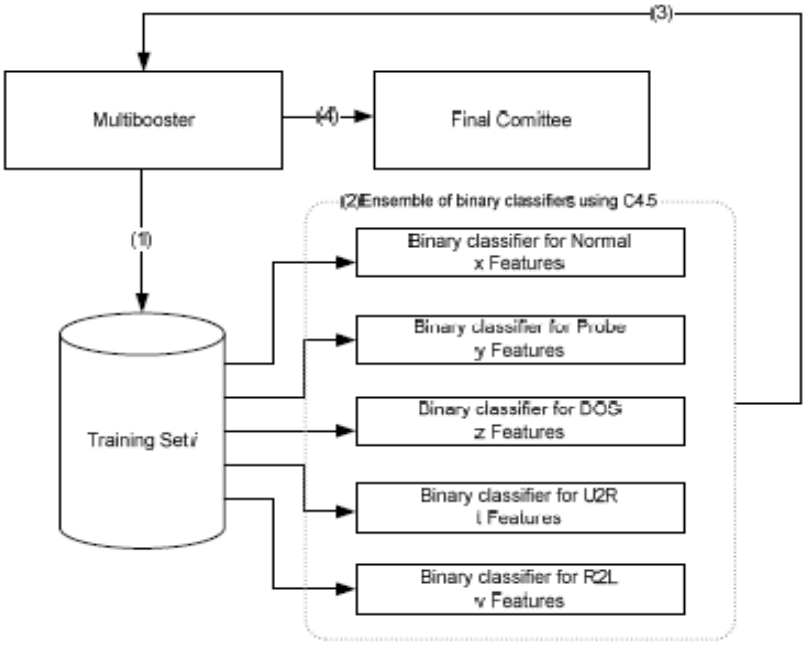

Fig 1 flow chart of proposed system

III. False Positive /Negative Assessment

FP and FN rates are two important metrics in measuring the accuracy of a network security system, such as an IDS or IPS. It has been demonstrated that even a small rate (1 in 10,000) of FPs could generate an unacceptable number of FPs in practical detections. The assessment is important to IDS/IPS developers trying to optimize the accuracy of detection by reducing both FPs and FNs, because the FP/FN rate limits the performance of network security systems due to the base-rate fallacy phenomenon. The statistical analyses in this work can elucidate the causes and rankings of FPs and FNs, thus allowing developers to avoid similar pitfalls during their product development.

Second, after detecting the potential FPs/FNs/TPs/TNs, this work replays the extracted packet trace according to the log of the DUTs (Device Under test)again. This step is called trace verification because it verifies whether this case is reproducible to the original DUTs. This case is a reproducible FP/FN/TP/TN when it meets the following two conditions.

I. For any DUT, it must produce an alert if it did last time.
II. The two alerts must be the same when one came from some DUT last time and the other is produced by the same DUT this time.

IV. C 4.5 Classification Algorithm

In general, steps in C4.5 algorithm to build Binary tree are:

1. Choose attribute for root node
2. Create branch for each value of that attribute
3. Split cases according to branches
4. Repeat process for each branch until all cases in the branch have the same class

Choosing which attribute to be a root is based on highest gain of each attribute. To count the gain, we use formula 1,below:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(Si) \quad (1)$$

With:
{S1,..., Si, …, Sn} = partition
s of S according to values of attribute A
n = number of attributes A
|Si| = number of cases in the partition Si
|S| = total number of cases in S

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log_2 pi \quad \text{.....(2)}$$

With:
S: Case Set
N: number of cases in the partition S
Pi: Proportion of Si to S

V. Experimental Setup

Our proposed model is implemented by LAN and tested. An examplethe entire incoming packet processed in LAN. With necessary details about the packets in the fig(1) respectively.

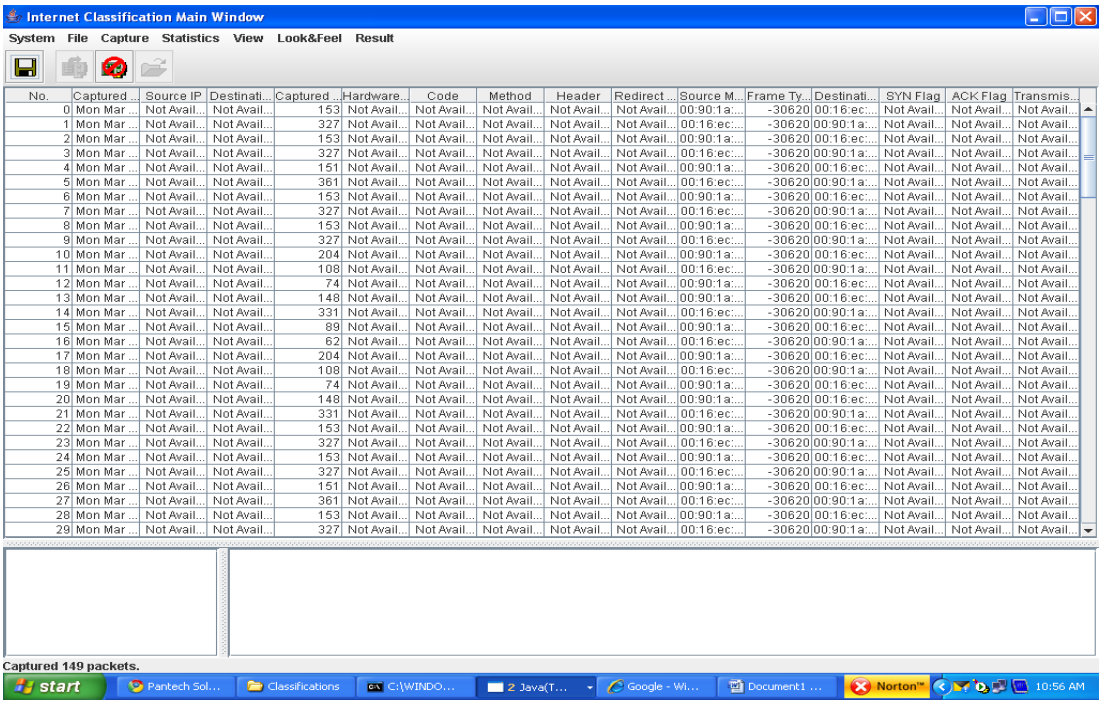

Fig 1:Processing of incoming packet in the LAN

During preprocessing cumulative pie graph for network layer protocol ratio measurements is shown in the fig(2) respectively. Each and every variation in the packet will reflect in the graph automatically.

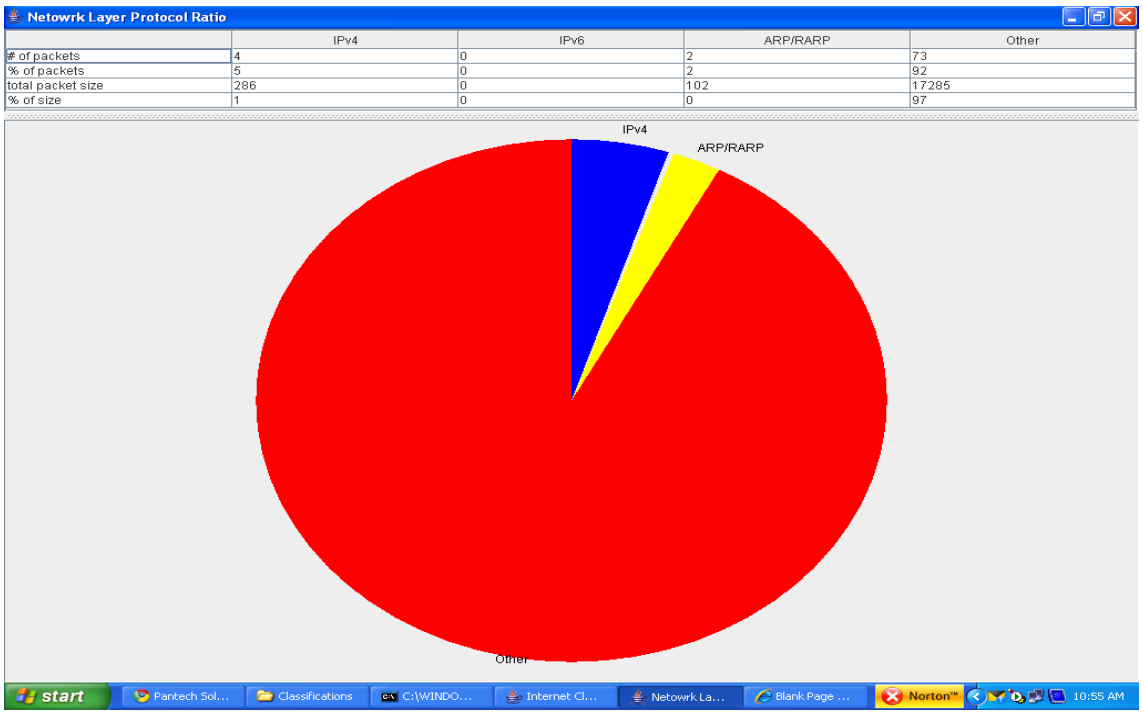

Fig 2: Network layer protocol ratio during preprocessing

During preprocessing continous graph for network layer protocol ratio is shown in the fig (3) respectively

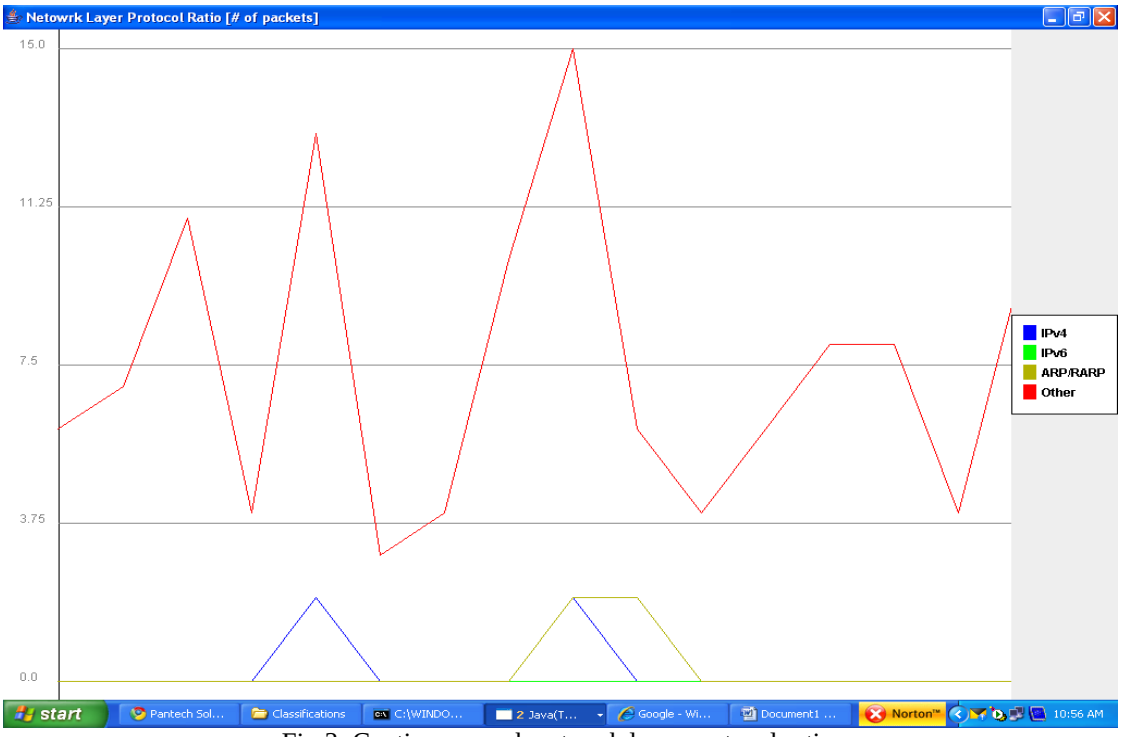

Fig 3: Continous graph network layer protocol ratio

VI. Conclusion

In this paper, we first propose a false positive/false negative assessment with C 4.5 classification algorithm based on binary tree. Packet has been preprocessed in the LAN. C4.5 binary classification done by pattern matching with KDD dataset.Multi boosting technique is used to check the features of the data. The advantages of our proposed system is

1. Implementation of C 4.5 classification algorithm is more efficient when compared to decision tree.
2. If some problem or attack occurs we can easily find out the location where the attack occurs exactly.
3. If frequent message or DOS attack occur we can logoff theremote machine.
4. Alert message in case of problem.
5. Multi boosting is data mining technique for checking the algorithm
6. For each operating system different applications have to be used, regardless they are doing exactly the same.

Finally, although our model focuses on FP/FN assessment based on preprocessing ,binary classifiers,Analysing the data set and multi boosting.The model is designed for dataset for all the features of datain attribute relative file format.Our result based on client operating system.Server operarting systemwill be studied in our future enhancements.

References

[1] H. G. Kayacık and A. N. Zincir-Heywood, “Using Intrusion Detection Systems with a Firewall: Evaluation on DARPA 99 Dataset”, Project in Dalhousie University, [Online]. Available:http://projects.cs.dal.ca/projectx/files/NIMS06-2003.pdf.
[2] DARPA 99 Intrusion Detection Data Set Attack Documentation. [Online]. Available: http://www.ll.mit.edu/IST/ideval/docs/1999/attackDB.html.
[3] V. Corey, C. Peterman, S. Shearin, M. S. Greenberg, J. V. Bokkelen, "Network Forensics Analysis," IEEE Internet Computing, vol.06, no.6, pp. 60-66, 2002.
[4] W. D. Yu, D. Aravind, P. Supthaweesuk, "Software Vulnerability Analysis for Web Services Software Systems," iscc, pp. 740-748, 11th IEEE Symposium on Computers and Communications (ISCC'06), 2006.
[5] M. Bailey, E. Cooke, F. Jahanian, D. Watson, Jose Nazario, "The Blaster Worm: Then and Now," IEEE Security and Privacy, vol. 03, no. 4, pp. 26-31, 2005.
[6] C. L. Schuba, I. V. Krsul, M. G. Kuhn, E. H. Spafford, A. Sundaram, D. Zamboni, "Analysis of a Denial of Service Attack on TCP," sp, p. 0208, 1997 IEEE Symposium on Security and Privacy, 1997.
[7] V. Paxson, “An analysis of using reflectors for distributed denial-of-service attacks” ACM SIGCOMM Computer Communication Review, 2001.
[8] M. Roesch, “Network Security: Snort - Lightweight Intrusion Detection for Networks”, Proceedings of the 13th USENIX conference on System administration, November. 1999.