RESEARCH ARTICLE OPEN ACCESS

A survey on anomaly and signature based intrusion detection system (IDS)

Mrs.Anshu Gangwar
M tech (CTA) Research scholar
Shri Raam Institute Of Technology,
Jabalpur, M.P.

Mr. Sandeep Sahu,
Assistant Professor & Head P.G. Dept.
of Computer Science & Engineering SRIT,
Jabalpur, M.P.

ABSTRACT
Security is considered as one of the most critical parameter for the acceptance of any networking technology. Information in transit must be protected from unauthorized release and modification, and the connection itself must be established and maintained securely malicious users have taken advantage of this to achieve financial gain or accomplish some corporate or personal agenda. Denial of Service (DoS) and distributed DoS (DDoS) attacks are evolving continuously. These attacks make network resources unavailable for legitimate users which results in massive loss of data, resources and money. Combination of Intrusion detection System and Firewall is used by Business Organizations to detect and p revent Organizations network from these attacks. Signatures to detect them are not available. This paper presents a light-Weight mechanism to detect novel DoS/DDoS (Resource Consumption) attacks and automatic signature generation process to represent them in real time. Experimental results are provided to support the proposed mechanism.

Keywords: Novel DoS attack detection, automatic Signature generation, Main Memory Database Management System

I. INTRODUCTION

Variability is a characteristic that is always present in traffic. It can be ignored or not, it can be smoothed or not, it can be enhanced or not, but it cannot be removed. Some of the variability depends on intrinsic causes, meaning that it exists whenever there is a communication between two or more parties. Such variability is mainly due to the communication protocols being used and the approaches used to generate and transfer data traffic: live or streaming audio and video, distance education, entertainment, peer-to-peer, telephony and video-conferencing, as well as numerous new and often still evolving communication protocols. Extrinsic variability results from the interaction between data flows. Connections are continuously being established over the Internet, ranging in capacity from hundreds of Mb/s to several Gb/s, and continuously sharing resources between them. However, such environments do not have a central authority regulating and checking communication quality, nor any feedback structure to throttle unfriendly practices or products. This significantly hardens data traffic control, and together with intrinsic variability makes network traffic characterization a challenging task. DoS/DDoS attack is attempt by attacker to prevent Internet site or Server from functioning efficiently or properly. There are several ways of launching DoS/DDoS attacks against a server. Every attack uses any one of the following technique:

I. Consume Server resources
II. Consume network bandwidth
III. Crash the server using vulnerability present the server
IV. Spoofing packets

Even though there are different ways to launch attack but every attack makes server either nonresponsive or extremely slow. Firewall and Intrusion Prevention System (IPS) can prevent Server from known DoS/DDoS attacks and sometimes from their variations; as their working mechanism is known in advance. But no one can build a prevention system which will prevent Server from every novel DoS/DDoS attack. One possible solution is to detect a novel attack in real time and automatically generate a signature to represent it. Once the signature of attack is available; defense mechanism against that attack can be developed. Network traffic information in order to increase the speed of novel attack signature generation module. Section 2 gives overview the related work and section 3 gives overview of KDD 99 dataset used for training and testing IDS. Section 4 describes advantages of attribute selection process in design of IDS and Section 5 describes requirement of automating the Signature Generation process. Section 6 describes proposed mechanism and section 7 presents the Experimental results. Section 8 compares the proposed mechanism with existing solutions and section 9 concludes the paper.

II. TYPES OF IDS

Host Intrusion Detection System HIDS is a software product, resides on a specific machine called host, and does its job by protecting the entire system and discloses if a system has been compromised. It monitors the file system integrity, system register state system logs of the host machine to find the evidence of suspicious activity if any. If any user attempts to access authorized content on the host in a shared network, HIDS identifies and collects the relevant data in a quickest possible manner. HIDS only look for the intrusions on the single host but not on the entire network system

Network Intrusion-detection System software process on a dedicated hardware system. The NIDS places the network interface card on the system into promiscuous mode, meaning that the card passes all traffic on the network to the NIDS software. The traffic is then analyzed according to a set of rules and attack signatures to determine if it is traffic of interest. If it is, an event is generated. The most common configuration for an NIDS is to use two network interface cards.

Signature-based methods - The signature-based methods monitor and compare network packets or connections with predetermined patterns known as signatures. This technique is a simple and efficient processing of the audit data. Although the false positive rate of these techniques can also be low, comparing packets or connections with large sets of signatures is a time consuming task and has limited predictive capabilities. The signature-based methods cannot detect novel anomalies that would not be defined in the signatures, and thus administrators frequently have to update the system signatures.

Anomaly based methods an anomaly detection [10, 11] system first creates a baseline profile of the normal system, network, or program activity. Thereafter, any activity that deviates from the baseline is treated as a possible intrusion. Anomaly detection systems offer several benefits. First, they have the capability to detect insider attacks. For instance, if a user or someone using a stolen account starts performing actions that are outside the normal user-profile, an anomaly detection system generates an alarm. Second, because the system is based on customized profiles, it is very difficult for an attacker to know with certainty what activity it can carry out without setting off an alarm. Third, an anomaly detection system has the ability to detect previously unknown attacks.

III. RELATED WORK

A.Swati Paliwal[1]this paper methodology based on Service and Remote to User attacks is proposed. The proposed approach aims at gaining maximum detection of the probing, R2L and DoS attacks with minimum false positive rate.

B.Yu-Xin Ding et al [2] proposed a Snort-Based Hybrid (Misuse-Anomaly) IDS. It is divided into three modules: misuse detection, anomaly detection and signature generation module. Snort is used as misuse detection module to detect known attacks. Anomaly detection module used uses Frequent Episode Rule mining algorithm with a sliding window to generate rules for Anomaly detection. Signatures of newly detected attacks by Anomaly detection module are generated by using Signature generation module. It uses A priori algorithm for signature generation. IT provides good performance in offline detection, but cannot be used for real- time detection

C.Gang Xiong and Minxia Zhang [3] proposed a clustering based outlier detection method to detect unknown (novel)intrusive activities. They considered intrusive activities as outliers and used DOExMi Cluster (proposed by them) to detect outliers of unknown type. The micro-cluster concept, data normalization and k-mean measure are used interactively to create sub micro-clusters of normal activities till two micro-clusters can be merged to create new micro-cluster. After this network activity instances which cannot fall into any micro-clusters are recognized as outliers.

D.Adetunmbi A.Olusola et al [4] performed experiments using KDD 99 dataset. Their experimental results show that, two network afficfeaturesnum_outbound_cmds and is_hot_login have no relevance in Intrusion Detection. Their resultsalso show that, derived featur esnamely num_compromised, su_ attempted, num_ file_ creations, is_guest_login, and dst_ host_ rerror_ rate are of little significance for Intrusion Detection process. So if these features are not used, it will increase speed of IDS and reduce the resource requirements without affecting the accuracy.

E.Jie Yang et al [5] proposed Hybrid

(Anomaly-Misuse) Intrusion Detection System (IDS) using network protocol analysis. It is consist of four modules: Data Preprocessing, Misuse detection, Anomaly Detection and Decision making module. Both Misuse and Anomaly detection modules are built using Decision tree. Decision making module classifies any network activity as intrusion if both (Misuse and Anomaly) detection modules classify it as intrusion.

F.Miner and Fuzzy Inference Engine.[6] Data Analyzer module analyzes network packets and performs packet grouping to get aggregate information. This aggregated information is then used by fuzzy data miner to generate the rules for Intrusion detection. These rules and network traffic data is given as an input to the fuzzy inference engine which determines whether there is any intrusive activity or not. It cannot be used for real time detection due to its high computational complexity and high false positive rate.

G.Imen Brahmi et al [7] proposed Hybrid (Misuse-Anomaly) IDS using data mining and mobile agent technology to detect known and novel attacks. It uses mobile agents to collect and analyze network traffic. Multiple copies of sniffing agents are created and distributed in the network to collect network traffic data in a file. Data present in this file is processed by filter agent to transform it into a form suitable for Misuse and Anomaly detection. Distribute clustering technique is used by H.Wei Wang et al [8] performed experiments using KDD 99 dataset and their experimental results show that, network traffic record containing only 10 relevant features with highest information gain can be used for Intrusion detection with same or improved detection rate.

I.Neveen I. Ghali [9] performed experiments using KDD 99 data set and experimental results show that, only 7 features are enough to detect DoS attack with high accuracy. This reduction in number of attributes for detection process reduces

i. Amount of data to be processed by 83%
ii. Mean square error in detection of novel attack by approximately 90%
iii. Memory and CPU time required to detect attacks

J. Güneş Kayacık et al [10] performed experiments using KDD 99 data set and used information gain to express Feature relevance with different attacks. Their experimental results show that, normal network traffic, neptune and smurf attacks are highly related to certain network traffic features compared to others. If only these are used for Intrusion detection, attack detection task becomes much easier and provides good results.

IV. KDD 99 DATA SET

Data set (KDD'99 and DARPA) AND TCPDUMP for capturing real network traffic The KDD'99 DARPA 1998 dataset [34], which is the most popular data set used to evaluate IDSs. The DARPA 1998 environment simulating a typical U.S Air Force LAN that contains 100's of users on 1,000's of hosts. Training data and two and made it available for the KDD'99 Classifier-Learning Contest [13]. Through the processing, the binary tcp dump data is transformed to connections that contain some context information for each network session Despite some drawbacks [45], the KDD'99 dataset is still the most widely used benchmark data set for evaluating machine learning-based IDSs. It can be used for testing machine learning algorithms without further time-consuming preprocessing.

V. AUTOMATIC SIGNATURE GENERATION (ASG)

Every activity (legitimate and intrusive) over network has a unique pattern. These patterns can be used to detect which activities are going on the network. So these patterns are also called as Signatures. Signatures of known intrusive activities are defined and used to detect their existence. But there are two major problems in this approach; both problems are related with the manual process usually carried out to create Signature of attack.

First, a detailed and precise knowledge about attacks process is required to define its Signature. If defined Signature is too simplified, it will generate high false positive rate. On other hand if it is too specific, it will result in high false negative rate. Second, some time is required to gain detailed and precise knowledge about attack. This introduces delay between the first time attack is reported and generation of signatures to detect it. Thus zero-day or novel attacks are serious threat for computer systems. [K]M. Soleimani et al [13] proposed some approaches to make this process easier by correlating and thus reducing the number of alerts to analyze, but major problems are still unsolved.

According to I. Qualys [14] approximately twenty to forty new vulnerabilities in commonly used networking and computer products are discovered and published every month by users or attackers. Such wide-spread availability of known vulnerabilities leads to launch of novel („Zero-day) attacks. Since Firewall and IDS cannot protect network against novel attack it leads to massive loss of data, resources and money. Thus, both the activities are very crucial

i. Detect novel attacks in real time without human involvement

ii. Generate signatures of novel attack in real time without human involvement Once the signature of attack is known, security expert

Anomaly Detection based Filter (ADF) and Signature Generator (SG) are used to generate signatures which can represent Novel attack. Known Attack Signature DB (KAS DB) contains signatures of known attacks and used by Signature based IDPS to detect them. LogDB contains all the connection records which do not match with known attacks. These records are used to generate signatures of novel attack after filtering them by using ADF.

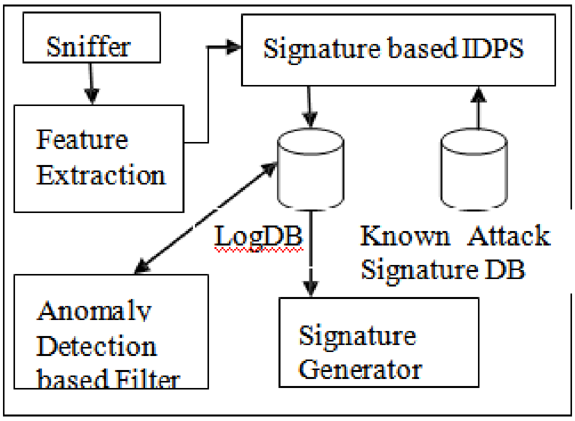

Fig 1: Proposed Mechanism

Author of [7] proposed a security solution based on KNN method. For better evaluation of unknown attacks and authors method our proposed method uses the same concept incremented one step further. In proposed work detection of suspicious traffic using clustering well be tested integrating the SVM filter on them.

Support Vector Machine (SVM) Support Vector Machine applications to classification and clustering of channel current data. SVMs are variation calculus based methods that are constrained to have structural risk minimization (SRM), i.e., they provide noise tolerant solutions for pattern recognition. solutions for pattern recognition. The SVM approach encapsulates a significant amount of model-fitting

can

VI. PROPOSED MECHANISM

Figure 1 shows the proposed mechanism. Signature-based IDPS is used to detect and prevent server from known DoS/DDoS attacks.

information in the choice of its kernel. In work thus far, novel, information-theoretic, kernels have been successfully employed for notably better performance over standard kernels. Currently there are two approaches for implementing multiclass SVMs.

K-Nearest Neighbor (KNN) Algorithm The K-Nearest Neighbor algorithm (KNN) is a method for classifying objects based upon the closest training samples in the feature space [7]. Assume n labeled samples are mapped in a feature space – an abstract space where each sample is represented as a point. Each dimension of the feature space represents one attribute (feature) of the sample. The space is partitioned into regions by the classes (labels) of the points. An unknown point is classified to the class whose labels are most frequent among the K nearest samples. An example of KNN classification. The unknown point (circle) belongs either to the first class (square) or to the second class (triangle

VII. COMPARISON WITH EXISTING APPROACHES

Clustering based IDS methods are based on two assumptions. The first assumption is; number of normal instances in dataset is much higher than number of intrusive instances and second is; enough difference exists between the intrusive and the normal instances. But these assumptions are not true in every situation. Proposed method does not rely on any assumption. It performs incremental analysis while remaining approaches does not. This difference makes it real time solution for detection of Novel attacks. Every other mechanism treat every network traffic record which does not match with Signature and Anomaly based IDS as possible attack which increases the computational complexity; this is not done by proposed mechanism. This reduces the processing time and increases the accuracy of Generated Signature. Comparison of proposed mechanism with existing solutions is shown in table 1.

Table 1. Comparison of proposed mechanism with existing solutions

Method Name	**Is Incremental Analysis**	**Assumptions Used**	**Data Processed for new attack detection**	**Relative Complexity**	**Real time detection**
Outliers Detection Based on clustering	No	Yes	Every data instance which does not match with Signature and Anomaly based IDS	Moderate	No
DecisionTree based Hybrid IDS	No	No	Every data instance Which does not match with Signature and Anomaly based IDS	Moderate	No
Fuzzy Logic based Hybrid IDS	No	No	Every data instance which does not match with Signature and Anomaly based IDS	High	No
MAD-IDS	No	Yes	Every data instance which does not match with Signature and Anomaly based IDS	High	Yes
Anomaly Detection by Clustering in the Network	No	Yes	Every data instance which does not match with Signature and Anomaly based IDS data	High	No
K-Means Clustering and NaïveBayes Classification	No	Yes	Every datainstance which does not match with Signature and Anomaly based IDS	High	No
Snort-Based Hybrid IDS	No	No	Every data instance which does not match with Signature and Anomaly based IDS	High	No
Proposed Mechanism	Yes	No	Much lesser compared to other mechanisms	Low	Yes

VIII. CONCLUSION AND FUTURE WORK

Defined as the ability of a program a system to learn and improve their performance we develop a hybrid intrusion detection model. Proposed model describe hybrid architecture is better instead of single approach. The model consists of a set of base feature selecting classifiers and each uses partial original feature space as well as a data mining classifier. The basic concept is using ensemble feature selection technique to increase the detection rate and data mining technique to reduce the number of false alarms. This paper presents a Light weight mechanism for novel DoS/DDoS attack detection and signature generation to represent those using MMDBMS. Condition based network connection records omission used for Novel attack Signature Generation increases the speed and accurate

REFERENCES

[1] Swati Paliwal,,Ravindra Gupta” Denial-of-Service,Probing& Remote to User(R2L) Attack Detection using Genetic Algorithm Volume60” No.19. December 2012

[2] Jie Yang, Xin Chen, Xudong Xiang, Jianxiong Wan, “HIDS-DT: An Effective Hybrid Intrusion Detection System Based on Decision Tree”, 2010 International Conference on Communications and Mobile Computing

[3] Yu-Xin Ding, Min Xiao, Ai-Wu Liu “Research And Implementation On Snort-Based Hybrid Intrusion Detection System”, Proceedingsof the Eighth International Conference on Machine Learning and Cybernetics, Baoding,12-15 July IEEE 2009, DOI: 1.1109/ICMLC.2009. 5212282

[4] Adetunmbi A. Olusola, Adeola S. Oladele, Daramola O.Abosede, “Analysis of KDD 99 Intrusion Detection Dataset for Selection of Relevance Features”, Proceedings of the World Congress on Engineering and Computer Science 2010 Volume I, IEEE 2010 AnomalousContents”,2010Sixth Advanced International Conference on Telecommunications

[5] Yu-Xin Ding, Min Xiao, Ai-Wu Liu, “Research And Implementation On Snort-Based Hybrid Intrusion Detectio System”, Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding,12-15 July IEEE2009, DOI: 10.1109/ICMLC.2009. 5212282

[6] Adetunmbi A. Olusola, Adeola S. Oladele, Daramola O.Abosede, “Analysis of KDD 99 Intrusion Detection Dataset for Selection of Relevance Features”, Proceedings of the World Congress on Engineering and Computer Science 2010 Volume I, IEEE 2010

[7] Imen Brahmi, Sadok Ben Yahia, Pascal Poncelet, “MAD-IDS: Novel Intrusion Detection System Using Mobile Agents and Data Mining Approaches”, Lecture Notes in Computer Science,2010, Volume 6122/2010, 73-76, DOI: 10.1007/978-3-642-13601-6_9

[8] Wei Wang, Sylvain Gombault, Thomas Guyet, “Towards fast detecting intrusions: using key attributes of network traffic”, The Third International Conference on Internet Monitoring and Protection, 978-0-7695-3189-2/08, 2008 IEEE, pp. 86 – 91

[9] Neveen I. Ghali, “Feature Selection for Effective Anomaly-Based Intrusion Detection”, International z Journal of Computer Science and Network Security, VOL.9 No.3, March 2009, pp. 285-289 H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm

[10]. Heywood, “Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets”, Proceedings of the Third Annual Conference on Privacy, Security and Trust, October 2005, St. Andrews, Canada

[11] Soleimani, E. Khosrowshahi, M. Doroud, M. Damanafshan, A. Behzadi, M. Abbaspour, “RAAS: A Reliable Analyzer and Archiver for Snort Intrusion Detection System,” ACM SAC, 2007

[12] Feng Guo, Yingzhen Yang , Lian duan , “Anomaly Detection by Clustering in the Network”, International Conference on Computational Intelligence and Software Engineering, 2009, ISBN: 978-1-4244-4507-3

[13] Adetunmbi A. Olusola, Adeola S. Oladele, Daramola O.Abosede, “Analysis of KDD 99 Intrusion Detection Dataset for Selection of Relevance Features”, Proceedings of the World Congress on Engineering and Computer Science 2010 Volume I, IEEE 2010

[14] Vijay Katkar, Rejo Mathew, “One Pass Incremental Association Rule Detection Algorithm For Network Intrusion Detection System”, International Journal of Engineering Science and Technology (IJEST), ISSN :0975-5462 Vol. 3 No. 4 Apr 2011

[15] Feng Guo, Yingzhen Yang , Lian duan , Anomaly Detection by Clustering in the Network”, International Conference on Computational Intelligence and Software