IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 1, March, 2013
ISSN: 2320 - 8791
www.ijreat.org

Analysis Of True Positives And True Negatives From Real Traffic Using intrusion

N.Gobinathan [1] ,V.G.Shanmuga priya [2] , M.Dhivya[3]

[1]Assistant Professor
[2&3]UG Scholar
Department of Computer Science and Engineering
VRS College of Engineering and Technology, Arasur-607107, India

Abstract

During the last several years, malicious traffic detection has been an active area of network security because the internet malicious traffic detection has been an active area of network security because the internet has witnessed a surge in malicious traffic generated by network attacks such as denial of service (DOS) probe, url, r2l, viruses, Trojan horses ,spyware and so on. There are multitude of malicious traffic detection techniques are available. In those Intrusion Detection/Prevention system are commonly used today. This work purposes a mechanism of true positive/negative assessment and true positive/negative assessment with multiple IDS/IPS to collect FP,FN,TP,TN cases. FP and FN cases are incorrectly identified and justified the malicious attacks arised in malicious traffic, it sometimes getting failure to detect the attacks properly. By using TP and TN cases, the malicious attacks should be identified and justified correctly. By using this techniques we reduces the problem occurred in malicious traffic. Moreover malicious traffic makes network performance inefficient and gives troubles to user which can be solved by using this technique.

1.0 INTRODUCTION

There are a multitude of malicious traffic detection techniques, and thus, vulnerabilities in common security components, such as firewalls, are unavoidable. Intrusion detection systems (IDSs) and intrusion prevention systems (IPSs) are commonly used today. They are used to detect different types of malicious traffic, network communications, and computer system usage with the mission of preserving systems from widespread damage; that is because other detection and prevention techniques, such as firewalls, access control, skepticism, and encryption have failed to fully protect networks and computer systems from increasingly sophisticated attacks and malware .

Statistically analyze the Fp and FN cases from real world traffic by FP/FN assessment with multiple IDS/IPS.Monitoring and analyzing the communication protocol between a connected device (a user/PC or system) because using protocol based system.An application protocol consists of a system or agent that would typically sit within a group of servers, monitoring and analyzing the communication on application specific protocols.A host-based intrusion detection system (HIDS) consists of an agent on a host which identifies intrusions by analyzing system calls, application logs, file-system modifications(binaries, password files, capability/acl databases) and other host activities and state. An example of a HIDS is OSSEC. The hackers recover the embedding data in original image because the data placed in particular bit position. To attack the hidden data using original image because referred the key value. The data extraction is not separable from the content descriptions.

An FP of the IDS/IPS will not result in an intrusion and it may be caused by two reasons: the detection mechanism of the IDS/IPS may be faulty or the IDS/IPS detects an anomaly that turns out to be benign. Therefore, an FP may cause security analysts to expend unnecessary effort. Moreover, if a hacker launches a *snowblind* attack, the challenge for security analysts is to somehow identify the real attack amidst the chaff caused by the hacker. This may create a potential vulnerability for the IDS. On the other hand, when an IPS has an FP, the primary concern is that legitimate traffic might be blocked. Most organizations consider blocking legitimate traffic as a much more serious problem than generating a false alert. Consequently, an FP of the IPS is a much more serious matter than that of the IDS. If the IPS blocks legitimate traffic a few times, it will be yanked out of the network.

An FN is simply a missed attack, which may put networks or computer systems in danger. Clearly an FN is undesirable, and every vendor strives to provide the most complete coverage possible. However, there is no silver bullet: no product detects all attacks. Hence, the goal becomes providing coverage against high priority attacks. Aside from lack of coverage, several other reasons may also cause an FN. For example, in order to evade the IDS or IPS, the attack may incorporate obfuscation techniques. Another possibility is overwhelming the IDS with traffic beyond its processing capacity, so the IDS will drop the packets needed to detect the attack.

For an IPS, overwhelming it has a different effect: it causes traffic to be dropped. The attack doesn't succeed because attack packets are dropped, but it is also not detected. Accordingly, the attack can be tried again. In practice, for a vendor of IDSs/IPSs, an FN is much more serious than an FP because of negative effects of an FN including reduced trust in the IDS/IPS, and because of damage caused by the intrusion. However, from a user's point of view, an FP may be more serious than an FN because an FP may cause the IPS to block the user's benign traffic. In addition, the user may allow some FNs as long as they're not too requent. Therefore, it is necessary to investigate and analyze FPs and FNs with IDSs/IPSs in detail.

2.0 PROPOSED WORK

IDSs/IPSs can identify a normal activity as amalicious one, causing a true positive (FP), or malicious traffic as normal, causing a true negative (FN) and then a variety of commercial products,open source, and research into IDSs were proposed.To create a pool of traffic traces causing possible FPs and FNs to IDSs because using Attack System Extraction(ASE).When securing a network, administrators have to use many different tools. Although functionality of them is similar, administrators have to spend a considerable amount of time to read documentation and learn how to use a new tool. Network Intrusion Detection Systems gain access to network traffic by connecting to a hub, network switch configured for port mirroring, or network tap.To minimize this effort a specialized tool securing network and checking available service. For each operating system different applications have to be used, regardless they are doing exactly the same. he ASE was expanded into a bigger system, called the PCAPLib system. The PCAPLib system not only extracted and classified the real-world traffic captured from CampusBeta Siteinto proper categories by leveraging multiple IDSs, but also anonymized users privacy in these FP and FN traffic traces out of security considerations.

2.1 PREPROCESSING

In this paper we are going to receive the network packet and extract attributes using the WinPcap and JPCap.

In information technology, a **packet** is a formatted unit of data carried by a packet mode computer network. Computer communications links that do not support packets, such as traditional point-to-point telecommunications links, simply transmit data as a series of bytes, characters, or bits alone. When data is formatted into packets, the bitrate of the communication medium can better be shared among users than if the network were circuit switched. By using packet switched networking it is also harder to guarantee a lowest possible bitrate.

A packet consists of two kinds of data: control information and user data (also known as *payload*). The control information provides data the network needs to deliver the user data, for example: source and destination addresses, error detection codes like checksums, and sequencing information. Typi cally, control information is found in packet headers and trailers, with user data in between.

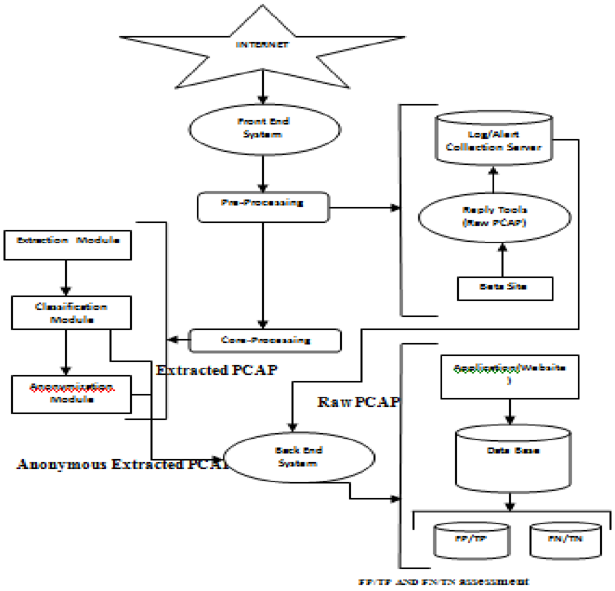

Fig 2.1 Architecture Diagram

Different communications protocols use different conventions for distinguishing between the elements and for formatting the data. In Binary Synchronous Transmission, the packet is formatted in 8-bit bytes, and special characters are used to delimit the different elements. Other protocols, like Ethernet, establish the start of the header and data elements by their location relative to the start of the packet. Some protocols format the information at a bit level instead of a byte level.

A good analogy is to consider a packet to be like a letter: the header is like the envelope, and the data area is whatever the person puts inside the envelope. A difference, however, is that some networks can break a larger packet into smaller packets when necessary (note that these smaller data elements are still formatted as packets).

A network design can achieve two major results by using packets: *error detection* and *multiple host addressing*.

A nominal traffic profile consists of single and joint distributions of various packet attributes that are considered unique for a site. Candidate packet attributes from
IP headers are:
1. packet size,
2. Time-to-Live (TTL) values,
3. protocol-type values, and
4. source IP prefixes.
Those from TCP headers are:

5. TCP flag patterns and
6. server port numbers, i.e., the smaller of the source port number and the destination port number.

Server port number is more stable than sort/destination port numbers because most of the well-known port numbers are small numbers (e.g., below 1,024) and a large portion of Internet traffic uses the well-known port numbers. To increase the number of attributes, we can employ joint distributions of the fraction of packets having various combinations, such as:

7. <packet-size and protocol-type>,
8. <server port number and protocol-type>, and
9. <source IP prefix, TCP flags and packet size>, etc.

Joint distributions often better represent the uniqueness of the traffic distribution for a site, and are harder to guess for the attackers. As many different combinations of single attributes as needed may be used while the storage space permits.

2.1.1WINPCAP

WinPcap is an open source library for packet capture and network analysis for the Win32 platforms. Most networking applications access the network through widely used operating system primitives such as sockets. It is easy to access data on the network with this approach since the operating system copes with the low level details (protocol handling, packet reassembly, etc.) and provides a familiar interface that is similar to the one used to read and write files.

The purpose of WinPcap is to give this kind of access to Win32 applications; it provides facilities to:

- Capture raw packets.
- Transmit raw packets to the network.
- Gather statistical information on the network traffic.

2.1.2 JPCAP

Jpcap is a Java class package that allows Java applications to capture and/or send packets to the network.

Jpcap is based on libpcap/winpcap and Raw Socket API. Therefore, Jpcap is supposed to work on any OS on which libpcap/winpcap has been implemented. Currently, Jpcap has been tested on FreeBSD 3.x, Linux RedHat 6.1, Fedora Core 4, Solaris, and Microsoft Windows 2000/XP.

2.2 ANALYZING THE DATA SET

A **data set** (or **dataset**) is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the data set in question. The data set may comprise data for one or more members, corresponding to the number of rows.The values may be numbers, such as real numbers or integers, for example representing a person's height in centimeters, but may also be nominal data (i.e., not consisting of numerical values), for example representing a person's ethnicity. More generally, values may be of any of the kinds described as a level of measurement.

3.0 MULTI BOOSTING

The effect of combining different classifiers can be explained with the theory of bias-variance decomposition. Bias refers to an error due to a learning algorithm while variance refers to an error due to the learned model. The total expected error of a classifier is the sum of the bias and the variance. In order to reduce bias and variation, some ensemble approaches have been introduced: Adaptive Boosting (AdaBoost),Bootstrap Aggregating(Bagging), Wagging and Multiboosting.This is why the idea emerged of combining both in order to profit from the advantages of both algorithms and obtain an overall error reduction

3.1 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.

4.CONCLUSION

This work proposes the FPNA and TPNA mechanism in the PCAP Lib system to provide statistical analysis of TP and TN cases.By using this techniques we reduces the problem occurred in malicious traffic as well as malicious traffic makes network performance inefficient and gives troubles to user which can be solved by using this technique.By proposing this work the performance of internet access will be provide faster access without any malicious attacks arised in malicious traffic. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts.False positives and false negatives happen to every intrusion detection and intrusion prevention system.This work proposes a mechanism or true positive/negative assessment with multiple IDSs/IPSs to collect TP and TN cases from real-world traffic and statistically analyze these cases

Furthermore, the FPNA as well as TPNA will continue to trace whether statistical results change when the DUTs update their engines and virus patterns.In summary, TPs/TNs are still the key issues for IDSs/IPSs which are less reliable today because of the limitations of the signature-based methodology.

5. REFERENCES

[1] K.-C. Lan, A. Hussain, and D. Dutta, "Effect of Malicious Traffic on The Network," *Proc. Passive and Active Measurement Wksp. (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 Anonymizing 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. 2010, pp. 83–91.
[9] TippingPoint Technologies, "IPS vs. IDS: Similar on the Surface, Polar Opposites Underneath," Whitepaper, http://rovingplanet.net/resources_whitepapers.html.
[10] "Global Market Share Statistics and News," http://marketshare.hitslink.com/os-market-share.aspx?qprid=9.