THE ROLE AND USE OF DATA MINING TECHNIQUES FOR INTRUSION DETECTION SYSTEMS

Amit Sharma*

ABSTRACT

Information access through Internet provides intruders various ways of attacking a computer system. More and more organizations have become vulnerable due to intrusion of cyber attackers which compromise the security of their network [1]. Establishment of a safe and strong network system for the secure information transmission has already become focus of research. One of the many ways used for this purpose is the use of firewalls. A firewall might prevent from many kinds of attacks such as using the protocol weakness, the source route, the address counterfeited, and so on, and provide safe data channel. But it could do nothing about the back door in application layer, the attack or stealing caused by authority exceeding of internal user and the information damaging. Moreover, because the firewall was at the network boundary, its own design flaws are inevitably exposed to attackers; firewall only is hard to resist the variety of attacks [2]. Thus in orders to ensure network security, data mining techniques are adopted for detecting abnormal or unauthorized behavior in the Intrusion Detection System (IDS.) These data mining techniques are an offline environment to add more depth to the network defense in order to determine the various attacks or threats to the network.

*Assistant Professor, Apeejay institute of Management, Jalandhar.

1. THE INTRUSION DETECTION SYSTEM

An intrusion can be defined as "any set of actions that attempt to compromise the integrity, confidentiality, or availability of a resource" [3]. An **intrusion detection system** (**IDS**) is a device or software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a Management Station [4].

Functions of an intrusion detection system are to:

- Monitor and analyze the user and system activities.
- Analyze system configurations and vulnerabilities.
- Assess system and file.

A secured network must have the following three features:

- **Confidentiality:** Only authorized people should be able to access the data that are being transferred through the network.
- **Integrity:** The integrity of the data should be maintained starting from its transmission until it is received by the receiver.
- **Availability:** The network should be resilient to any kind of attacks.

IDS comprises of several components such as: sensors to generate security events, a console to monitor the events and alerts and control the sensors, and a central engine to record events logged by the sensors in a database and uses a system of rules for the generation of rules to generate alerts from security events received. Components of IDS are shown in Figure 1.

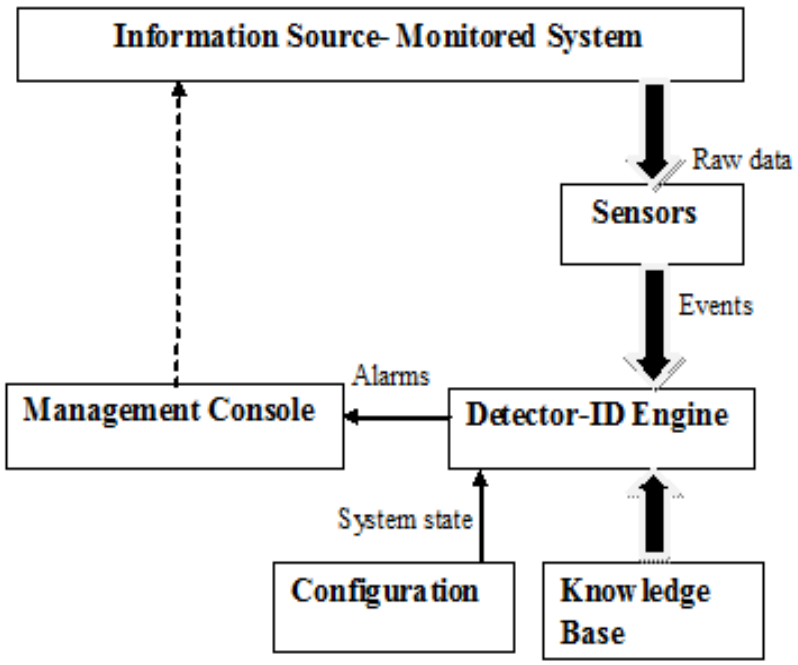

Figure 1: IDS Model

Intrusion detection can broadly be categorized as:

1. **Misuse Detection vs Anomaly Detection:** In **misuse detection**, the IDS identifies illegal invasions and compares it to large database of attack signatures. Basically, the IDS looks for an already documented specific attack. The main disadvantage of this method is that if an unknown intrusion appears then it cannot be detected. The detection efficiency of this method is quite high. In **anomaly detection,** the IDS monitors the network segments and compare their state to the normal baseline to detect anomalies.
2. **Network-based vs Host-based Systems:** A **network-based** intrusion detection system (NIDS) identifies intrusions by examining network traffic and monitoring multiple hosts. A **host-based** intrusion detection system examines the activity of each individual computer or host.

2. THE DATA MINING TECHNOLOGY

Data Mining

Data mining is the process of sorting through large database or data warehouse and extracting knowledge interested by the people. The extracted knowledge may be represented as concept, rule, law and model. The purpose of data mining is to help the decision- maker in order to find potential association between data, found neglected elements which might be very useful for trends and decision- making behavior. It has been described as “the nontrivial extraction of implicit, previously unknown, and potentially useful information from data” [5] and “the science of extracting useful information from large data sets or databases” [6].

Data mining identifies trends within data that go beyond simple analysis. Through the use of sophisticated algorithms, non-statistician users have the opportunity to identify key attributes of any kind of real life problems like Intrusion Detection Activities, Face recognition problem, Image processing, business processes and any other target opportunities. However abdicating control on these process from the statistician to the machine may or may not result in positives or useful results [1] until one can assure that the data on which the operations are supposed to be performed are complete in all respect. Figure 2 shows the basic approach of Data Mining.

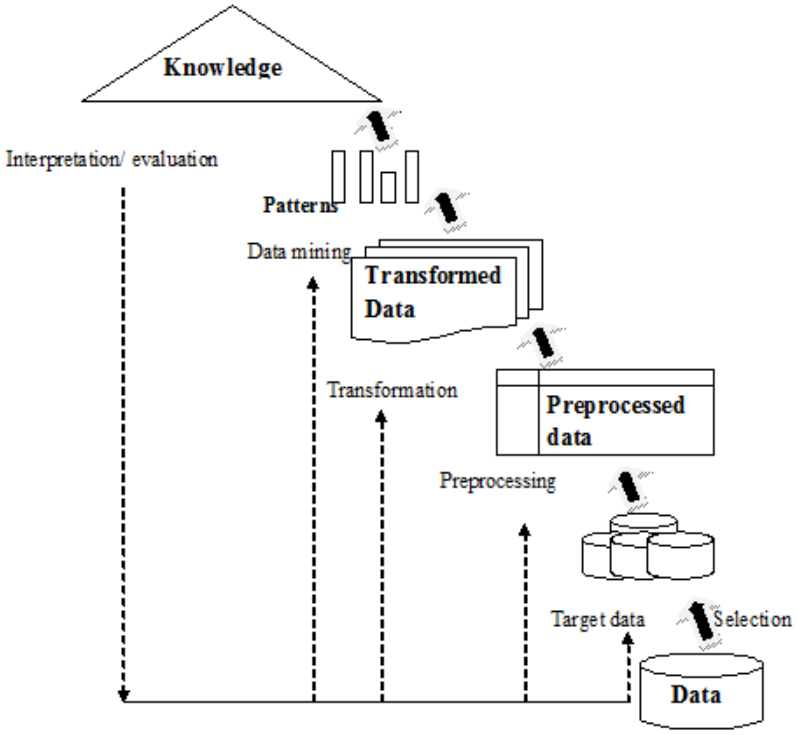

Figure 2: Data Mining

2.1 Data Mining Techniques

- **Correlation Analysis-** Finding item set model knowledge frequently appeared from given data set for the purpose of excavating the relationship that was hidden in the data.
- **Feature Selection-** A subset of features available from the data is selected for the application of a learning algorithm. It is used in machine learning.
- **Machine Learning-** It is the study of computer algorithms which automatically improve through experience.
- **Sequential patterns-** It is used to excavate connection between data, time series analysis gains more focus on the relationship of data in times.
- **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.
- **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.
- **Deviation analysis**- Finding abnormal data from the database.
- **Forecast**- Finding certain laws according to historical data, establishing models and predicting types, characteristics of the future data, etc based on the model.

2.2 Data Mining Technique in Intrusion Detection System

Data mining applied into intrusion detection can generate many concise and exact detection modes automatically from a great deal of audit data [7] collected through secondary/ primary source. In the analysis of intrusion detection system, the data circulating in network has the following characteristics: mass data, even if a small commercial website, the number of data message sent and received are quite impressive and incomplete whose transportation is busy, data message which overweigh network carry will be discarded: noisy, when network is unstable, data information may get changed in the transportation of message, we can see that these data is in accordance with the feature of object of data mining, naturally, we want to apply data mining technology to intrusion detection system[8].

Intrusion detection system is a passive method in the security field, it monitors information system and sends out warning when it does detect intrusion, but data mining technology can analyze these data when network message is acquired, it can forecast for visit on its own initiative, thus reduce the frequency of matching, and thus achieve the function of active defense [8]. Data mining is the process of discovering meaningful correlations, patterns and trend among the data by applying statistical, mathematical and machine learning techniques.

Data mining technology covered under descriptive and predictive methodology, for instance, 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 [8].

REFERENCES

1. Yusufovna S.F., “Integrating Intrusion Detection System and Data Mining”, International Symposium on Ubiquitous Multimedia Computing, 978-0-7695-3427-5/08, IEEE, 2008, pp.256-259.
2. Miao Chunyu, Chen Wei, “A Study of Intrusion Detection System Based on Data Mining”, 978-1-4244-6943-7/10, IEEE,2010, pp.186-189.
3. Heady et.al. “The architecture of a network level intrusion detection system”. Technical report, Computer Science Department, University of New Mexico, August 1990.
4. Scarfone, Karen; Mell, Peter (February 2007). "Guide to Intrusion Detection and Prevention Systems (IDPS)". *Computer Security Resource Center* (National Institute of Standards and Technology) (800-94). http://csrc.ncsl.nist.gov/publications/nistpubs/800-94/SP800-94.pdf.

5. Frawley et.al. "Knowledge Discovery in Databases: An Overview". AI Magazine, ISSN 0738-4602,pp.213-228.
6. Hand et.al." Principles of Data Mining", MIT Press, Cambridge, MA. ISBN 0-262-08290-X, 2001.
7. Lee W, Salvatore Department" Data Mining Approaches for Intrusion Detection [M]", New York, NY: Computer Science Department, Columbia University, 1996.
8. Liu Wei, "Research of Data Mining in Intrusion Detection System and the uncertainty of the attack", 978-1-4244-5273-6/09, IEEE, 2009.