A Study on Data Mining Based Intrusion Detection System

Anthony Raj.A
Department of computer Science,
Sri Bhagawan Mahaveer Jain College, Bangalore University
KGF Karnataka, India,Research Scholar / CSE
PRIST University, Thanjavur, INDIA
anthonyraj171@gmail.com

Abstract— In recent years security has remained unsecured for computers as well as data network systems. Intrusion detecting system used to safeguard the data confidentiality, integrity and system availability from various types of attacks. Data mining techniques that can be applied to intrusion detection system to detect normal and abnormal behavior patterns. This paper studies nature of** network attacks and **the current trends of data mining based intrusion detection techniques.

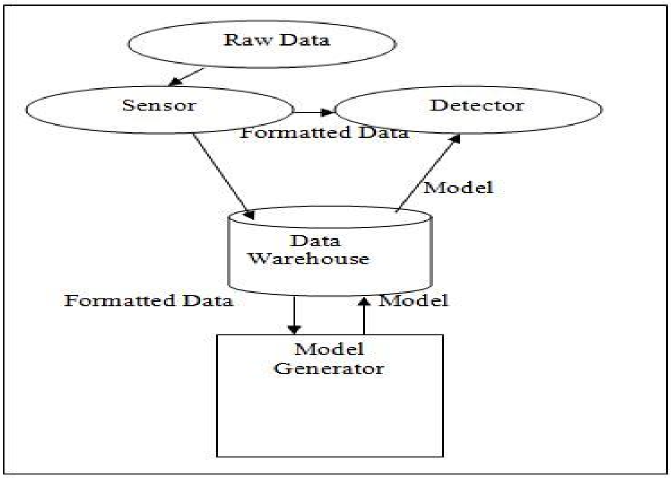
Keywords— Data Mining, Intrusion detection system, Anomaly Detection, Supervised Learning, Classification, Support Vector Machine

I. INTRODUCTION

Intrusions are the activities that violate the security norms of the system. An Intrusion Detection system is Mechanism used to identify, monitor network or system actions for malicious activities and produces reports to a management departments. The development of IDS is motivated by the following factors: Most existing systems have security was that render them susceptible to intrusions, and finding and fixing all these deficiencies are not feasible. Prevention techniques cannot be sufficient. It is almost impossible to have an absolutely secure system. Even the most secure systems are vulnerable to insider attacks. New intrusions continually emerge and new techniques are needed to defend against them [1].

Data mining is a process to extracting (mining) knowledge from large amount of data.(KDD: Knowledge discovery from data) using computational technique and it is used for strategic decision making, wealth generation,anlyzing trends and security purpose. Data mining overlaps with many diciplines like statistics, Machine learning, information retrieval, distributed computing. In each dicipline approach of data mining keeps different ways. Information in Data mining process basically categorized in to two braod categories and threse are descriptive and predictive information. In Descriptive we find pattern that human can intrepretable. In predictive information is a process of finding a value of an attribute using value of other attributes. This paper surveys current existing trends of data mining based intrusion detection techniuqes which are being used for intrusion detection in a data networks and internets.

II. DATA MINING BASED IDS ARCHITECTURE



The Architecture consist following major components [2], [4]

1. Sensors: Used to detect raw data on a monitored system and compute fetures for use in model evaluation
2. Detectors: Recive processed data from sensors and use detection model to evaluate the data and examine if it is an attack

3. A data warehouse: Served as centralized data source for data and models. Integrates data from multiple sensors.
4. Model Generator: Facilitates the rapid development and distribution of new or updated intrusion detection models

III. CURRENT TRENDS OF IDS TECHNIQUES

3. Signature based method:

This is the traditional mehtod for intrusion detection. It requires extensive knowldege of signature of previously known attacks. In this process monitored events are matched with the signature to detect intrusion. Data is extrated from various different audit database and comparing these features with to a set of attack signature provide by human expert for intrusion detection [1], [4]

3.1. Most Popular Techniques

A. Anomaly Detecion

Detects the abnormal behaviours of host or network. It stores the features of user's usual behaviours hooked on database, and then it compares user's present behaviour with database. The deviation of the monitored traffic from the normal profile is measured. [1], [4]

B. Misuse Detection

It works by searching for the traces or patterns of well-known attacks.
There are two steps to be followed:
Step one: Define abnormal system behaviour:
Step Two: Define any other behaviour, as normal behaviour.
Deviations from these rules indicate an attack on the network

3.2. Recently emerged methods from Machine learning

A. Supervised Learning-Based Approaches: Can detect known attack and pattern recognition have been utilized to detect intrusions.

B. Unsupervised Learning-Based Approaches: Can detect the intrusions that have not been previously learned. [3]

3.3. Classification of IDS

Based on Sources of audit information IDS divided into different types [1],[5]

A. Host Based IDS: Audit data held on individual computer that serve as hosts. Intrusion detection takes place on a single host system.

B. Network Based IDS: Network traffic considered as audit data source. Used to provide normal computing services and detect attacks from network.

C. Distributed IDS: Gather audit data from multiple hosts that connected by the network. Used to detect attacks involving multiple hosts

D. Hybrid intrusion Detection: It is a combination of both host-based and network-based IDS. It provides flexibility and increases the security level.

3.4. Major Problems with Current IDS

A. Data overload: Datasource/Audit data which needs to be analysed for intrusion detection must be discrete volume/size for efficient and effective analyze. Data overload is major problems of the Current IDS

B. False Positives : When IDS treats normal attack as malicious then it considered as False poitives

C. False Negatives: When IDS does't generate an alert/alarm when intrusion actually taking place then it is considered as False Negatives[1],[5]

IV. TYPES OF NETWORK ATTACKS.

A Network attack or Security or Security incident is defined as a threat, intrusion, and denial of service or other attack on a network infrastructure that will analyze your network and gain information eventually and cause your network to crash or to become corrupted. There are at least seven types of network attacks. [5], [6]

A. **Spoofing (Identity spoofing or IP Address Spoofing)**
Any internet connected device necessarily sends IP datagrams into the network. Such internet data packets carry the sender's IP address as well as application-layer data. If the attacker obtains control over the software running on a network device, they can then easily modify the device's protocols to place an arbitrary IP address into the data packet's source address field. This is known as IP spoofing

B. **Sniffing.**
Packet sniffing is the interception of data packets traversing a network.

C. **Mapping.**
Before attacking a network, attackers would like to know the IP address of machines on the network, the operating systems they use, and the services that they offer. With this information, their attacks can be more focused and are less likely to cause alarm. The process of gathering this information is known as mapping.

D. **Hijacking.**
This is a technique that takes advantage of a weakness in the TCP/IP protocol stack, and the way headers are constructed. Hijacking occurs when someone between you and the person with whom you are communicating is actively monitoring, capturing, and controlling your communication transparently

E. **Trojans.**
These are programs that look like ordinary software, but actually perform unintended or malicious actions behind the scenes when launched.

F. **DoS and DDoS.**
A denial of service attack is a special kind of Internet attack aimed at large websites. It is a type of attack on a network that is designed to bring the network to its knees by flooding it with useless traffic. Denial of Service can result when a system, such as a Web server, has been flooded with illegitimate requests, thus making it impossible to respond to real requests or taks.

G. **Social engineering.**
Social engineering is the use of persuasion or deception to gain access to information systems. The medium is usually a telephone or e-mail message.

H. **User to Root Attack(U2R)**
The Attaker take down user level password and finally achieves root to access the system

I. **Remote to User Attack(R2U)**
Class of attacks where attaker sends packets to a machine over a network, then exploits machine vulnerability to illegal gain local access as a user.

V. APPLIED DATA MINING BASED INTRUSION DETECTION TECHNIQUES

Data mining look for hidden patterns and trends in data warehouse that is not immediately apparent from summarizing the data, and there is no query involved but use the concept interestingness criteria i.e specification of data such as Frequency,Rarity,Correlation, Length of occurrence, Consistency, Repeating/ periodicity,abnormal behaviour, and other patters of interestingness.

The algorithms which are used for intrusion detection based on datamining techniques are listed as follows

5.1 Association rule

Association rules mining identifies association among database attribures and their values. It is a pattern-discovery technique which does not serve to solve classification problems nor predict problems. Association rule mining requires two thresholds i.e Minimum support and Minimum Confidence. Example: Apriori for mining Association rules Algorithm. [Agrawal and Srikant, 1994] [1] and [2]

5.2 Classification

Classification is the process of learning a function that maps data objects to a subset of a given class set. There are two goals of classification, First fininding a good general mapping that can predict the class of so far unkown data objectswith high accuracy. Second to find a compact and understandable class model for each othe classes [1] and [2]

5.3 Clustering techniques

Clustering group’s data elements into different groups based on the similarity between within a single group Cluster partitions the data set into clusters or equivalance classes. Cluster methods divided into two categories based on the cluster structure namely Non Hierachical and Hierarchical –connection oriented. [1], [2], [3]

5.4 Decision Tree

Decision tree initially builds a tree with classification. Each node represents a binary predicate on one attribute, one branch represents the positive instances of the predicate and the other brach represents the negative instances. Construction of Decision Tree does not require any domain knowledge and can handle high dimensional data. [3][4]

5.5 Genetic Algorithms

Method: learning examples are stored in relational database that are represented as relational tuples. It solves the problems with multiple solutions and easily transferred to existing models [3][4]

5.6 K Nearest Neighbour

An object classification process is achieved by the majority vote of its neighbours. The object is being assigned to the class most common amongst its k nearest neighbours. If k=1, then the object is simply assigned to the class of its nearby neighbour. Its Implementation taks is simple and Easy for parallel implementations.[2]

5.7 Support vector Machine

Method: A support vector machine is a classification and regression technique it constructs a hyperplane or set of hyper planes in a high or infinite dimensional space. It is able to model complex and nonlinear decision boundaries. [4]

5.8 Neural Network

Method: A Neural Network is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Implicitly detect the complex nonlinear relationships between dependent and independent variables. Highly tolerant the noisy data. Availability of multiple training algorithms.[3] [4]

5.9 Bayesian Method

Bayesian classifier based on the rules. It uses the joint probabilities of sample classes and observations. The algorithm tries to estimate the conditinal probabilities of classes given an observation. Naïve Bayesian Classifier simplifies the computaions It exhibit high accuracy and speed when applied to large databases.[4]

5.10 Fuzzy Logic

The Fuzzy logic has been used for both anomaly and misuse intrusion detection. It uses linguistic variables and allows imprecise inputs, permits fuzzy thresholds. Rule base or Fuzzy sets easily modified.[3][4]

VI. COMPARATIVE STUDY ON APPLIED DATA MINING BASED INTRUSION DETECTION TECHNIQUES

YEAR	PAPER NAME	TECHNIQUE	MERITS	DEMERITS
Dec 2012	A Survey on Intrusion Detection using Data Mining Techniques [1]	I. Association rule or Dependency Mining II. Classification & clustering	Used in transaction data analysis Applied for KDD task Unsupervised technique	----
Feb 2013	A Review of Data Mining based Intrusion Detection Techniques[2]	I. Novel IDS II. K-means Clustering Algorithm III. Data Dependency weighted sequence Mining IV. Hybrid IDS KDD, Anomaly Detection	Used to detect Dos Attack Detect black hole attack Used to filter out extra rules generated by this appraoch Combines the filter and wrapper models for selecting relevent features Investigate more efficient methods against intrusions	Not reliable Doesn't provide sufficient mining method Architecture needs to be enhanced for improvement Needs to be enhanced for Cryptographic mechanism. Required to survey more recent techniques
April 2013	Survey paper on Data Mining techniques of Intrusion Detecion[3]	I. Feature Selection II. Machine learning II. Hybrid approach	Used on finite data set Improve automatically through experience ----	
June 2013	A Survey on Intrusion Detection System in Data Mining[4]	I. Data Mining , Feature Selection, Multiboosting II. K- means clustering Distributed IDS	Find high detection rates for U2R and R2L and also to detect attacks False alarm rate has been decreased also clustering helps in to identify the attacked data.	
Oct 2013	A Survey: Network Intrusion Detection system Based on Data Mining Techniques[5]	I. Support Vector Machine II. Genetic Algorithm III. K nearest Neighbor IV. Neural Network V. Bayesian Method	High Accuracy solves optimization problem Simple and highly adaptive behaviour Implicitly detect the complex nonlinear Relationships between dependent and independent variables Simplifies the computations Exhibit high accuracy and speed when applied to large database	More Time space complexity No global optimum High Storage requirement It requires long training time The assumptions made in class conditional independence Lack of available probability data

VII. CONCLUSIONS

The paper showed several intrusion detection tools to uncover known and unknown attacks and to model the behaviour of the user. The shortcomings of each tools either using independently or combining with other tool for targetting specific attacks reveals that the outcome have other side effects or shortcomings inturn. Thre is a much research socpe involved for the research community in this field to find the right kind of generalization of the IDS model .i.e it should not be neither too general nor too specific which is not reliable. The challenges to find solution to the new emerging attacks with using current data mining based intrusion detecion techniques in different fields is a new research domain

ACKNOWLEDGMENT

I Give thanks to God and Father of our Lord Jesus Christ, the merciful Father, the God from whom all help Comes! And wish to acknowledge Dr.A.Arul L.S., my Advisor and other known & unkown research scholars for their support and sharing ideas, views in their respective papers which really gave me an inspiration to do A Study on Data Mining Based Intrusion Detection System

REFERENCES

[1] R.Venkatesan, Dr.R.Ganeshan, Dr. A.Arul Lawrence Selvakumar “A Survey on Intrusion Detecion using Data Mining Techniques” in International journel of Computers and Distributed System, December 2012

[2] Kamini Maheshwar and Divakar Singh “A Review of Data Mining based Intrusion Detection Techniques” in International Journal or Innovation in Engineering & Management (IJAIEM) Feb 2013

[3] Harshna and NavneetKaur “Survey paper on Data Mining techniques of Intrusion Detecion “ in International Journal of Science, Engineering and Technology Research (IJSETR), April 2013

[4] Sahilpreet Singh, Meenakshi Bansal “A Survey on Intrusion Detection System in Data Mining “ in International Journal of Research in Computer Engineering & Technology (IJARCET), June 2013

[5] Subaira.A.S and Anitha.P "A Survey: Network Intrusion Detection system based on Data Mining Techniques” in International Journal of Computer Science and Mobile Computing, October, 2013

[6] http://ayurveda.hubpages.com/hub/Types-of-Network-Attacks