A Review on Intrusion Detection System in Data Mining

Monika Boora [#1], Dr. Chhavi Rana *[2], Tarika Verma [#3]
[#] M.Tech Student, CSE Department, UIET, MD University, Rohtak, Haryana, India
[1]Email: monikaboora93@gmail.com
[3]Email: tarika.verma@yahoo.co.in
[*] Assistant Professor, CSE Department, UIET, MD University, Rohtak, Haryana, India
[2]Email: chhavi1jan@yahoo.com

***Abstract*---Abstract Various IDSs has been developed by manual encoding of expert knowledge. To change the IDS require high cost and speed of changes is also slow. This paper shows a data mining framework to build Intrusion Detection (ID) models. The survey also aimed to import ideas and opinions of members who feel how much is the real need of data mining in a particular field. Since data warehouses are the basics of data mining, different points concentrated on present support and future ideas for data warehousing.**

Keywords: Algorithm, Current technique, Intrusion Detection system, Meta –Classification, Proposed Technique.

1. INTRODUCTION

Intrusion defined as a malware that causes harm to the system. An Intrusion detection system is a method used for malicious actions to monitor them after finding the intrusion, the work of fixing the intrusion is to be performed. Even the most protected systems are facing insider attacks. New intrusions continually emerging and new operations are needed to fight against them [1].

The term Data mining has successful applications in different fields like marketing & retail, e-business etc. It is the method of extraction of the required data from the large data. It co-operates different industries & Organisations to focus on the most important tasks rather than on to the whole warehouse. Systems are in favour of mining of data concept of various hardware & software platforms for enhancing their existing information resources & Upgrading with New Products and the motive of this survey is to determine the extent to which data mining technology is being helpful in different detection system. The paper also aimed to elicit ideas and opinions of members who feel how much is the real need of mining in a particular field. As we know data warehouse is a backbone of data mining for storage of the data. Present support and future ideas for data warehousing focused on several points. Mining of data is said to be a discovery of required knowledge as it analyzes the various large set of data and extract the unknown and necessary data.

The term data mining came from the equivalent features b/w searching for valuable business information in large databases. It decreases the work of the users who wants to search the predictive data & automates the searching process so that the time must be saved. The evolution of mining of data has begun when businessmen's first stored the data of their business onto the computers. Data mining works in different fields of statistics, machine learning, information The paper reviews the existing trends of detection of intrusion techniques which are depend upon data mining and which are being used for detection of attacks in a data networks and internets.

2. CURRENT TECHNIQUES OF INTRUSION DETECTION SYSTEM

AMost Popular Techniques

1. Anomaly Detection

It is the detection of unknown behaviours of host or network. Then it compares user's present behaviour with the database. The deviation of the analysed traffic pattern with the defined patterns is measured [1] [2].

2. Misuse Detection

It includes the searching of the traces or patterns of well-known attacks.

There are two steps to be followed:

Step 1: Define abnormal system behaviour:

Step 2: Define normal behaviour as any other behaviour

Deviations from these rules indicate an attack on the network.

B. Signature based method

It is the traditional method for detection of intrusions. It requires large information of signature of previously known attacks. In this process analysed events are matched with the signature to detect intrusion. Data is taken from different audit database and comparing these features with previously defined signatures provide by the human expert for intrusion detection.

C.Types of IDS

IDS divided into different types based on Sources of audit information [1][3].

1. Host Based IDS*:* An individual computer that serves as hosts held on audit data. Intrusion detection works on a single host system.

2. Network Based *IDS:* Audit data source takes place on the network traffic. Used to serve normal computing services and detect attacks from the network.

3. Distributed IDS: The data that is audited is collected from multiple hosts connected by the network is gathered. Attacks involving multiple hosts are detected.

4. Hybrid intrusion Detection: It is a combination of both host-based and network-based system IDS. It provides tensile feature and enlarges the security purposes.

D. Algorithm for Mining Audit Data

In this paragraph, we discuss about algorithms of mining of data, and discuss to apply these algorithms to generate detection models from the audited data. Here audit data is the previously processed audit records which have time stamped, each with a number of features (i.e., fields).

1. Classification:

The mapping of the data item into one of the several predefined categories is known to be classification. These algorithms which we us in generally outputting “classifiers”, for example, in the way of decision tree or rules. To extract “normal” and “abnormal data”, the data which is audited is used by programmer is an application of IDS. Then, a classification algorithm will be applied to learn a classifier that can label or predict new unseen audit data that belongs to the audit data that belongs to the normal class or the abnormal class.

.2.Link analysis:

In link analysis, fields related to the database records are determined. System features in audited data that are correlated founded in link analysis, for example, the common features between command and argument in the shell command history data of a user can perform as the basis for constructing normal usage profiles. The person which is doing the programming, for example, may have included “emacs” highly linked with “C” files.

3 Sequence analysis:

This analysis works on the sequential patterns. These algorithms also discover the data which comes repeatedly together at time-based sequence audit events. These event patterns which occur frequently provide the guidelines for temporal statistical measures which are incorporated into intrusion detection models. For example, the data of audit have patterns of network-based denial of service(DOS) attacks suggest that several per-host and per-measures should be included.

4. Meta-Classification*:*

In this, we study the set of multiple detection models. First, we takes the performance issue & try to reduce them and an easy target of “subversion”, an IDS should consist of multiple cooperative lightweight subsystems that each monitor a separate part (e.g., the access point) of the whole network environment.

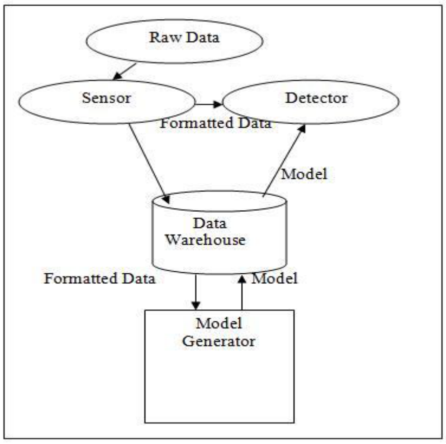

Fig 1: IDS Architecture based on Data mining

The Architecture has following major components [2][4]

1. Sensors: A monitored system is used to detect/sense the source (raw) data & extract features to use in the model evaluation.

2. Detectors: These detectors takes processed data from sensors and apply detection model to extract the data and examine if it is an attack.

3. A data warehouse: It works as a central system for all the data and models. It integrates the data received from multiple sensors.

4. Model Generator: Evaluate the fast development and distribution of new or updated intrusion detection models.

Privacy-Preserving Methods

The privacy preserving is used for the privacy of the data which is under intrusion detection. This can be done by working on the different types of approaches. The approaches use modified values that are generated by using custom code.

There are various methods for PPM as follows:

Value class membership

Value distortion technique

III. PROPOSED TECHNIQUE FOR IDS

The Proposed technique is that there should be a design in which the data directly after mining comes to the user for analysis of attacks. This whole process will be done in a sequence. Let it clear by an example suppose anyone wants to download any file from the internet. User clicks to download the data, then during the downloading, the intrusion detection also done with it. Some errors are removed in the mining process. Now only the useful data is to be scanned for the removal of attacks. If there is any intrusion in that data, it will be removed after scanning. This process will also have the authentication & security features so that any attacker could not able to attack or read our useful important data. The Detection system will be effective in both offline & online mode. As we have seen that various systems have the detection property in online mode only. So, I am developing a system such that offline users also have the safe data and the intrusions will also remove in the offline mode also. The proposed model is under the processing as I m working on it by the software weka tool.

1. Weka tool*:-*

Weka is a tool used for the data mining process. It contains the set of algorithms for data analysis & predictive modelling. The Proposed model is under the processing as it is not designed up to now but working on it that it should be like this technique. Weka supports various tasks like arrangement, grouping, & visualization etc.

IV. CONCLUSION

The challenge is to find a new research technique for mining of only the required data in the IDS field. The key idea is to apply this model cover the various intrusion detection tools to uncover known and unknown attacks to model the behaviour of the user. There are different shortcomings are also present in the paper with the various approaches used in the model. For data mining, the IDS system is surely a better idea in my view as this is not data mining programs to audit data to determine abnormal and anomaly detection models, according to the observed behaviour in the different types of data.

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] 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

[4] 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

[5] W. W. Cohen. Fast effective rule induction. In Machine Learning: the 12th International Conference, Lake Taho, CA, 1995. Morgan Kaufmann.

[6] Principles of Data Mining by David Hand, Heikki Mannila and Padhraic Smyth

[7] Data mining for intrusion detection system by Aleksandar Lazarević, Jaideep Srivastava, Vipin Kumar

[8] 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

[9] Privacy-Preserving Data Mining Rakesh Agrawal Ramakrishnan Srikant IBM Almaden Research Center 650 Harry Road, San Jose.

[10] A Data Mining Framework for Building Intrusion Detection Models_ Wenke Lee Salvatore J. Stolfo Kui W. Mok Computer Science Department, Columbia University

AUTHOR PROFILE

Monika Boora is currently pursuing Master in Technology degree at MD University, Rohtak, Haryana, India. She has also got TEQUIP World Bank scholarship for the meritorious students and has been teaching under the same scheme since last two years. She has been interested in the field of Computer Networks, Data Mining based Systems and Big Data Analysis. She has attended conference and also has presented papers related to this field.

Dr. Chhavi Rana Balhara has an experience of over 10 years teaching Data Mining and Web Development subjects at various engineering institutes. She has been interested in the area of Data Mining research from the past 8 years attending around 25 conferences and presenting papers related to this field. Also, She has published 30 papers in reputed journals including Springer and Elsevier. Besides Data Mining, her research interests also include information management, information retrieval, and ICT. She has supervised 20 M.Tech thesis and currently supervising 4 M.Tech thesis and 4 Ph.D. students. She has also won DST Travel grant Twice to present the paper in USA and Spain as well as TEQUIP World Bank travel grant to present the paper in University of Sydney, Australia. She has also been a Reviewer on IEEE Transaction's on Systems, Man and Cybernetics: Systems, Artificial Intelligence Review, Springer as well as Inderscience Publishers. She has also published 4 books on her research work.