Volume 7, Issue 5, May 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Intrusion Detection System Using Data Mining Techniques– A Review

**Ranju Marwaha**
Assistant Professor/CSE, SSIET Derabassi,
Punjab, India

***Abstract: Security is one of the most challenging areas for computers and networks. Intrusion Detection System tools aim to detect computer attacks, computer misuse and to alert the proper individuals upon detection. But still they face challenges in robust and changing environment. Data Mining based IDS can provide more accuracy of results and these results can be used in automated Decision Support System or by Analyst. This will focus on role of Data Mining Techniques in IDS.***

***Keywords: Data Mining, Intrusion Detection System, Data Mining Techniques***

## I. INTRODUCTION

Security means degree of protection given to the network or system. The main goals of security are confidentiality, integrity and availability of data. Attacks on network can be referred as Intrusion. Intrusion means any set of malicious activities that attempt to compromise the security goals of the information. In early days, only conventional approaches were used for network such as encryption, firewalls, virtual private network etc but they were not enough to secure network completely. It is difficult to depend completely on static defense techniques. This increases the need for dynamic technique that can be monitor system and identify illegal activities. Thus to enhance the network security dynamic approach is introduced and known as Intrusion Detection System. Intrusion detection system collects online information from the network , monitors and analyzes this information and partitions it into normal & malicious activities, provide the result to system administrator.

## II. WHY WE NEED IDS?

Of the security incidents that occur on a network, the vast majority come from inside the network.These attacks may consist of otherwise authorized users who are disgruntled employees. The remainder come from the outside, in the form of denial of service attacks or attempts to penetrate a network infrastructure .IDS tools allow for complete supervision of networks, regardless of the action being taken, such that information will always exist to determine the nature of the security incident and its sources.

The main function of IDS includes:

- Monitoring and analyzing the information gathered from both user and system activities.
- Analyzing configurations of system and evaluating the file integrity and system integrity.
- For static records, it finds out the abnormal pattern.
- To recognize abnormal pattern, it use static records and alert to system administrator

## CLASSIFICATION OF IDS

*Signature based*

In signature based detection mechanism the attack patterns are saved in the database. Each packet of the network traffic is compared with the attack patterns to detect abnormal behavior. Signature based intrusion detection system detects only know attacks. If attack signatures are clearly defined then it has low false positive.

Requires specific knowledge of intrusion behavior and collect data before the intrusion could be out of date. Difficult to detect unknown attacks. Raises alerts regardless of the outcome. Example if a windows worm tries to attack a Linux system then the IDS sends many alerts of unsuccessful attack.

The knowledge of the attacks is dependent on the specific environment.

*Anomaly based intrusion detection system*

Anomaly based intrusion detection system is based on the network behavior. The network behavior is defined by the administrator or is learned by the dataset during the training phase of the development of IDS. Rules are defined for normal behavior and abnormal behavior. Example, Snort and Bro-IDS are anomaly based intrusion detection system. It has the ability to detect unknown attacks.Defining the rule set for intrusion detection is difficult.Efficiency of system depends on the fitness of the rules and its testing on the testing datasets.

 

## III. INTRUSION DETECTION APPROACHES

IDS vendors implement their products in different ways and there are consequently several ways to categorize intrusion detection systems. The first is based on the scope of the IDS's monitoring; that is, whether it is installed on and uses data from a single host computer, or is a network-based product that monitors traffic on the network as a whole, as well as analyzes data from individual computers.

Another difference in implementation has to do with how the vendor markets the system, either as a software product or as an integrated hardware device (appliance).

*Host-based intrusion detection*

A host-based IDS is one in which the software is installed on a single system and the data from that system is used to detect intrusions. Because the host-based IDS protect the server "at the source," it can more intensely protect that specific computer. The host-based system usually examines log files on the computer to search for attack signatures. Important system files and executables may also be checked periodically for unexpected changes. A host based system will also monitor ports and trigger an alert if certain ports are accessed.

*Network-based intrusion detection*

A network-based IDS monitors data from network traffic as well as data from one or more host computers to detect intrusions. A network-based IDS analyzes data packets sent over the network, and generally uses a "promiscuous" network adapter (one that is capable of reading all of the packets sent over the network, rather than just those packets addressed to it). The network-based IDS examines packet headers, which are generally not seen by the host-based IDS. This allows the detection of Denial of Service (DoS) and other types of attacks that may not be detected by a host-based IDS.

## IV. DATA MINING BASED INTRUSION DETECTION SYSTEM

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Network traffic is huge and information comes from different sources, so the dataset for IDS becomes large. Hence the analysis of data is very hard in case of large dataset. Data mining techniques are applied on IDS because it can extract the hidden in formation and deals with large dataset. Presently Data mining techniques plays a vital role in IDS. By using Data mining techniques, IDS helps to detect abnormal and normal patterns.

The various data mining techniques that are used in the context of intrusion detection.

1. *Correlation analysis*: *Correlation* is often used as a preliminary technique to discover relationships between variables. More precisely, the correlation is a measure of the linear relationship between two variables.
2. *Feature selection* : A subset of features available from the data is selected for the application of a learning algorithm.
3. *Machine learning* : Machine learning explores the study and construction of algorithms that can learn from and make predictions on data
4. *Sequential patterns* :It is used to excavate connection between data, time series analysis gains more focus on the relationship of data in times.
5. *Classification*: It is a technique of taking each instance of a dataset and assigning it to a particular class. Typical classification techniques are: inductive rule generation, genetic algorithms, fuzzy logic, neural networks and immunological based techniques.
6. *Clustering* : It is a technique for statistical data analysis. It is the classification of similar objects into a series of meaningful subset according to certain rules, so that the data in each subset share some common trait.
7. *Deviation analysis*: Deviation analysis can reveal surprising facts hidden inside data
8. *Forecast:* Finding certain laws according to historical data, establishing models and predicting types, characteristics of the future data, etc based on the model.

With the increase in computerization and storage of more and more sensitive data on the data servers, the security of the data servers is a major issue.

As the intrusion detection systems are being used for monitoring networked devices where they look for the behavior patterns of various anomalous and malicious behaviors in the audit data. Making comprehensive IDS requires more time and expertise.

On the other hand Data mining based IDS require less expert knowledge and give better performance (Barbara et al.,2001; Noel et al.,2002; Eskin et al ., 2002 ; Markou and Singh, 2003).They can generalize new and unknown attacks in a better way. The methods used for finding knowledge can be mathematical or non - mathematical; it can be deductive or inductive. The available knowledge can be used for optimizing enquiry, manage information, control progress and make intellectual decision.

Given databases of sufficient size and quality, data mining technology can generate

 

1. Automated prediction of trends and behaviors :

Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data. It also provides various models that help in forecasting.

2. Automated discovery of previously unknown patterns:

Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions

## V. CONCLUSION

Therefore, data mining helps people in application of data from low and simple inquiry to discovery knowledge in data and support decision. In the analysis of intrusion detection system, the data circulating in network has the following characteristics:

1. Mass data : the number of data message sent and received are quite impressive.
2. Noisy: when network is unstable, data information may get changed in the transportation of message ( Liu, 2009 ).

It can be seen that these data is in accordance with the feature of data mining, naturally data mining need to be applied to intrusion detection system. The various detection models need log data as training collection whose accuracy will largely influence intrusion detection system. Because of the density and accuracy of visit network, it is difficult to acquire completely no attack action. In addition, it is also uneasy to log attack behavior. Data mining technology can resolve this problem, in the analysis of general network visit; isolated point is an invasive behavior to reduce the difficulty of acquisition of training data. Data mining technology, for example Clustering, Classification, Feature Summary, association rules can be applied in the intrusion detection system. It has been proved that data mining technology improves the property of intrusion detection system, the processing rate and reduces the rate of misreporting.

## REFERENCES

[1] Deepthy K Denatious & Anita John, "Survey on Data Mining Techniques to Enhance IntrusionDetection", International Conference on Computer Communication and Informatics (ICCCI 2012), Jan. 10,2012, Coimbatore, INDIA

[2] Rung - Ching Chen , Kai - Fan Cheng and Chia - Fen Hsieh," Using Rough Set And Support Vector Machine For Network Intrusion Detection ",International Journal of Network Security & Its Applications (IJNSA), Vol 1, No 1, April 2009

[3] David Ndumiyana, Richard Gotora and Hilton Chikwiriro, "Data Mining Techniques in Intrusion Detection: Tightening Network Security" , International Journal of Engineering Research & Technology (IJERT) , Vol. 2 Issue 5, May – 2013

[4] Singhal, "ANSWER: Network monitoring using object oriented rules, in Proceedings of the Tenth Conference on Innovative Application of Artificial Intelligence, G. Weiss and J. Ros, Madison, Wisconsin, July 1998

[5] Barbara et al.,2001; Noel et al.,2002; Eskin et al ., 2002 ; Markou and Singh, 2003

 