ISSN No.: 2454- 2024 (online)

International Journal of Technical Research & Science

REVIEW OF INTRUSION DETECTION SYSTEM TECHNIQUE USING DATA MINING

Salona Ranga
Email Id: salonaranga@gmail.com
Deenbandhu Chhotu Ram University of Science and Technology, Sonipat, Haryana (India)

Abstract-With the high usage of Internet in our day to-day life, security of network has become the key foundation all web applications. Therefore every entity wants to protect its data from both internal and external attackers. Firewall, encryption, and authentication serve as the first line of security Intrusion Detection serves as the second line of security. An intrusion detection system (IDS) monitors networked devices and looks for malicious behavior so as to detect anomaly and misuse. Misuse detection is usually conducted by signature matching of known attacks. It effectively protect against common attacks but cannot detect 'newly invented' attacks. Anomaly detection is generally based on machine learning but anomaly detection may have a higher rate of false positives.

1. INTRODUCTION

Intrusion Detection System (IDS) is a active component of any network in today's world of Internet because IDS are an real way to detect different kinds of attacks in interconnected network and it requires high accuracy and detection rate as well as low false alarm rate[3]. Intrusion detection technology refers to identify any activities of loss to the computer system security, integrity and confidentiality [5]. The aim of an intrusion detection systems (IDS) is to detect various types of malicious network traffic and computer usage, which cannot be detected by a conventional firewall.

Index terms: Ids, Security, Challanges, function of Ids, types of Ids

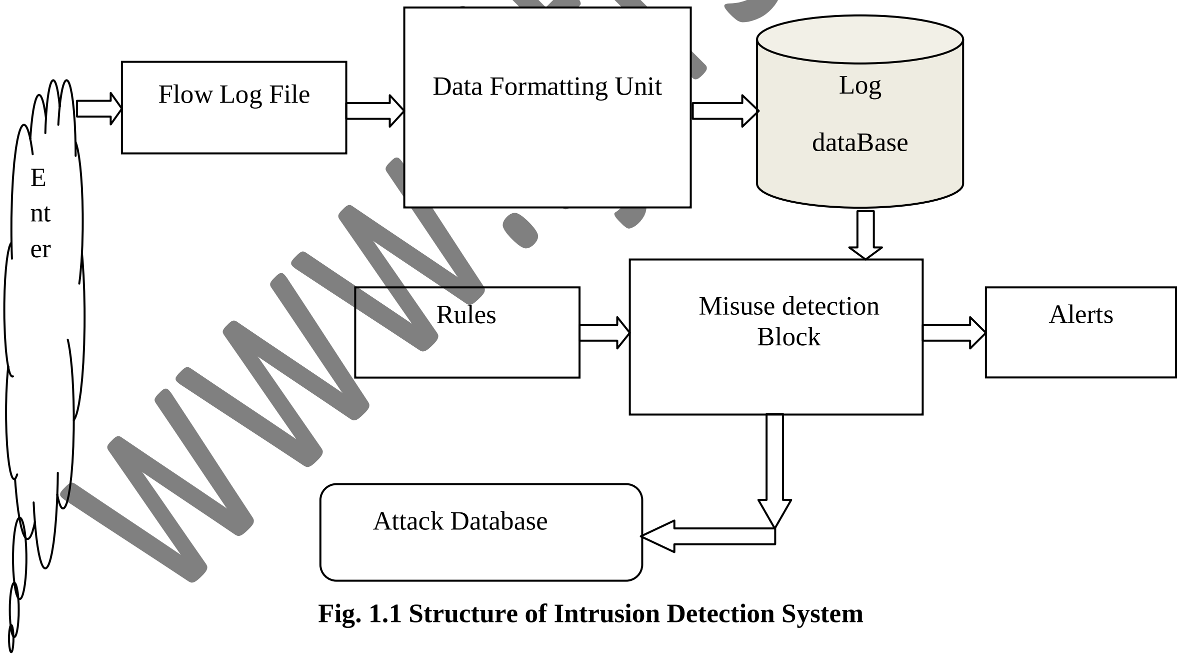

Fig. 1.1 Structure of Intrusion Detection System

2. FUNCTION OF INTRUSION DETECTION SYSTEMS

Intrusion detection system performs many functions which are dynamic for the system. These are as follows:

- It Collect appropriate data and establishing an event database.
- Establishing a knowledge base to characterize the normal behavior or the attack signatures.
- Designing appropriate algorithms to analyze monitored events and perform intrusion detection based on certain criteria or further evidence inference[1].

3.CHALLANGES IN INTRUSION DETECTION SYSTEM

Challenges faced by IDS are discussed below:

3.1 Current Intrusion Detection System

These systems are usually changed to detect known service level network attacks. This leaves them vulnerable to original and novel malicious attacks.

3.2 Data Overload

Another aspect which does not related directly to misuse detection but is particularly important is how much data an analyst can efficiently analyze. The amount of data that needs to be analyzed seems to be growing rapidly.
Depending on the intrusion detection tools working by a company and its size there is the possibility for logs to reach millions of records per day.

3.3 False Positives

A common objection is the amount of false positives an IDS will generate. A false positive follows when normal attack is incorrectly classified as malicious and treated accordingly.

3.4 False Negatives

This is the case where an Intrusion Detection System does not generate an alert when an Intrusion is actually taking place. (Classification of malicious traffic as normal) Data mining can help improve Intrusion Detection by addressing each and every one of the above mentioned problems.

3.5 Current Intrusion Detection

System does not detect the Novel Intruders.

3.6 Intrusion Detection Systems are expensive and slow [10].

- Centralized view of the data
- Data transformation capabilities
- Analytic and data mining methods
- Flexible sensor deployment, including scheduling that enables periodic model relation and distribution Real-time detection and alert infrastructure reporting skills distributed processing high system availability scalability with system load[3].

4. RELEVANCE OF INTRUSION DETECTION SYSTEM

Firewalls cannot ward off all outside attacks on the network and are useless to defend against inside attacks. Authentication systems cannot prevent valid users from carrying out damaging operations on a computer system. Intrusion detection is a technology for discovering attacks against computer network systems from both outside and inside. Intrusion detection together with firewall and authentication technology constitutes most of the state-of-the-art defense-in-depth framework and is considered as effective for computer network security[1].

5. SECURITY SUPPORT BY INTRUSION DETECTION SYSTEM

Provide securities for the relevant system these securities are [8].

5.1 Data Confidentiality

It checks whether the information stored on the system is protected by unauthorized access. Since systems are sometimes used to manage sensitive information, data privacy is often a gauge of the ability of the system to protect its data.

5.2 Availability

The network should be rough to Denial of Service attacks. Intrusion detection system based on sources of examination information it can be divided into 3 subcategories.

5.3 Data Integrity

It refers to maintaining and assuring the correctness and consistency of data over it entire life-cycle. No corruption or data loss is acknowledged either from random events or malicious movement.

6. TYPES OF INTRUSION DETECTION SYSTEM ARE DISCUSSED

6.1 Misuse Detection

The intrusions in terms of the characteristics of known attacks or system vulnerabilities. It extract feature from known intrusions and integrate the Human knowledge. The rules are pre-defined however, it has disadvantage that it cannot detect novel or unknown attacks.

6.2 Anomaly Detection

Detect any action that significantly differs from the normal behavior is considered intrusion. It has a disadvantage that when a noise (intrusion) data is in training data, it will make a misclassification. Data Mining has found wide applications for last two decades and Network Security is not left untouched. The talk will focus on applying classification and association rule mining for anomaly-based intrusion detection in the network[4].

7. LITERATURE REVIEW

Jabez Ja, Dr.B.Muthukumarb[9] An Intrusion Detection System (IDS) is a software application or device that monitors the system or activities of network for policy damages or malicious activities and generates reports to the management system. A number of systems may try to stop an intrusion go but this is neither required nor expected of a monitoring system. The main focus of Intrusion detection and prevention systems (IDPS) is to identify the possible actions, logging information about them and in report efforts. In addition, organizations use IDPS for other purposes, like identifying problems with security policies, preventing individuals and documenting current threats from infringing security policies. Many methods can be used to detect intrusions but each one is particular to a specific method. The main goal of an intrusion detection system is to detect the attacks efficiently[9].

Wenke Lee Salvatore J. StolfoKui W. Mok, [13] There is frequently the need to update an installed Intrusion Detection System (IDS) due to new attack methods or improved computing environments. Since many current Intrusion Detection Systems are constructed by guide encoding of expert knowledge, changes to Intrusion Detection System are expensive and slow. We describe a data mining structure for adaptively building Intrusion Detection (ID) models. The vital knowledge is to utilize auditing programs to remove an extensive set of features that describe each network connection or host session, and apply data mining programs to learn rules that correctly capture the behavior of intrusions and common activities. These rules can then be used for misuse detection and anomaly detection. New detection models are combined into an existing Intrusion Detection System through a meta-learning(or co-operative learning) procedure, which produces a meta detection model that combines evidence from multiple models. We discuss the powers of our data mining programs, namely, classification, meta-learning, association rules, and frequent episodes [13].

Zhongmin Cai,1Xiaohong Guan,1,2 Ping Shao,1 QingkePeng1 and GuojiSun,[14] Intrusion Detection is important in the protection in- depth network security framework. An effective method for anomaly intrusion detection with low overhead and high efficiency. The method is based on rough set theory to remove a set of detection rules with a minimal size as the normal behavior model. It is capable of detecting the abnormal working status of a process and thus reporting a possible intrusion. Compared with other methods, the method needs a smaller size of training data set and less effort to collect training data and is more suitable for real-time detection. Network security is gaining worldwide attention. The security of computer network systems cannot be sure if it merely depends on conventional peripheral protection mechanisms such as firewalls and various authentication methods. Firewalls cannot area off all outside attacks on the network and are useless to defend against inside attacks. Authentication systems cannot prevent real users from carrying out harmful operations on a computer system. Intrusion detection is a technology for detecting attacks against computer network systems from both outside and inside[14]**.**

Peyman Kabiri and Ali A. Ghorbani,[15] Research on Intrusion Detection and Response: A Survey[]In the past two decades with the fast progress in the Internet based technology, new application areas for computer network have emerged. At the same time, widespread progress in the Local Area Network (LAN) an Wide Area Network (WAN) application areas in business, financial, industry, security and healthcare sectors made us more dependent on the computer networks. All of these application areas complete the network an attractive target for the misuse and a big weakness for the community. A fun to do job or a challenge to win action for some people became a nightmare for the others. In many cases malicious acts complete this nightmare to become a reality. In addition to the hacking, new entities like worms, Trojans and viruses introduced more alarm into the networked society. As the current state is a relatively new phenomenon, network defenses are weak. Securing such an important infrastructure has become the priority one research area for many researchers [15].

Chakchai So,[16] As a huge increase of Internet user population to more than two billions , more people are victims of cyber attacks, estimated at more than five millions a year. This problem has increased the maintenance to cyber crimes into both research and industry in order to secure the network. Several techniques, methods, policies, and systems have been proposed to improve the threat and attacks through the Intrusion Detection Systems (IDS) used to counter measure against intrusions and malicious efforts. IDS can monitor actions at the endpoints or on the network, i.e., host and network–based methods [16].

Wang Pu,[17] The latest developments in computer systems and the internet have changed the way people think and do things. A process like sending old mail that normally takes hours or even days can now be completed in a click of a mouse or a touch of a finger complete electronic mail or e-mail. People communicate with each other from is future in order to lessen the false negative rate and a system for automatically identifying the number of clusters may be developed. Different places through integrated relay chat, or video conferencing as a much suitable mode of communication. However, risks are associated with these technologies. Over the last two times, computer threats and cybercrimes have multiplied at the disadvantage of the general public, threats are introduced each day that compromise the integrity, validity and confidentiality of data. Companies, populations, and individual persons can be losses of malicious activities in the internet there by secure the network. A successful Intrusion Detection System requires high correctness and detection rate as well as low false alarm rate. This focuses on a hybrid move on for Intrusion Detection System (IDS) based on data mining techniques[17].

G.V. Nadiammai, M. Hemalatha 2014 [6] In this work, data mining concept is integrated with an Intrusion Detection System to identify the relevant, hidden data of interest for the user successfully and with less execution time. Four issues such as Classification of data, High Level of Human Interface, Lack of Considered Data, and Effectiveness of Distributed Denial of Service Attack are being solved using the future algorithms like EDADT algorithm, Hybrid IDS model, Semi-Supervised Method and Varying HOPERAA Algorithm respectively. Our proposed algorithm has been tested using KDD Cup dataset[6].

Ashish Dutt, Maizatul Akmar Ismail, and Tutut Herawan[18] In this field of study, Educational Data Mining (EDM) applies machine-learning, statistics, Data Mining (DM), psycho-pedagogy, information retrieval, cognitive psychology, and recommender systems methods and techniques to various educational data sets so as to resolve educational issues. While Data Mining, also referred to as Knowledge Discovery in Databases (KDDs), is a known field of study in life sciences and commerce, yet, the application of Data Mining to educational context is limited [3]. It is an unsupervised approach for analyzing data in statistics, machine learning, pattern recognition, and Data Mining. It refers to collecting similar objects together to form a group or cluster. Each cluster contains objects that are similar to each other but dissimilar to the objects of other groups. This approach when applied to analyze the dataset derived from educational system is termed as Educational Data Clustering (EDC[18].

S. SobinSoniya[19] These networks are the supports of the industries like banking, transport, healthcare, defense, communication etc. So securing the data passed finished these networks is necessary. Organizations are investing more and more money to secure their data from the attackers. On the other hand, the attackers are success stronger day by day. This Intrusion Detection System techniques are used to protect the network from the attackers. Also in the upcoming days our research will focus on structure an improved system to detect the intruders and to secure the network from the attackers. It involves allowing only the authorized person to access the data in a network which is controlled by the network administrator. Network security helps to prevent and monitor unauthorized access to the system, misusing the system, modification of information, or denial of a data services and network accessible resources [19].

Urvashi Modi1 and Anurag Jain[11] An intrusion detection system detects various malicious behaviors and abnormal activities that the force harm security and trust of computer system. IDS operate either on host or network level via using anomaly detection or misuse detection. Main problem is to correctly detect intruder attack against computer network .To resolve the problems of IDS scheme this research work suggest “an improved method to detect intrusion using machine learning algorithms”. Intrusion Detection System (IDS) has been experiential as the ”silver bullet” that assurances safety of an organization system beside possible attacks. When this method, it’s not successfully utilized due to the huge measure of false alarms that it generates. For example, the well identified open source Intrusion Detection System Sniff technique is performs on a network with few hundred machines and it generates thousands of warnings daily, which holds a bulk of false alarms[11].

Srikanth Yadav [12] Thus to analyze data and to determine various kind of attack data mining techniques have occurred to make it less vulnerable. Anomaly detection uses these data mining techniques to detect the unforeseen behavior hidden within data increasing the chances of being intruder or attacked. This approach is similar to other measures such as antivirus software, firewalls and access control schemes. Usually, these systems have been categorized as a signature detection system, an anomaly detection system or a hybrid detection system. In signature based detection, the system identifies patterns of traffic or application data is acknowledged to be malicious while anomaly detection systems compare activities compared to a normal defined behavior. Hybrid Intrusion Detection Systems combine the techniques of both these methods. Firstly, they are capable of detecting insider attacks. Secondly, the detection system is based on practice made profiles. It develops very difficult for an attacker to carry out any activity without set off an alarm. Finally, it can detect the attacks that are previously not known. Anomaly detection systems look for anomalous events rather than the attacks[12].

8. PROCESS OF INTRUSION DETECTION SYSTEM ARE DISCUSSED

8.1 Data Mining, KDD and related fields

Data Mining(DM),also called Knowledge Discovery and Data Mining, is the process of repeatedly searching large volumes of data for patterns using association rules. Data mining is the process of removing useful information from large databases. The term knowledge discovery in databases (KDD) is used to denote the process of extracting beneficial knowledge from large data sets. Here, we mostly outline some of the most basic KDD steps:

- Understanding the application domain: First is developing an understanding of the application domain, the relevant background knowledge, and the specific goals of the KDD endeavor.
- Data integration and selection: After understanding the application domain, the integration of multiple data sources and then selection of the subset of data that is related to the analysis task.
- Data mining: Third is the application of particular algorithms for extracting patterns from data.
- Pattern evaluation: Fourth is the analysis and validation of the discovered patterns. The goal of the step is to assurance that actual knowledge is being discovered [6].

8.2 Data Mining and Intrusion Detection System

8.2.1Anomaly Detection or Profile Matching

This technique is based on the normal behavior of a subject (e.g., a user or a system) any action that significantly differs from the normal behavior is measured as an intrusive action. Misuse detection catches intrusions in terms of the features of known attacks or system vulnerabilities any action that conforms to the pattern of a known attack or vulnerability is measured intrusive. The anomaly approach is focused on normal behaviors patterns. When a new kind of activity becomes acceptable (does not challenge to security policy), the normal behavior pattern database must be efficient otherwise the activity will be treated as an intrusion and will result in false positives. A general problem of all anomaly detection methods, with the exception of the specification based technique, is that the subject's normal behavior is modelled on the basis of the (audit) data collected over a period of normal operation. In addition, because a subject's normal behaviour usually changes over time the Intrusion Detect System use the above method usually allow the subject's profile to slowly change. So, this gives an intruder the chance to slowly train the IDS and trick it into accepting intrusive activities as normal. Also, because these methods are all based on brief information, they are indifferent to quiet attacks. Because of some technical reasons, the current anomaly detection methods usually suffer from a high false alarm rate. Another difficult problem in building such representations is how to decide the features to be used as the input of the models (e.g., the statistical models). So, it is not sure that all the features related to intrusion detection will be selected as input limitations. Missing important intrusion related features makes it difficult to distinguish attacks from normal activities.

8.2.2 Misuse Detection or Signature Matching

Misuse detection is said to be corresponding to anomaly detection. Its main advantage is simplicity of adding known attacks to the model. Therefore, this systems look for well-defined pattern of known attacks or vulnerabilities. They can catch an intrusive activity even if it is so negligible that the anomaly detection methods be likely to to ignore it. Attacks and deviations from normal behaviour are taken as anomalies. The disadvantage of misuse detection is that it cannot detect novel or unknown attacks. As a result, the computer systems protected only by misuse detection systems face the risk of being comprised without detecting the attacks. When the requirement of clear representation of attacks, the detection system requires the environment of the attacks to be understood. It implies that human experts must work on the analysis and representation of attacks. So, it is time consuming and error disposed to. Additionally, Intrusion detection systems (IDSs) are categorized according to the kind of input information they analyze. This leads to the distinction between host-based and network-based IDSs. Host-based IDSs analyze host-bound inventory sources such as operating system inventory trails, system logs, or application logs. Network-based IDSs analyze network packets that are captured on a network.

9. LIMITATIONS OF INTRUSION DETECTION SYSTEMS (IDS)

This component in security systems as they allow network administrators to detect policy damages and these policy damages range from external attackers trying to gain unauthorized access to insiders abusing their access. But these IDSs also have some drawbacks which are described as under as :

- Current IDS are usually altered to detect known service level network attacks. This leaves them vulnerable to original and unknown or novel malicious attacks.
- Data overload: Another phase which does not relate directly to misuse detection but is extremely important is how much data an analyst can efficiently analyze. Depending on the Intrusion Detection tools employed by a company and its size there is the probability for logs to reach millions of records per day.
- False positives: A false positive occurs when normal attack is incorrectly classified as malicious and treated hence.
- False negatives: In this case, an IDS does not generate an alert when an intrusion is actually taking place. (Classification of malicious traffic as common) Data mining can help improve intrusion detection by addressing each and every one of the above mentioned problems [6].

Table-9.1Comparison of Intrusion Detection Techniques

S. No	Authors	Advantge	Techniqes	Methods	Limitations
1	S.V.Shirbhate	Discovered unknown attack	Hybrid anomaly detection technique [5]	K-means clustering.	Require more cluster
2	A.M. Chandrasekhar	High Accuracy	Support Vector Machine (SVM) [19]	Classification	The training and testing speed is show
3	Yang Yong	It solves the problems with multiple solutions	Genetic Algorithms[18]	REGAL System	No constant optir mizati On response Time.
4	Subaira.A.S	Highly tolerate the noisy data.	Neural Networks[4]	MLP(Multilayer perceptron)	It Requires long Training time.
5	G. J. Klir	5.Rule base or fuzzy sets easily modified	Fuzzy Logic[17]	Classification	Hard to develop a Model from a fuzzy System.
6	Niken Prasasti	Can handle high dimensional Data.	Decision Tree[19]	Classification	Limited to one output Attribute.

CONCLUSTION

This paper discussed functions, securities, techniques and challenges Intrusion Detection System are discussed. Intrusion Detection System current industrial Intrusion Detection Systems make use of Misuse Detection. As such, they completely are short of the ability to detect new attacks. It is impossible to prevent security violation completely by using the exciting security technology. Accordingly, Intrusion. Afterwards we introduced the data-mining and machine learning algorithms and it advantages of the IDS based on data mining and machine learning approach such as clustering and classification.

REFERENCES

[1] Zhongmin Cai, Xiaohong Guan, Ping Shao,Qingke Peng and Guoji Sun,” A rough set theory based method for anomaly intrusion detection in computer network systems”IEEE 2003.

[2] LI Yunl , LIU Xue-cheng and ZHU Feng,” Application of Data Mining in Intrusion Detection”IEEE 2010.

[3] M. Moorthy, Dr. S. Sathiyabama” A Study of Intrusion Detection using Data Mining” IEEE2012.

[4] Sunil Kumar Khatri,”Intrusion Detection Using Data Mining “,2012.

[5] Kapil Wankhade, Sadiya Patka”, An Overview of Intrusion Detection System based on Data Mining Techinque “ IEEE 2013.

[6] Harshna, NavneetKaur” Survey paper on Data Mining techniques of Intrusion Detection”IEEE2013.

[7] Chakchai So–In, Member, Nutakarn Mongkonchai, Phet Aimtongkham,” An Evaluation of Data Mining Classification Models for Network Intrusion Detection” 2014 IEEE.

[8] Kalpana Jaswal, Seema Rawat,Praveen Kumar “Design and Development of a prototype Application for Intrusion Detection using Data mining” 2015 IEEE.

[9] Jabez Ja,.B.Muthukumarb,”Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection”.

[10] Wenke Lee Salvatore J. StolfoKui W. Mok, A Data Mining Framework for Building Intrusion Detection Model.

[11] Urvashi Modi1 and Anurag Jain “An Improved method to detect intrusion Using machine learning algorithm”2016.

[12] Srikanth Yadav “A Study on Intrusion Detection using Mining Techniques”2016.

[13] Wenke Lee Salvatore J. StolfoKui W. Mok, A Data Mining Framework for Building Intrusion Detection Model.

[14] Zhongmin Cai,1Xiaohong Guan,1,2 Ping Shao,1 QingkePeng1 and GuojiSun , A rough set theory based method for anomaly intrusion detection in computer network systems.

[15] PeymanKabiri and Ali A.Ghorbani,"Research on Intrusion Detection and Response A Survey".

[16] Chakchai So–In,"An Evaluation of Data Mining Classification Models for Network Intrusion Detection".

[17] Wang Pu,"Intrusion Detection System with the Data Minining technology".

[18] Ashish Dutt, Maizatul Akmar Ismail, and Tutut Herawan A Systematic Review on Educational Data Mining .

[19] S. SobinSoniya**,** Intrusion Detection System Classification and Technique.