EFFECTIVE INTRUSION DETECTION SYSTEMS USING HYBRID APPROACH

Mr. Prakash N Kalavadekar  
Research Scholar K.K Wagh Institute of Engineering Education & Research, Nashik Kprak3004@gmail.com  
Dr. Shirish S. Sane  
Research Guide K.K Wagh Institute of Engineering Education & Research, Nashik sssane@kkwagh.edu.in

Abstract: Conventional intrusion prevention methods such as firewalls, access management schemes or cryptography techniques, have not proved themselves to completely defend networks and systems from refined malwares and attacks. The Intrusion Detection Systems (IDS) proved to be the right salvage to the current issues and became an important element of any security infrastructure to sight these threats before they induce widespread harm. The basic aim of IDS is to detect attacks and their nature that may harm the computer system. Several different approaches for intrusion detection have been reported in the literature. These approaches are broadly categorized into three approaches: i) Signature-based approach ii) anomaly based approach and iii) hybrid approach that combines signature and anomaly detection approaches. Hybrid approach has been found to be superior that either signature based or anomaly based approaches. However, despite superior performance, the hybrid approach has till date failed to provide desired detection rate and time needed for detection.

Keywords: Intrusion Detection, Security, Signature, Anomaly, Data Mining

I. Introduction

Security attacks are classified into 2 main branches: passive and active. The passive attackers are usually invisible (hidden) and either faucet the communication link to gather data; or destroy the functioning parts of the network. Passive attacks are classified into eavesdropping, node wrong, and node tampering / destruction and traffic analysis sorts. In active attacks, somebody really affects the operations within the attacked network. This result could also be the target of the attack and might be detected [1].

As an example, the networking services could also be degraded or terminated as a result of these attacks. Active attacks is classified as hole attacks (blackhole, wormhole, sinkhole, etc.). Denial-of-
techniques (such as encoding, authentication, access management, secure routing, etc.) area unit bestowed because the initial phase of defense against intrusions. However, there in any quite security system, intrusions can't be entirely prevented. The intrusion and compromising of a node ends up in counseling like security keys being discovered to the intruders. This may ends up in the failure of the designed preventive security mechanism. Therefore, IDSs area unit designed to reveal intrusions, before they'll disclose the secured system resources. IDSs area unit permutually thought of as a defense second wall from the protection purpose of read. IDSs area unit Internet equivalent of the stealer alarm that area unit being employed in physical security systems these days. Because the expected operational demand of IDSs is given as: “low false positive rate, calculated because the proportion of normalcy variations detected as anomalies, and high true positive rate, calculated because the proportion of anomalies detected”. Thus there's plenty a lot of scope for analysis in up detection performance for unknown attacks & detection speed.

II. Motivation and Related Work

Misuse or signatures based detection: - The signatures (profiles) of the previously known attacks are generated and are used as a reference to detect future attacks. The advantage of this type of detection is that it can accurately and efficiently detect known attacks; hence they have a low false positive rate [1]-[5]. The disadvantage is that if the attack is a new kind (that was not profiled before), then the misuse detection would not be able to catch it.

Sobh pointed out that these systems are very much like the anti-virus systems, which can detect most or all known attack patterns, but are of little use for the attack methods that are unknown yet [18]. These systems used known attack dataset like KDD Cup 99 which contains 41 attributes for each signature of different types (DOS, R2L, U2R, and Probe) attacks [5]. Shun and Malki gives neural network-based IDS for detective of internet-based mostly attacks on a network. Neural networks are used to determine and predict current and possibly future attacks. In this feed forward type neural network with the back propagation training algorithmic was used to detect intrusion. For training & testing of classifier KDD Cup (1999) dataset is used. This method is only used for signature detection [4][13].

Sahana Devi K. J., Bharathi gives information of systems based on misuse model like SNORT and Bro [1]. Siva s Sivatha Sindhu, S.Geetha , A. Kannan given decision tree based light weight signature based detection (nerotree) using a wrapper approach. As well it used genetic algorithm for optimizing selection of signature features from given 41 features in KDD Cup 99 dataset [5] [13].

Anomaly based detection: - This is based on statistical behavior modeling. Normal operations of the members are profiled and a certain amount of deviation from the normal behavior is flagged as an anomaly [1] [7]. The disadvantage of this detection type is that the normal profiles must be updated periodically, since the network behavior may change rapidly. The advantage of this detection type is that it is well suited to detect unknown or previously not encountered attacks.

According to Garcia Teodoro et al. anomaly based IDSs are further divided into three categories according to the nature of the processing involved in the behavioral model considered[16] [18].

1] Statistical based: In statistical based anomaly IDSs, a profile representing its stochastic behavior is generated. After that, the network is monitored and profiles are generated periodically and an anomaly score is generated by comparing it to the reference profile. If the score passes a certain threshold, the IDS will flag an occurrence of the anomaly.

2] Knowledge based: Knowledge based anomaly IDSs rely on the availability of the prior knowledge (data) of the network parameters in normal operating condition as well as the one under certain attacks.

3] Machine learning based: In machine learning based anomaly IDSs, an explicit or implicit model of the analyzed patterns is generated. These models are updated periodically, in order to improve the intrusion detection performance on the basis of the previous results.

Hybrid Approach: - This approach combines signature and anomaly based detection approaches so that advantages of both approaches will improve the performance of the system. This approach works for detection of known & unknown attacks [1] [2] [4].

Koutsouitos, Christou, and Efremidis gives solution using neural network type classifier design system using a combination of more than one neural network which is capable of detecting network attacks on web servers. The system can detect unseen attacks and make categorization.
Prema Rajeswari and Kannan discusses a rule based approach using enhanced C4.5 algorithm for intrusion detection in order to detect abnormal behaviors of internal attackers through classification and decision making in networks[9].

D. Barbara gives sensitivity of both signature-based and anomaly-based IDSs with respect to the attack characteristics, system training history, services provided, and underlying network conditions. Data mining techniques are also useful to construct classification models from labeled attacks [5] [8].

Lee et al. gives information about a framework to specify rules for anomaly detection against normal track problems [18].

Fan et al. extended Lee et al.’s work to discover accurate boundaries between known attacks and unknown anomalies [18].

Kai Hwang, Fellow, IEEE, Min Cai, Member, IEEE, Ying Chen, Student Member, IEEE, and Min Qin suggest data mining techniques where association rules were used to build IDS. They have found the differences between single connection and multi connection attacks. They also give information of systems based on misuse model like SNORT and Bro [1].

Gisung Kim, Seungmin Lee, Sehun Kim (2014) done the analysis on a brand new hybrid intrusion detection technique that hierarchically integrates a misuse detection model and an anomaly detection model. First, the C4.5 decision tree (DT) is used to produce the misuse detection model which decomposes the traditional training information into smaller subsets. Then, the one-class support vector machine (1-class SVM) was used to produce associate anomaly detection model [2].

The experiments were conducted with the NSL-KDD data set, which is a modified version of well-known KDD Cup 99 data set. They demonstrate that their method is better than the conventional methods in terms of the detection rate for both unknown and known attacks while it maintains a low false positive rate. And significantly reduces the high time complexity of the training and testing processes.

The advantage of one-class SVM is it does not require the labeled information. However, there is downside to using one-class SVM: it is difficult to use the one-class SVM in the real world, due to its high false positive rate.

Wenyiing Fenga,b,Qinglei Zhangc, Gongzhu Hud, Jimmy Xiangji Huangc (2014) combines the SVM method with CSOACNs( Clustering based on Self-Organized Ant Colony Network) to take the advantages of both while avoiding their weaknesses. The algorithm is implemented and evaluated using a standard benchmark KDD99 data set. Experiments show that CSVAC (Combining Support Vectors with Ant Colony) outperforms SVM alone or CSOACN alone in terms of both classification rate and run-time efficiency [19]. The CSVAC algorithm outperforms pure SVM in their experiments with higher average detection rate, less training time, and lower rates of both false negative and false positive.

It is better than pure CSOACN in terms of less training time with comparable detection rate and false alarm rates.

The effectiveness and the flexibility of IDS system is less & required to improve. As per the survey it can be observed that there is scope for research by combining advantages of both signature & anomaly detection model to improve the detection rate & time efficiency with real time traffic & multi connection in network.

**Challenges in Intrusion Detection:**

- Low detection efficiency due to the high false positive rate usually obtained.
- Low throughput and high cost, mainly due to the high data rates (Gbps).
- Detection with single connection & multi connection.
- Only signature detection is inefficient because it works only for known attack.
- Most of the IDS systems perform poorly in defending themselves from attacks.

**III. Implementation Methodology**

1. **Signature based IDS or Misused IDS**

Signature based IDS can be trained by using previously known attack pattern. Whenever new record comes to system it compares that pattern with previously known attack pattern and based on comparison decision will be given.

2. **Anomaly (Behavior) based IDS**

The proposed work to mix the benefits of minimum or less false positive detection rate of signature-based intrusion detection system (SIDS) and anomaly-based intrusion detection system (AIDS).Anomaly-based detection systems are generally used for the detection of unknown attacks. Signature based mostly system leverages manually characterized attack signatures to observe notable attacks in period traffic. These two subsystems be
part of hands to hide all traffic events initiated by each legitimate and malicious user. Figure 1 shows proposed architecture of Effective IDS, in which combination of signature based & anomaly detection system will be used for detection of known & unknown attacks. In this, focus will be use of efficient algorithm for signature machine engine in order to decompose the normal data evenly into each subset without degrading the misuse detection performance, which will be expected to significantly improve the overall performance of the system [2]. In signature machine engine NSL-KDD dataset which is a changed version of renowned KDD Cup 99 [13] dataset will be used for experimental analysis. In anomaly detection, data mining engine will be used for generating Frequent Episode Rules (FER’s) [2][3][16] to form normal profile & abnormal profile of the behavior of the system. The forming rules is a major research work so that performance of detection & speed will be increased. The mining engines will produce output as anomaly with the help of normality score. These detected anomalies will be used for further in generation of new signatures using data mining algorithms to update the existing database. The generation of new signature will improve the performance of detection of various types of unknown attacks. So lot much work is expected to design & develop an effective Intrusion Detection system using combination of signature & anomaly detection system.

IV. Experimental Work

1) Collection of dataset KDD Cup -99 & its study.

The KDD'99 dataset was used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD'99 dataset, the fifth International Conference on Knowledge Discovery and Data Mining. The competition task was to build a network intrusion detector. This database was acquired from the 1998 DARPA intrusion detection evaluation program. An environment was set up to acquire raw TCP/IP dump data for a local-area network (LAN) simulating a typical U.S. Air Force LAN, which was operated as if it was a true environment, but blasted with multiple attacks. There are totally 4,898,431 connections recorded, of which 3,925,650 are attacks. For each TCP/IP connection, 41 various quantitative and qualitative features were extracted. The simulated attacks fall in one of the following four categories:

Denial of Service Attack (DOS): In this category the attacker makes some computing or memory resources too busy or too full to handle legitimate request, or deny legitimate users access to machine.

Users to Root Attack (U2R): In this category the attacker starts out with access to a normal user account on the system and is able to exploit some vulnerability to gain root access to the system.

Remote to Local Attack: In this category the attacker sends packets to machine over a network but who does not have an account on that machine and exploits some vulnerability to gain local access as a user of that machine.

Probing Attack: In this category the attacker attempts to gather information about network of computers for the apparent purpose of circumventing its security.

<table>
<thead>
<tr>
<th>Attack Category</th>
<th>List of Attacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOS</td>
<td>'neptune', 'back', 'smurf', 'pod', 'land', 'teardrop'</td>
</tr>
<tr>
<td>U2R</td>
<td>'buffer_overflow', 'loadmodule', 'rootkit', 'perl'</td>
</tr>
<tr>
<td>R2L</td>
<td>'warezclient', 'multihop', 'ftp_write', 'spy'</td>
</tr>
<tr>
<td>PROBE</td>
<td>'portsweep', 'satan', 'nmap', 'ipsweep'</td>
</tr>
</tbody>
</table>

Table 1: Types of Attack Signature in KDD
The KDD’99 dataset has 41 attributes for each record. Some of these attributes are irrelevant and redundant. Irrelevant attributes simply add noise to the dataset and affect the accuracy of proposed models using it. Another important point that has to be considered is the computation cost. Using a dataset with a large number of attributes results in a lengthy training and detection processes, and hence degrading the performance of an intrusion detection system.

So it is important to find relevant attributes with the help of some data mining tools like Weka or Oracle. Data mining that ranks the attributes in a dataset based on their significance using the Minimum Description Length (MDL) algorithm. It is found that 13 attributes out of the 41 attributes of the KDD’99 dataset have an importance value above zero, and the rest have an importance of zero.

2] Collection of dataset NSL-KDD a new version of KDD Cup -99

KDDCUP’99 is the mostly widely used data set for the anomaly detection. But researchers conducted a statistical Analysis on this data set and found two important issues which highly affect the performance of evaluated systems, and results in a very poor evaluation of anomaly detection approaches. To solve these issues, they have proposed a new data set, NSL-KDD, which consists of selected records of the complete KDD data set. The following are the advantages of NSL-KDD over the original KDD data set:

□ It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records.
□ The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set. As a result, the classification rates of distinct machine learning methods vary in a wider range, which makes it more efficient to have an accurate evaluation of different learning techniques.
□ The numbers of records in the train and test sets are reasonable, which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion. Consequently, evaluation results of different research works will be consistent and comparable.

The following Table shows analysis for 156 packets tested using signature matching algorithm which contains different types attack packets & normal packets.

Table 4: Analysis of Signature matching

<table>
<thead>
<tr>
<th>Number of packets</th>
<th>Number of packets containing attacks</th>
<th>Number of packets detected attacks</th>
<th>Detection rate in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>156</td>
<td>114</td>
<td>106</td>
<td>92.98</td>
</tr>
</tbody>
</table>

Number of instances in NSL -KDD is as per below table.
4] Implementation & Testing of Anomaly Based Detection

Frequent Episode Rules Database is generated as per normal profile using the following expression:
L1.L2,… Ln -> R1,...,Rm E2->E1, E3

Events (E1, E2, E3) are nothing but services like TCP, UDP, Authentication

Rule: for normal sequence E2->E1, E3
(service = authentication) → (services = smtp)(service = smtp)

Table 5: Some Sample Rules

<table>
<thead>
<tr>
<th>Rule no.</th>
<th>Rules for Normal Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Service=TCP-&gt;Service=SMTP-</td>
</tr>
<tr>
<td>2</td>
<td>Service=TCP-IP-&gt;Service=SMTP-</td>
</tr>
<tr>
<td>3</td>
<td>Service=TCP-IP-&gt;Service=SMTP- &gt;Service=SMTP-&gt;Service=SMTP-</td>
</tr>
<tr>
<td>4</td>
<td>Service=TCP-IP-&gt;Service=SMTP- &gt;Service=SMTP-&gt;Service=SMTP- &gt;Service=TCP-IP-&gt;Service=TCP-IP-</td>
</tr>
<tr>
<td>5</td>
<td>Service=TCP-IP-&gt;Service=SMTP- &gt;Service=SMTP-&gt;Service=SMTP- Service=TCP-IP-Service=TCP-IP-</td>
</tr>
</tbody>
</table>

Table 6: Anomaly Detection

<table>
<thead>
<tr>
<th>Rules generated for Normal Profile</th>
<th>Number of packet sequence send</th>
<th>Number of Abnormal sequence send</th>
<th>Number of Anomaly Detected</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>50</td>
<td>16</td>
<td>16</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 7: Analysis of EIDS

<table>
<thead>
<tr>
<th>For implemented System</th>
<th>IDS</th>
<th>ADS</th>
<th>HIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection Rate</td>
<td>68%</td>
<td>100%</td>
<td>84.17%</td>
</tr>
<tr>
<td>False Alarm Rate</td>
<td>6.2%</td>
<td>4%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

V. Conclusion & Future Scope

Since the current IDS technologies are not sufficient enough to provide a reliable detection rate so work should be carried on to improve the rate. Another major problem in this research area is the speed of detection. Most of the research works are aimed to introduce the most time efficient methodologies. The goal is to make the implemented methods suitable for the real time implementation.

The information presented constitutes an important starting point for addressing R&D in the field of IDS. Faster and more effective countermeasures are needed to cope with the ever-growing number of detected attacks.

The EIDS advantages will come from using dynamic data mining threshold and automated signature generation. Generating more signatures by ADS will further enhance the overall performance of the effective intrusion detection system.

VI. References


AUTHOR'S BIBLIOGRAPHY

Mr. Prakash Kalavadekar received ME degree in Branch of Computer Science & Engineering from Walchand College of Engineering, Sangli in 2007 and pursuing PhD in Computer Engineering from Savitribai Phule Pune University, Pune. Presently working as Associate Professor in Department of Computer Engineering in SRES College of Engineering, Kopargaon (MH).