IJRRECS/October 2013/Volume-1/Issue-6/1514-1524 **ISSN 2321-5461**

INTERNATIONAL JOURNAL OF REVIEWS ON RECENT ELECTRONICS AND COMPUTER SCIENCE

AN INTRUSION DETECTION SYSTEM FOR REAL NETWORK TRAFFIC CLASSIFICATION

P.Bharani Padma[1], P.Kiran Sree[2]

[1]M.tech, Dept of CSE, Bonam Venkata Chalamaiah (BVC) Engineering Collegeodelarevu, Amalapuram, Andhra Pradesh Email Id: bujjubharani@gmail.com

[2]Professor, Dept of CSE, Bonam Venkata Chalamaiah (BVC) Engineering Collegeodelarevu, Amalapuram, Andhra Pradesh Email Id: profkiran@yahoo.com

ABSTRACT:

An intrusion detection system (IDS) may be a device or software package application that monitors network or system activities for malicious activities or policy violations and produces reports to a management station. organizations use IDPS for different functions, like characteristic issues with security policies, documenting existing threats, and detecting people from violating security policies. Accordingly, once this application meets an IDS/IPS with strict detection rules, its traffic will be thought to be malicious traffic, leading to a lot of Federal Protective Service. Second, regarding ninety one % of FP alerts, equal to regarding eighty five % of false cases, square measure not related to security problems, however to management policy. For example, some firms and campuses limit or forbid their workers and students from victimization peer-to-peer applications; thus, in order to simply notice P2P traffic, associate IDS/IPS is designed to be sensitive thereto. SQL server attacks, and worm slammer attacks account for ninety three % of FNs, even though they're aged attacks. This means that these attacks forever have new variations to every IDS/IPS detection.

Keywords: IDS/IPS, FP, FN, RFC, P2P traffic.

1. INTRODUCTION:

During the previous years, the Internet has witness a surge in malicious traffic, such as that generate by denial of service (DDoS) attacks and propagation of worm traffic . Most previous attacks has focused on studying the reasons behind the malicious traffic but not their effects on the normal background traffic. We define normal traffic as network traffic generated due to well-known services and applications, for example, web, ftp, nntp, and smtp. An

increasing number of organizations use information systems to conduct their core business activities. As a result, the frequency and magnitude of intrusion incidents have increased considerably. IDS attacks have many causes, such as malware (e.g., worms, spyware),unauthorized access to systems and misuse of privilege or attempted to gain additional privilege. Traditional intrusion preclusion technique, such as firewall, access manage and encryption, have failed to fully protect networks and systems from increasingly sophisticated attacks and malwares. As a result, IDS have become an essential component of security infrastructure used to detect these threats before they inflict wide spread damage. When building an IDS one needs to consider many issues, such as data collection, data pre-process for intrusion recognition reporting file and response. Among them, IDS recognition is at the heart. Audit data are examine and compare with detection models, which described the pattern of invasive or benevolent behavior, so that both successful and un successful ids attempts can be identified. That is mainly because the presentation of application or the arrangement of the application content is self defined; that is, there is not completed conformance to the specification of RFCs. According, when this application meets an Intrusion detection system/instruction prediction system with the detection rules, its traffic will be regard as malicious traffic, resulting in a lot of False Positives. Second, about 91 percent of False Positives alerts, equal to about 85 percent of FP cases, are not related to safety issue, but to management policy. For example, some company and campuses limit or forbid their employees , students from using peer to peer applications; there fore, in order to easily detect Pear to Pear traffic, an Intrusion detection system/instruction prediction system is configured to be sensitive to it. This causes alerts to be trigger easily regardless of whether the Pear to Pear applications has malicious traffic or not traffic. The last finding shows that buffer spread out, SQL server attacks, and worm slammer attacks account for Ninety Three percent of False Negatives, even though they are aged attacks. The all indicates of attacks always have new variations to evade intrusion detection system/instruction prediction system detection.

I. Intrusion Detection System:

An intrusion detection system is software that automates the process of monitoring the

events occurring in a computer system or network analyzing them for signs of achievable incident, which are violation or imminent threats of infringement of computer protection policies, acceptable use Intrusion Detection Systems vs. Intrusion Prevention Systems policies or standard security practices. The primary purpose of IDS is to help prevent the consequences of undetected intrusions by monitoring network and system activities in real time and identifying and responding to unauthorized activities. The real-time detection requires a watchdog system that sits in the background and monitors all activities, distinguishes various types of incidents and diagnoses actual attacks. Intrusion detection system also allow analysis of current activity in comparison to past activity to identify unusual trends and problems. Most Intrusion detection system take one of the two principal approaches, n/w based approach and host based approach. Both types look for attack signature, exact patterns that usually specify malicious intent or suspicious activity. In addition, intrusion detection system with identification capability was recently introduced. An intrusion detection system dynamically monitors the events taking place in a monitored system, and decides whether these events are symptomatic of an attack or constitute a legitimate use of the system . Figure 1 depicts the organization of an IDS where solid arrows indicate data/control flow while dotted arrows indicate a response to intrusive activities.

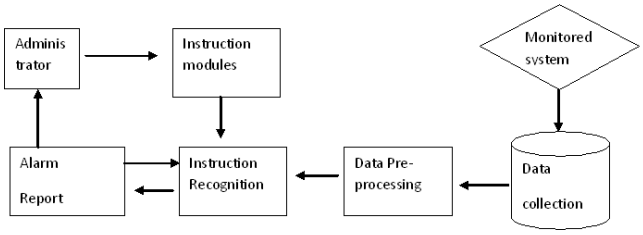


Figure1- Response to Instruction

However, IDS are usually polymorph, and evolve constantly. Misuse detection will fail easily when facing unknown IDS. One way to address this trouble is to regularly updated the knowledge based, either manually which is time consuming and laborious, or automatically with the help of supervise learning algorithm.

II. Implementation:

Classification is achieved by various means prototype. Matching bit patterns of data to those of known protocol is a simple, yet widely used technique. For example to match the BitTorrent protocol hand-shaking phase would be a check to see if a packet began with character 19 which was then

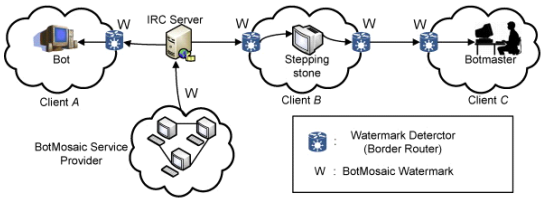
followed by the 19-byte string 'BitTorrent protocols.Further advance traffic classifications technique rely on statistical analysis of attribute such as byte frequencies, packet size and packet inter arrival times. leading classify a traffic flow using a particular protocols, a pre-determined policies can be applied to it and other flows to either guarantee a curtained quality as with VoIP or media streaming services or to provided best effort deliveries. This may be applied at the ingress point (the point at which traffic enters the network) with a granularity that allows traffic management mechanism to divide traffic into individual flows and line, police and shape them in a different way.

III. Typical traffic classes:

Operators often distinguish three broad types of network traffic. Sensitive, Best-Effort, Undesired.

A. Sensitive traffic: Sensitive traffic is traffic the operators have an expectation to distribute on time. Th**is** includes Voice over internet protocol, online-gaming and video-conferencing web-browsing. Traffic management's schemes are typically modified in such a way that the quality of services of these selected uses is certain, or at smallest amount prioritized over other classes of traffic. This can be accomplish by the absence of shaping for traffic class, or by prioritizing responsive traffic above other class.

B. Best-effort traffic: All other kinds of non detrimental traffic. This is traffic that the Internet system protocol deems isn't sensitive to Quality of Service metrics (jitter, packet loss, latency).



A typical Ex would be P2P and email applications. Traffic management schemes are generally tailored so best effort traffic gets what is left after sensitive traffic.

C. Undesired traffic: This category is generally limited to the delivery of spam and traffic create by worms, bonnets, and other malicious attack. In some N/W, this definition can include such traffic as non-local Voice over internet protocol (for example, Windows Skype) or video streaming service to protect the market for the services of the same type.

In these cases, traffic categorization mechanism identify this traffic, allowing the network operator to either block this traffic entirely, or severely hamper its operation.

IV. Security: The degree of security to protect a nation, union of nations, persons or person against damages, losses, dangerous, and crime. Security as a form of security is structure and processes that provide or improve safety as a conditional. The Institute for Security and Open Methodologies (ISECOM) in the OSSTMM Three defines security as a form of protections where a division is create between the assets and the threats. This includes but is not restricted to the eliminations of either the asset or the threats. Security as a national conditional was defined in a United Nations study (1986) so that countries and states can developed and progress securely. Security has to be compared to related concepts: safety, continuity, reliable. The key variation between security and reliability is that protection must take into account the actions of people attempting to cause devastation.

Analyzing the Data set:

A data set is a collection of data, usually accessible in tabular form. Each column represents a particular variables. Each row and columns corresponds to a given members of the data set in questions. It lists values for each of the variables, such as height and weight of an objects or values of random numbers. Each values is known as a datum. That data set may comprised data for one or more members, matching to the number of rows. Each row corresponds to a given members of the data set in questions. It lists values for each of the variables, such as height & weight of an objects or value of random numbers.. The values may be numbers, such as integers ,real numbers , An example representing a person's height in cms, but may also be nominal data (i.e., not consisting of numerical value), for ex representing a person's traditions. More generally, values may be of any of the kinds described as a level of measurement. In each variable, the values will usually all be of the same kind. However, there may also be missing the values which need to be indicated in some way.

V. WinPcap:

The WinPcap is an open source library for packets captured and n/w analysis for the Win32 platforms.Most networking application access the N/W through widely used OS primitives such as socket. It is Easley to access data on the network with this approaches since the OS copes with the lower level details and provides a familiar interface, similar to the one used to read and write files. The purpose of WinPcap is to give this kind of access to Win32 application; it provides facilities to Capture raw packets. Transmit raw packet to the N/w. Gather round statistical information on the N/W traffic.

VI. Jpcap:

Jpcap is a Java class package that allows Java applications to capture and/or send packets to the n/w. Jpcap is based on winpcap and Raw Socket API. Therefore, Jpcap is invented to work on any OS on which libpcap/winpcap has been implemented. at this time, Jpcap is tested on FreeBSD 3.x, Linux RedHat6.1, Solaris, and Microsoft Windows 2000/XP.

VII. Data mining using binary classifier (c4 Algorithm):

Binary classifiers are generated for each class of event using relevant features for the class and classification algorithm .Binary classifiers are derived from the training sample by considering all classes other than the current class as other, e.g., Cnormal will consider two classes: normal and other. The purpose of this phase 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. Moreover, applying the information gain or gain ratio will return all the features that contain more information for separating the current class from all other classes. The output of this ensemble of binary classifiers will be decided using arbitration function based on the confidence level of the output of individual binary classifiers.

A. Multi boosting

The effect of combining different classifiers can be explained with the theory of bias-variance disintegration. Bias refer to an errors due to a learning algorithm while variance refers to an error due to the learned model. The whole expected error of a classifier is the sum of the bias and the difference. In order to reduce bias and difference, some ensemble approaches have

been introduced: Adaptive Boosting Bootstrap Aggregating Wagging and Multi-boosting. This is why the idea emerge of combining both in order to profit from the advantages of both algorithms and achieve an overall error reduction.

.STATISTICAL RESULTS

This subsection analyzes what kinds of False Positives or False Negatives happen easily to IDS/IPS with real world traffic and investigate their frequencies across all False Positives or False Negatives. There are two hierarchies of classification work. One is by protocol, such as HTTP and FTP and NetBIOS and IRC and the other is by Intrusion Detection System policy types like DDoS, buffer overflow, Web attacks, scan, and so on.We can observe that the number of FPs is 13 times that of FNs. In other words, more than 92.85 percent of false cases are FPs. However, when we calculate how many kinds of attack there are in FPs and FNs, we find that the number of kinds of attack in FN cases, 27, is close to that in FPs cases,

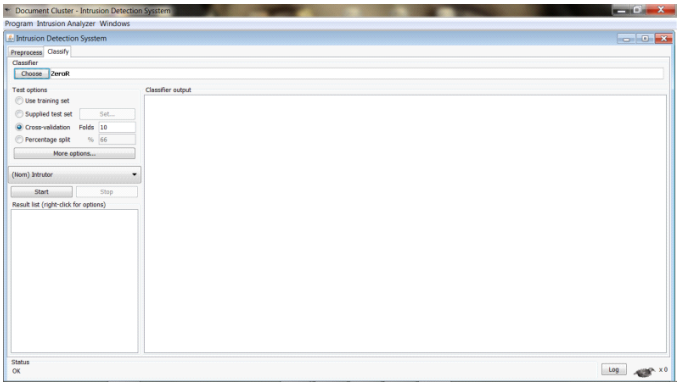


Fig-5: IDS File

we guess that FP cases have many cases with traffic similarity, meaning that network traffic of a certain protocol happens to exhibit some characteristics belonging to other protocols .

Naive Biases Technique:

In simple terms, a naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class changeable. An example, a fruit may be considered to be an apple if it is red, round, and about in diameter. A Naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, apart from of the attendance or absence of the other features.

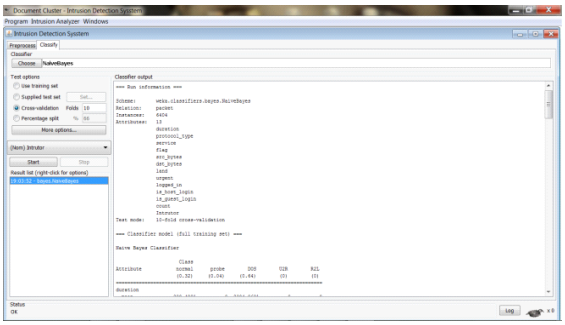


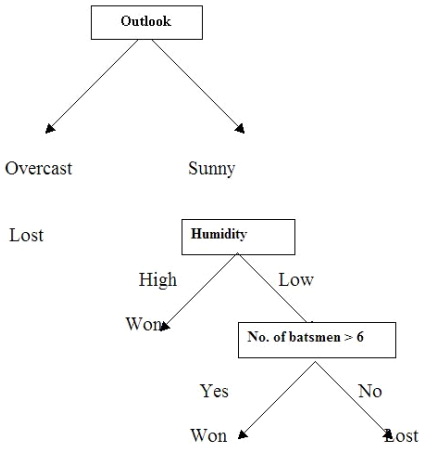
Fig-6: Naive Biases Result

For some types of probability models, Naive Bayes classifiers can be train very resourcefullyin a supervised learning setting. In many practical applications, parameter opinion for Naive Bayes models uses the method of maximum likelihood; in other words, one can work with the Naive Bayes model with no accept Bayesian probability or using any Bayesian methods.

J48 Decision Trees:

A decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. The internal nodes of a decision tree denote the different attributes, the branches between the nodes tell us the possible values that these attributes can have in the observed samples, while the terminal nodes tell us the final value (classification) of the dependent variable.

The J48 Decision tree classifier follows the following simple algorithm. In order to classify a new item, it first needs to create a decision tree based on the attribute values of the available training data. So, whenever it encounters a set of items (training set) it identifies the attribute that discriminates the various instances most clearly. This feature that is able to tell us most about the data instances so that we can classify them the best is said to have the highest information gain. Now, among the possible values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling within its category have the same value for the target variable, then we terminate that branch and assign to it the target value that we have obtained.



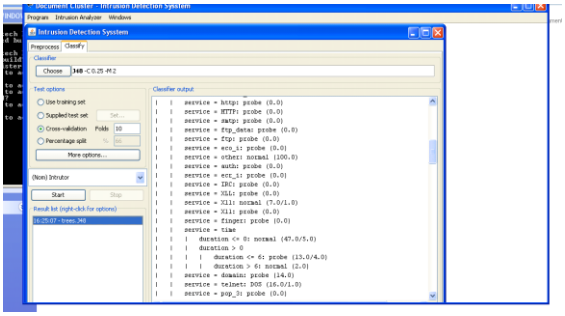


Fig-7: J48 Decision Running

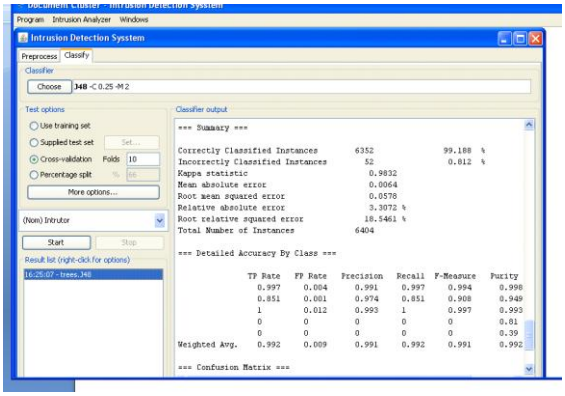


Fig-8: J48 Decision Result

Random Forest:

In method for classification (and regression) that function by constructing a multitude of decision trees at education time and output the class that is the mode of the classes output by individual trees. The word came from random decision forests that was first proposed by Tin Kam Ho of Bell Labs in 1995. The method combines Breiman's "bagging" idea and the random selection of features, introduced separately by Ho and Amit and Geman in order to .assemble a collected works of decision trees with controlled difference. The early increase of random forests was influenced by the works of Amit and Geman which introduced the idea of penetrating over a random subset of the offered decisions when splitting a node, in the background of increasing a single tree.

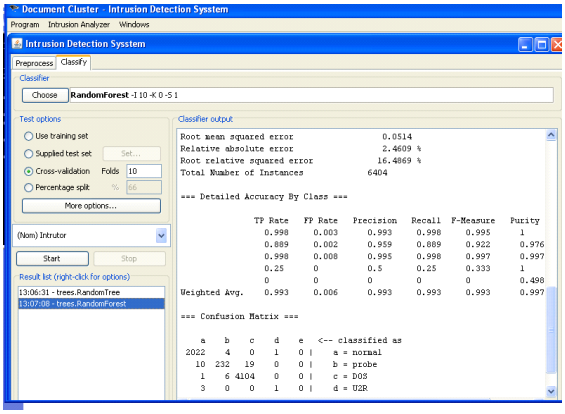


Fig-9: Random Forest Result

In this method a forest of trees is full-grown, and variation amongst the trees is introduce by project that training data into a randomly chosen subspace before fitting each tree. as a final point, the suggestion of randomized node optimization, where the decision at all each nodes are selected by a randomize formula, rather than a deterministic optimization was first introduced by Dietetics.

CONCLUSION:

"It is probably going not realistic to expect that associate IDS be capable of properly classify each event that happens on a given

system. Good detection, like good protection, is just not associate potential goal given the complexness and fast evolution of recent systems.Associate Intrusion detection Scan, however, try to lift the bar for attacks by scale back an oversized categories of attacks and increase the work issue needed to attain a system compromises. The totally eradicate the Federal Protective Service is analogous to confirm whole protection, because it isn't it's not potential to list all vulnerabilities. Then the alternate strategies area unit thought of to handle Federal Protective Service. The simplest strategies to attenuate FP is to fine tune behavior or signatures. Intrusion is associate action that violates protection policy of associate system, ordinarily attributable to a system outcast UN agency enters this method to perform execution however generally used additional ordinarily to perform defiance of policy. The false positives area unit as a result of the intrusion detection systems area unit designed for general conditions and to not take into thought of the precise policy of the organization. Systems Administrator cannot modify intrusion detection system method, will solely bespoke by modification parameters. The alarms have to be compelled to be analyzing to notice the Federal Protective Service establish the connected rules and set the parameters or pass the foundations supported the protection policy. Results area unit promising and any the customization can turn out effectual results. The at the moment part of experiments is to stay the intrusion detection system before the Firewall protocol and to watch the section of the Networks, examine the model by mix with alternative strategies to addition to form effective management of Federal Protective Service.

ACKNOWLEDGMENT

We wish to acknowledge the efforts of **Pantech Solution Pvt ltd., Hyderabad,** for guidance which helped us work hard towards producing this research work.

REFERENCES:

[1] "Effect of Malicious Traffic on The Networks," K.-C. Lan, A. Hussain, and D. Dutta, Proc. (PAM), San Diego, CA, Apr. 2003.

[2] S.-X. Wu and W. Banzhaf, "The Use of Computational Intelligence in Intrusion Detection Systems: A Review," Elsevier Applied Soft Computing, vol. 10, issue 1, Jan. 2010, pp. 1–35.

[3] H. T. Elshoush and I. M. Osman, "Reducing False Positives through Fuzzy Alert Correlation in Collaborative Intelligent Intrusion Detection Systems — A Review," Prof. IEEE Int'l.Conf. Fuzzy Systems, July 2000, pp. 1–8.

[4] M. Sourour, B. Adel, and A. Tarek, "Environmental Awareness Intrusion Detection and Prevention System toward Reducing False Positives and False Negatives," Proc. IEEE Symp. Computational Intelligence in Cyber Security, Apr. 2009.

[5] G. P. Spathoulas and S. K. Katsikas, "Using a Fuzzy Inference System to Reduce False Positives in Intrusion Detection," Proc. 16th Int'l. Conf. Systems, Signals and Image Processing, June 2009.

[6] I.-W. Chen et al., "Extracting Attack Sessions from Real Traffic with Intrusion Prevention Systems," Proc. IEEE ICC, June 2009.

[7] S.-H. Wang, "Extracting, Classifying and Anonym zing Packet Traces with Case Studies on False Positives/Negatives Assessment," M.S. thesis, Dept. Comp. Sci., Nat'l. Chiao Tung Univ., Taiwan, 2010.

[8] Y.-D. Lin et al., "On Campus Beta Site: Architecture Designs, Operational Experience, and Top Product Defects," IEEE Commun. Mag., vol. 48, no. 12, Dec.