ISSN 2320 - 2602
Volume 4 No.10, October 2015
International Journal of Advances in Computer Science and Technology
Available Online at http://www.warse.org/IJACST/static/pdf/file/ijacst014102015.pdf

A Study on Intrusion Detection Using Data Mining Techniques

Lalit N. Dhande [1], Prof. Yogesh S. Patil[2]
[1]ME CSE Student at SSGBCOET, Bhusawal, lalitdhande11@gmail.com
[2]ME CSE Professor at SSGBCOET, Bhusawal, yogeshe146@gmail.com

ABSTRACT

Recently, more and more attention is paid to cyber security; intrusion detection. I our current society, the threat of cyber intrusion is increasingly high and harmful. With the increase in use of computer, criminal activity has also shifted from physical intrusion into cyber intrusion. Since maintainsthe excellent property that the original flow of transmission can be losslesslycarried out throughout the communication. Current method for these systems include using anomaly detection or signature database. In this paper we use both anomaly detection and signature database using data mining techniques. The proposed method can provides a tool that would run data mining tools against a log file to detect patterns that may be considered an unauthorized activity. The tool gains addition patterns as time goes by and grows more effective.

Keywords–Cyber security; Intrusion Detection; Data Mining& Data Mining Techniques

1. INTRODUCTION

With the evolution of the Internet over the last few years, the need for security has been rising with it mainly due to the openness and connectivity nature of the Web. People and organizations are faced with more challenges every day to secure their data and allother assets of value to them. Prevention, detection, and response are part of the Network Defense model. Intrusion detection systems are important components of defensive measures protecting computer networks from abuse. In this project a data mining tool has been developed that describes the behavioral forensics in intrusion detection.

1.1 Intrusion Detection

An intrusion detection system (IDS) in Figure1 attempts to detect an intruder breaking into the system or a legitimate user misusing system resources. Network based intrusion detection system and host based intrusion detection system are two primary intrusion detection models. A network intrusion detection system (NIDS) monitors traffic on the network wire and attempts to discover if a hacker/cracker is attempting to break into a system or cause a denial of service attack. A host based intrusion detection system audits data from a single host to detect intrusion.

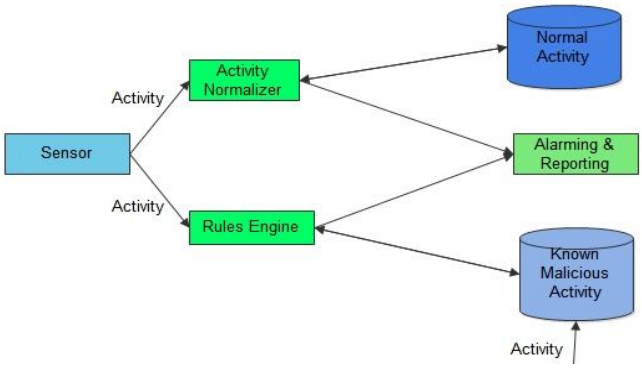

Figure 1: Components of IDS

The network provides the data to the rules engine and the activity normalizer. The rules engine searches the data for patterns from the known malicious activity database. The activity normalizer performs analysis of the data. The sensor is a network packet capturer.

2. LITERATURE SURVEY

2.1Intrusion Detection Techniques

Intrusion detection techniques can be categorized into misuse detection and anomaly detection, based on the strategy for detection [1]. Misuse detection finds intrusions Sensor Activity Normalizer Rules Engine Alarming & Reporting Known Malicious Activity Normal Activity by looking for activity corresponding to known techniques for intrusions [7]. This involves the monitoring of network traffic in search of direct matches to known patterns of attack called signatures. This is a rule based approach. The disadvantage of this approach is that it can only detect intrusions that follow pre-defined pattern [8]. In anomaly detection, the system defines the expected behavior of the network or profile in advance. Any significant deviations from this expected behavior are then reported as possible attacks [4].

Intrusion detection systems based on misuse detection lack the ability to detect attacks that do not fit a pre-defined signature. Those shortcomings can be overcome by using anomaly detection. It has potential ability to recognize unforeseen attacks. A critical issue for anomaly detection is the need to

reduce false alarms. Data mining is a major approach to anomaly detection. This helps in identifying the anomalous pattern from the audit data [10].

2.2Data Mining

Data mining attempts to extract implicit, previously unknown and potentially useful information from data. Data mining of intrusion detection involves processing large quantities of collected data to look for patterns of interest.

3.DATA MINING TOOL FOR INTRUSION DETECTION

As previously stated, data mining has been defined as the nontrivial extraction of implicit, previously unknown and potentially useful information from data. The goal of mining is to derive multi-feature correlations from audit data.

3.1 Data Collection

Data collection is an important part of this tool. Without the audit data this tool is of no significance. Snort is used to collect the audit data.

3.2 Snort

Snort is an intrusion detection system tool. Snort is a cross-platform, lightweight network intrusion detection tool that can be deployed to monitor small TCP/IP networks and detect a wide variety of suspicious network traffic as well as outright attacks.

Snort is appropriate for a network in a school or a small business.

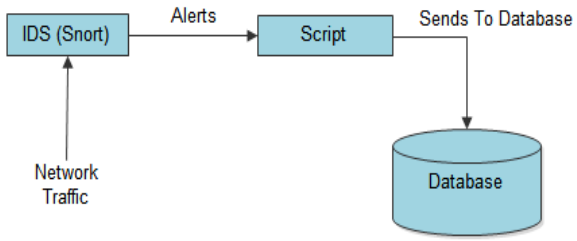

Figure 2: Collection of Data

4. SIGNATURE BASED DETECTION V/S ANOMALY BASED DETECTION

4.1Signature based detection

This detection technique uses specifically known patterns to detect malicious code. These specific patterns are called signatures. Identifying the worms in the network is an example of signature based detection.

4.2 Anomaly Detection

These techniques are designed to detect abnormal behavior in the system. The normal usage pattern is base lined and alerts are generated when usage deviates from the normal behavior.

Example, if a user logs on and off 20 times a day while the normal behavior is 1-2 time.

5. METHODOLOGY

5.1 Project Goal

The research presented in this is part of our on-going work which is to investigate the role of data mining algorithms in an Intrusion Detection System. Most importantly we would like to find out if discovering patterns within a log file will provide patterns of intrusion attacks. In this we test the ability of Data Mining using a clustering technique to discover DoS Attacks. Clustering refers to the grouping of similar data. This grouping allows users to see patterns of reoccurring activities or popular trends. The description of reoccurring activities is highly compatible with the description of a denial of service attack. We also investigate the use of Data Mining to incorporate both the signature and anomaly database scheme.

5.2 PreProcessing Log Files

The first step in designing and implementing our data mining tool for intrusion detection, was to analyze and parse the network log files. In order to do so, the following steps were implemented. 1) Extract the date and time, 2) Extract data until first colon, 3) Extract data within parenthesis, 4) Decode remaining information. The program adds a new column when a part of the line is split based on special characters, thus the header has the column information. The extracted data is added to the body of the file.

5.3 Data Mining Tool for Detecting Attacks

We categorize any action that may bring down a system or retrieve information as an attack whether it is honest or malicious. In the current version of our tool we have implemented a clustering algorithm which matches connections that appear multiple times. This enables us to detect possible password guessing or DoS attacks.

6. FUTURE SCOPE

The future implementation with this techniques. This tool can be upgraded withadditional features. Some of them are listed here:

1. Include services while clustering the data. A Source host is vulnerable to attacks because of the services running on it. So it

will be a good idea to cluster the databased on the services, a host is running.

2. Another data mining technique: Classification, can be used to classify the intrusioninto one of the known intrusion categories.

7. CONCLUSION

A data mining technique for intrusion detection that is robust and has sufficiently goodperformance is developed. This techniques helps the network administrator to identifyanomalies in the network traffic, create rules to detect intrusion based on the analysis andto distinguish false alarms from positive alarms.This techniques provides a better insight into differentiating false intrusions from the realones. It also provides with more rules/patterns to perform better intrusion detection. Theobservations and rules mined from audit data with the help of this techniques are merged andadded into an aggregate rule set to detect intrusion. This makes the existing intrusiondetection system more robust.

8. REFERENCES

1. Jonathon Ng, Deepti Joshi, Shankar M. Banik "**Applying Data Mining Techniques to IDS** "*IEEE 2015 "2015 12th International Conference on Information Technology*

2. Anthony Raj. A "**Study on Exploring Data Mining Techniques for Network Intrusion Detection**" *International Journal of Innovative Research in Information Security (IJIRIS) ISSN: 2349-7017(O) Issue 1, Volume 2 (January 2015)*

3. Sandip Sonawane, ShailendraPardeshi and Ganesh Prasad "**A survey on intrusion detection techniques**" **International Journal 21 April 2012**

4.Yingbing Yu, Han Wu"**Anomaly Intrusion Detection Based Upon Data Mining Techniques and Fuzzy Logic**"2012 *IEEE International Conference on Systems, Man, and Cybernetics October 14-17, 2012, COEX, Seoul, Korea*

5. L. Vokorokos, A. Balaz "**Host-based Intrusion Detection System***" 2010 IEEE INES 2010, 14th International Conference on Intelligent Engineering Systems*

6.Sushil Kumar Chaturvedi, Prof. VineetRichariya, Prof. Nirupama Tiwari "**Anomaly Detection in Network using Data mining Techniques**" *International Journal of Emerging Technology and Advanced Engineering April 2012.*

7. Sathish S.N Project Manager, "**Using Data Mining Techniques for Intrusion Detection**" *Infosys Limited, Mysore, India International Journal of Innovative Research in Science,Engineering and Technology*

8.QingqingZhang"**Research on the Intrusion Detection Technology with Hybrid Model**" *2010 2nd Conference on Environmental Science and Information Application Technology.*

9. SuratSrinoy and WerasakKurutachSuanDusit"**Anomaly Detection Model Based on Bio-Inspired Algorithm and Independent Component Analysis**" *Rajabhat University 295 Ratchasima Road, Dusit, IEEE 2006*

10. LI Yunl, LIU Xue-cheng, and ZHU Feng "**Application of DataMining in Intrusion Detection**" 2010 International Conference on Computer Application and System Modeling (ICCASM 2010)

11. Zhang Qu, Huang Wen-jie"**Research on Data Mining Technologies Appling Intrusion Detection**"f*by 2010 IEEE*