***A Hybrid Intrusion Detection System:***

***A Review***

Akash Garg Prachi Maheshwari

Department of Computer Science and Engineering, Department of Computer Science and Engineering,

Ajay Kumar Garg Engineering College, Ajay Kumar Garg Engineering College,

Ghaziabad, India Ghaziabad, India

garg.theakash92@gmail.com prashi0110@gmail.com

***Abstract* - The most important purpose of intrusion detection system is to identify attacks against information systems. It is a security method attempting to identify various attacks. In this paper, we reviewed snort as misuse based intrusion detection system as well as ALAD, PHAD, LERAD, NETAD as anomaly based statistical algorithms.**

***Keywords - Intrusion Detection System; Hybrid Intrusion Detection System; Misuse; Anomaly; Snort; Network.***

I. INTRODUCTION

Now a days, it is main problem to maintain the network security. As computer network is growing day by day, Security is the most powerful mechanism for a computer network. Firewalls are not much capable to secure network from attacks because firewall can only detect the attacks which come from outside of the network [1]. With the rapid use of computers and ease of access to internet in the world, the ways to attack and deceive a system has also rapidly increased. Intrusion, in other words, is an illegal process of entering, rights, or taking possession of another's assets.

**What is intrusion detection?**

Intrusion detection is the method of figure out the actions happening in a computer system as well as justifying them for the symbol of intrusion.

II. INTRUSION DETECTION SYSTEM

Intrusion detection is a technique in computer network which play an important role in detecting different type of attacks. It is the procedure of observing the actions which passes in a computer system. Intrusion detection provides three main security phenomenons such as monitoring, detecting, and responding. The motive of Intrusion Detection System is to identify inner as well as outer attacks. In common we can say that, an IDS’s is the composition of intrusion detection hardware & software [2].

Working of Intrusion Detection System is like the security guard. The two assumptions in the field of intrusion detection are 1) user and program events are monitored by computer systems and 2) traditional and intrusion activities can have totally different behaviour [3].

The main benefits can be summarized as:-

1. Detecting attacks and various security violations.
2. Identifying the harm and affected systems.
3. It doesn't suffer with the harmful security.
4. For security model and implementation, works as quality control.
5. Saving problem-behaviors by increasing the obtained risk of discovery.
6. Forensic analysis.
7. Presenting elements of intrusions, granting corrected identification, recovery and corrective measures.
8. Listing the present malware from among and outdoors a system [4].
9. Observing and analyzing computer and/or network system activity.
10. Checking the system configurations and vulnerabilities.
11. Evaluating the integrity of vital system and information files.
12. Estimating irregular activities.

Ideal IDS should possess the following features:-

1. Timeliness: Timeliness should identify intrusions either while they are occurring or shortly afterwards.

2. High probability of detection: It should acknowledge all or most intrusions.

3. Low false-alarm rate: It should contain a few numbers of false alarms.

4. Specificity: In analyzing attacks, specificity should give acceptable characterization information to support a good feedback.

5. Scalability: Scalability should be appropriate to massive or infinite networks.

6. Low a priori information: It should need a least of priori information about potential attackers and their strategies [4].

III. LITERATURE SURVEY

Mahoney described the two models for anomaly detection system for checking doubtful traffic. First of all for passing only the data packets of most requirement, e.g. first some packets of incoming server requests, the traffic were filtered. Second, at the packet byte stage to flag events that have not been found for a long span of time, the most common usable protocols (IP, TCP, telnet, FTP, SMTP, HTTP) were designed [9].

Mahoney and Chan characterize an empirical PHAD (Packet Header Anomaly Detector) that determines the ordinary range of values for 33 fields of the Ethernet, IP, TCP, UDP, and ICMP protocols. On the 1999 DARPA IDEVAL data set, PHAD identify 72 of 201 objects of attacks, together with all but 3 types that accomplishment the protocols tested, at a speed of 10 false alarms per day performing the training on 7 days of attack less internal network traffic. PHAD studied in various ways, and the better results were obtained by exploring packets and fields separately, and by using uncomplicated nonstationary structures [14].

Mahoney and Chan introduced a learning algorithm which structures design of normal nature from anomalies free network traffic. Nature that bifurcates from the known normal design signals possible novel attacks. Their intrusion detection system is special in two aspects. In first, the nonstationary model is presented in which the designing probabilities based on the span of time since the occurrence of last event instead of the rate. Now in the second, the intrusion detection system monitors the protocol collection in order to identify the unknown attacks that try to harm design faults in poorly monitored characteristics of the target software. On the 1999 Defense Advanced Research Projects Agency (DARPA) intrusion detection system evaluation information set, they identified 70 of 180 attacks, and portioned among user behavioural anomalies and protocol anomalies. As their ways are unconventional, there is a symbolic non-overlap of their intrusion detection system with the genuine DARPA participants, which symbolise that they can be taken overall to enhance the coverage [10].

Mahoney and Chan introduced an algorithm called LERAD that operates principles for identifying rare occurrence in normal time series information with long order dependencies. They used LERAD to identifying anomalies in network packets and TCP sessions to identify novel intrusions. LERAD results the actual participants in the 1999 DARPA or Lincoln Laboratory intrusion detection evaluation, and identified almost all attacks that arise a firewall. LERAD is well-organized for three causes. First, only a small part of the traffic has been tested. Second, the principles using only a small sample of the training information has been generated. Third, for building a small collection of principles that mostly covers the information, a coverage test has been used [11].

Aydin et al proposed a hybrid intrusion detection system which is the combination of misuse and anomaly based intrusion detection. In this paper they took snort as misuse based with PHAD (packet header anomaly detection) and NETAD (network traffic anomaly detection) as anomaly based intrusion detection. PHAD and NETAD are the anomaly based statistical method. Firstly, snort is tasted on IDEVAL dataset then it detects 27 attacks out of 201 attacks, secondly PHAD is added to the snort as a preprocessor (Snort + PHAD) is tested on same IDEVAL dataset then the number of attacks detected is increases up to 51 out of 201 attacks, finally NETAD is added to the snort and PHAD as a preprocessor (Snort + PHAD + NETAD) is tested on same IDEVAL dataset then the number of attacks detected is increases up to 146 out of 201 attacks [1].

Nandiammai and Hemalatha proposed a method named as hybrid intrusion detection in which first they used the statistical based anomaly methods such as ALAD (Application Layer Anomaly Detector), LERAD (Learning Rules for Anomaly Detection) and PHAD ( Packet Header Anomaly Detection) then combine these method with snort which is misuse based. Firstly snort is tested on KDD Cup 99 dataset then it detects 77 attacks out of 180 attacks after that PHAD is added to the snort as a preprocessor (Snort + PHAD) is tested on the same KDD Cup 99 dataset then the number of detected attacks increases to 105 out of 180 attacks after that ALAD is added to the snort and PHAD as a preprocessor (Snort + PHAD + ALAD) is tested on the same KDD Cup 99 dataset then the number of attacks detected increases to 124 out of 180 attacks after that LERAD and ALAD is added to the snort as a preprocessor (Snort + LERAD + ALAD) is tested on the same KDD Cup 99 dataset then the number of attacks detected increases up to 149 out of 180 attacks. Secondly, the advantage of both supervised and unsupervised methods has been used to develop a semi-supervised method. Semi supervised approach requires less amount of labeled data with heavy amount of unlabeled data. For semi supervised approach 5000 dataset are taken, in that 2500 taken as training phase and least is taken as testing phase. Training phase includes both the labeled and unlabeled data together. The result of semi supervised approach shows 98.88 % detection rate and 0.5529 % false alarm rate [12].

Nandiammai and Hemalatha proposed an intrusion detection system which is the combination of four approaches such as classification of data named as EDADT (combination of hybrid PSO with C4.5), snort based processing named as hybrid IDS (combination of snort which is misuse based IDS with ALAD and LERAD which are anomaly based statistical algorithm), semi-supervised approach, migrating DDoS attacks named as Varying HOPERAA. Firstly EDADT algorithm gives result as 92.51% sensitivity, 88.39% specificity, 95.37% accuracy, 0.72% false alarm rate. Secondly hybrid IDS gives result as discussed above and Third semi supervised gives result as also discussed above. Finally in HOPERAA algorithm a variable clock drift method is proposed to avoid the client waiting time for server and at the same time message loss is avoided greatly. Thus HOPERAA can minimize the message transfer delay as well as execution time [13].

IV. IDS TYPES

There are two ways of IDS’s. These are misuse-based IDS and anomaly-based IDS. Misuse approach detects the better known attacks that are predefined however fails to identify the unknown attacks. The main advantage of using misuse approach is to produce less false alarms. In anomaly approach, it detects the unknown attacks with high false alarms. Once an anomaly based attack is detected it becomes a signature based or misuse based attack.

**1. Misuse-based or Signature Based IDS**

Misuse based IDS is used to detect the known attacks which are predefined. In the signature based detection, predetermined attack patterns in the form of signatures are presented and these signatures are further used to determine the network attacks [5].

**Working of Misuse-based or Signature Based IDS:-**

Misuse-based or Signature Based IDS works when a person sends data to the network. Firstly all data goes to the server and server check them, if any harmful data is found then server discards the packet else sends it to the network. When the data arrives to server, server uses comparing tool to check that packet from the database of signature stored in the server and if server identify the packet that is matched to the database then it discards the packet else sends the data to the network [5].

**Advantages**

1. Alarm is raised when the signature is matched.
2. Signatures are simply created and understand if we know what network behavior we are trying to identify.
3. Generate low false positive alarm rate.

**Disadvantages**

1. They can only detect the attacks which are previously stored in the database.
2. They cannot detect the live attacks which are created by human.
3. The information about attack is much reliant on the operating system, application, and version.

**1.1 Snort**

Snort is a signature or misuse based intrusion detection system. It is open source IDS developed by Martin Roesch in 1990. It is fast and lightweight application. It can identify the real time traffic psychoanalysis and can identify different kind of attacks. For detecting network packets, it uses binary tcpdump-formatted files or plain text files. Tcpdump is a software program that detects network packets from the computer network and save them in tcpdump-formatted files. Snort is rule-based and it has a language to define new rules. Snort rule can be written in any language and it can be modified if required [5].

Rule Header Rule Options

Fig. 1. Snort IDS rule structure

Figure 1, shows the basic structure of the Snort-IDS rules which are portioned into two parts: first, rule header and second, rule option.

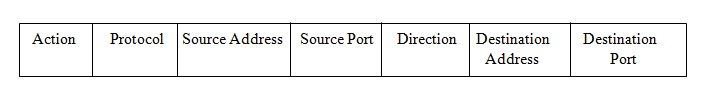


Fig. 2. Snort IDS rules header structure

Figure 2, shows each field in the rule header of the Snort-IDS rules.

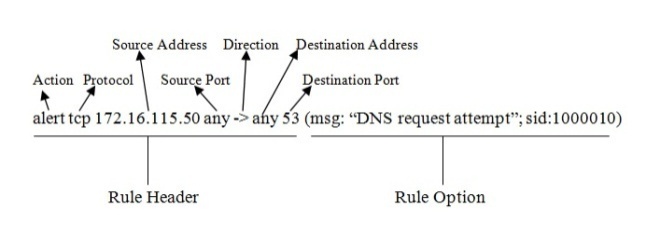


Fig. 3. The snort IDS rules example

Figure 3, shows an example of the Snort-IDS rule. This Snort-IDS rule will generate the alert. When a tcp packet with source IP address number 172.16.115.50 is sent from any port to any destination IP address with destination port number is 53 (DNS). In addition, it also shows the message “DNS request attempt” and the number of the rule is sid:1000010 [6].

**Component of Snort**

Snort is the combination of various components which shows in figure. 4,

1. Packet Decoder
2. Preprocessor
3. Detection Engine
4. Logging and Alerting System
5. Output Modules

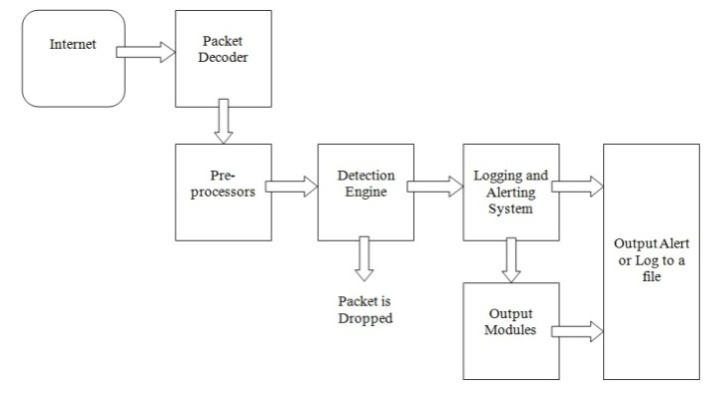


Fig. 4. Component of Snort

**1. Packet Decoder**

This component is used to extract packet from network and then send those packets to the preprocessor [5].

**2. Preprocessor**

If the packet is corrupted, it is used to modify them using some operation & then resends them to the detection engine. It generates alert if any attack is found in the packet [5].

**3. Detection Engine**

This component is mainly used to detect intrusion activity which exists in the packet by using snort rules. If the intrusion is found then it applies appropriate rule otherwise drops the packet. The time taken by detection engine depends upon the number of rules defined in the system [5].

**4. Logging and Alerting System**

This component is used to generate the alarm or log activity of intrusion that is detected by detection engine [5].

**5. Output Modules**

It is used to save the output generated by logging and alerting system. Mainly, it controls different output due to logging and alerting system [5].

**2. Anomaly Based IDS**

Anomaly based IDS is used to detect the unknown attacks. It identifies the abnormal behavior. Anomaly detection characterizes the forecasted performance of the network. Any notable variations from such outline forecasted performance are reported as possible attacks [7]. It is an effort to recognize malevolent traffic based on aberrations from reputable usual network traffic patterns [8].

**Advantages**

1. They can be used to acquire the signature information used by misuse-based IDS.
2. They identify attacks still when complete information of the attack does not be present.
3. No rules needed to be written [9].

**Disadvantages**

1. They generate high false alarms rate.
2. Defining rule set is difficult.
3. They do not define the nature of attack [9].
4. Low detection rate [7].

**2.1 PHAD**

PHAD has 34 attributes comparable to Ethernet (LAN), IP, TCP, UDP and ICMP packet header fields. It creates only a single structure of all network traffic, incoming or outgoing [9]. tn/r is used for calculating the anomaly score. Here n (signifies the number of packets in which the linked characteristic field is included where an irregular value is explored) and r (signifies the number of values which are normally accepted) are calculated at the time of the training period. Here t is the time since the last anomaly [1]. For each packet, PHAD determines anomaly scores and establish no difference between incoming and outgoing traffic.

Considering the network based anomaly detection system, PHAD is different in two ways: First, it structures the protocol apart from behavior of the user as the attacks in majority are exploited in the protocol accomplishment germs which can only be explained by identifying improper input and output. Second, it uses time based structure, supposing an immediate change in a little span of time in the network statistics [1].

**2.2 ALAD**

ALAD structures inward server TCP requests: source and destination addresses and ports, opening and closing TCP flags, and the list of commands in the application payload. It creates split structures for every target host, post number or host/port mixture [9]. In its place of appointing anomaly scores to each packet, it set a score to an incoming server TCP connection. Packets are reassembled from TCP connections [10].

**2.3 LERAD**

LERAD also structures TCP connections, but sample the training data to advise large subset to model [9]. To obtain good rules from the training set, it uses a learning algorithm rather than using a fixed set of rules. It requires two stages: First, for generating rules sampling is required from the training data. Second, for training, validation, testing [11].

The aim of LERAD is to discover provisional rules that recognize unforeseen proceedings in a time series of tuples of unordered attributes (e.g. packet field values, or words in a TCP session). There is no method in LERAD for differentiating true and false alarm [11].

**2.4 NETAD**

In NETAD, there are nine distinct models each for most common protocols such as IP, TCP, HTTP, etc. As any other network anomaly detector, NETAD does not explains the type of an attack, or even indicate either an event is hostile or not. Rather, the unusual or interesting events in a large content of information have been found by NETAD. It takes them to the concern of a network security expert for further analysis [9].

NETAD also structure packets same as PHAD and it functions in two stages: First, here filtering of the incoming client sessions to differentiate beginning of the sessions are allowed. Second, modeling stage. The traffic is eliminated up to 98-99% in the filtering stage. This elimination makes the traffic easier for the modeling stage. In the modeling stage, only those traffic data whose evidence of attacks are included in, are passed through the respective stage [1]. In the second stage, the nine types of packet are modeled.

The formula tna(1 - r/256)/r + tin(ni+ r/W) is used to calculate the anomaly score. Here na represent the number of normal packets from where last training anomaly seen to the end of training period, ti indicate the time when i is last seen and ni represent the number of times i is seen at the time of training period.

REFERENCES

1. M. Ali. Aydin, A. Halim Zaim and K. Gokhan Celyan, “A hybrid intrusion detection system design for computer network security”, Computer and Electrical Engineering 35(2009) 517-526.
2. Qingqing Zhang, Hongbian Yang, kai Li and Qian Zhang, “Research on the intrusion detection technology with hybrid model”, 2nd Conference on environmental science and information application technology, IEEE, 2010.
3. Sumaiya Thaseen and Aswani Kumar, “Intrusion detection model using fusion of PCA and optimized SVM”, IEEE, 2014.
4. Divya and Surendra Lakra, “HSNORT: A Hybrid intrusion detection system using artificial intelligence with snort”, International journal computer technology & application, Vol 4(3), 466-470, 2013.
5. Vinod Kumar and Dr. Om Prakash Sangwan, “Signature based intrusion detection system using SNORT”, International Journal of computer application & information technology, 2012.
6. Natttawat Khamphakdee, Nunnapus Benjamas and Saiyan Saiyod, “Improving intrusion detection system based on snort rules for network probe attack detection”, International conference on information and communication technology, IEEE, 2014.
7. Kapil Wankhade, Sadia Patka and Ravinndra Thool, “An efficient approach for intrusion detection using data mining methods”, IEEE, 2013.
8. Mohammadreza Ektefa, Sara Memar, Fatimah Sidi and Lilly Suriani Affendey, “Intrusion detection using data mining techniques”, IEEE, 2010.
9. Matthew V. Mahoney, “Network traffic anomaly detection based on packet bytes”, ACM, 2003.
10. Matthew V. Mahoney and Philip K. Chan, “Learning nonstationary models of normal network traffic for detecting novel attacks”, ACM, 2002.
11. Matthew V. Mahoney and Philip K. Chan, “Learning Rules for Anomaly Detection of Hostile Network Traffic”, Florida Institute of Technology, Melbourne, FL 32901.
12. G. V. Nadiammai and M. Hemalatha, “Handling intrusion detection system using snort based statistical algorithm and semi-supervised approach”, Research Journal of Applied Sciences, Engineering and Technology 6(16): 2914-2922, 2013.
13. G. V. Nadiammai and M. Hemalatha, “Effective approach toward intrusion detection system using data mining techniques”, Egyptian Informatics Journal (2014) 15, 37–50.
14. Matthew V. Mahoney and Philip K. Chan, “PHAD: Packet Header Anomaly Detection for Identifying Hostile Network Traffic, Florida Institute of Technology”, Melbourne, FL 32901.