ISSN No. 0976-5697

Volume 8, No. 1, Jan-Feb 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Data Mining Approaches on Network Data: Intrusion Detection System

Aasia Abdullah
Student (M.Tech)
Maulana Azad National Urdu University
Hyderabad, Telangana, India

Khaleda Afroaz
Assistant Professor, Dept. of CS&IT
Maulana Azad National Urdu University
Hyderabad, Telangana, India

Abstract: In the modern era having tremendous growth and the usage of networking over computer systems and its applications mainly focuses to make our system as secure as possible. So with the help of intrusion detection system we can implement security on the system that protects our system data to access from unauthorized persons. Intrusion Detection System (IDS) takes responsibility to monitor the networks or host packets to find the malicious activities which occurs in the system. This paper consists data mining techniques to implement on IDS to identify both of the attacks that is known and unknown attacking patterns, so IDS helps the user to secure the information system. A Network Intrusion Detection System (NIDS) must be installed into the system to work as software application for detecting and monitoring the network activities and also protects from malicious and illegal access of devices. Data mining provides a way to analyze, classify, clean and eliminate the large amount of network data. Therefore, to come out from huge volume of dataset we use different data mining techniques like as classification, clustering along with association rules to analyze the network traffic and make the information as confidentiality and integrity.

Keywords: Intrusion Detection System, Classification, Clustering, Data Mining, Misuse Detection, Anomaly Detection, False Alarm Rate, Network Security

I. INTRODUCTION

In the new era of digital network, most of the users are facing electronic attacks where Intrusion Detection System (IDS) provides an effective guidelines that gives an alternative solution to solve out the problems which are being effected. These days many researchers are going in this field to eliminate the problem. An intrusion maintains a bunch of actions to secure our system information on implementation of various conditions like integrity, availability, and confidentiality [13] into the system like as file systems [16], user accounts, system kernels etc. [1, 2]. Intrusion defines the performing wrong activity into the system and entering into another machine [6] without any permission [10]. Intrusion system gathers the data/information and after that analyze it for misbehavior or abnormal events [16]. Intrusion detection facilitates the techniques for monitor and analyze the different types of network activities which is generated in the computer system [6, 17]. To find the signals of security issues in the past few years intrusion detections system and other security technology like as firewalls, cryptography and authentication are performing important role in digital data. Intrusion detection performs as the data analysis process to work on information.

Data mining provides a process to search the matching pattern from huge set of already stored data [16]. These days the main purpose to implement the data mining into intrusion detection system is to eliminate the large volume of data that may be newly or existing [6] stored. There are so many limitations in traditional IDS, so data mining is responsible to increase detection rate, maintain false alarm rate and decrease false dismissals [1].

II. BACKGRUOUND

Data mining defines as a Knowledge Discovery from Database (KDD) [2] to fetch the knowledge through the online. The data that have been founded from existing data set that will help us to increase the data profit. Data mining gives a convenient way to learn new patterns automatically having huge amount of data availability [2]. Data mining techniques are used to apply for detecting the hidden knowledge with intrusion prevention mechanism. It is beneficial to find intrusion as well as vulnerabilities [13] and also detect unknown pattern which occurs previously [3].

Security in network is very important issue to manipulate and store sensitive information through outlier [3]. Generally an intrusion describes as a group of events. IDS protects the network system from various attacks, monitor all the network activities automatically [7] and detect the malicious attacks. Data mining performs on three transactions to obtain the data i.e. Extract, Transform and at last Load [ETL] the data. It stores the data in multidimensional system for manage and analyze data to represent in specific format such as graph or table.

There are various different data mining techniques such as classification, clustering and association rule that helps to obtain useful and frequently used information which is acquired by intrusion after analyzing and monitor the network data [18]. Several different classifiers are also applying to generate a hybrid learning approach. That is a group of classification and clustering technique to obtain huge detection rate as well as low false alarm rate [3].

III. INTRUSION DETECTION SYSTEM (IDS)

Intrusion detection system defines as a process for controlling the events which occurs in a network and analyze it for signal through various incidents [1]. The

concept of IDS was initially come in 1980 by James P. Anderson [5, 19]. So, due to huge growth of the internet and more availability of tools which are attacking on network, that intrusion detection makes a critical part for the network administrator. The main use of IDS is to find the security interference in information system. To control information system and raise the alarm [7] intrusion detection performs as passive approach for security when security interference are detected. Security violations contains the abuse of privileges or attacks for software exploits or protocol vulnerabilities. So, the tool is required to automatically find the intrusion in the digital network. Therefore an IDS is a software which automatically finds the intrusions which raised in the system [1].

3.1. Intrusion detection methodology:-

Intrusion detection system uses so many techniques to list out the incidents [5] .When event occurs it may have many reasons like as, trojan, worms , attack, spyware, unauthorized access of computer system that misuses their privileges and try to attempt damage.

Initially intrusion detection system techniques have only two classification ([1, 2].

a. Misuse detection

b. Anomaly detection

a. Misuse intrusion detection system: -

This is defined as some specific set of rules and guidelines [2] to apply on the system network. Misuse detection searches all the patterns and also ties of known attacks [7] those are stored as signature. These signatures are generated by human experts which is based on their huge knowledge of intrusion techniques. In this technique if pattern is exact matched then alarm is raised for the signal of the event. So that security analyst able to evaluate the alarm to decide which action should be taken for example shutdown the system, close the internet connection to stop traffic. Misuse detection contains trained dataset having learning algorithms which create difficulties to collect it [9]. So overall only known attacks can be found which leaves their characteristics traces. So this is the pit false of misuse detection. Therefore we have required to change and update the signature whenever new version of software comes or we can also change software configuration because system are dynamic. The main advantages of misuse detection technique is having more degree of the accuracy to detect known attacks and its variants [1].

b. Anomaly intrusion detection system: -

Anomaly intrusion detection system provides a way to find unknown attacks as well as known attacks which occurred in the system [2]. It actually stores the features of the user's that usually occurred, and then compare and check with the current behavior of the events which is available in the database. If variations is huge then more chances of some abnormal events [15]. The main weakness of anomaly intrusion detection system is having more chances of false positives [1] or high false alarm rate [2] in the system.

Apart from the Misuse detection and Anomaly detection, intrusion detection system can also implements various detection mechanism to find the events which are performing on networking or in the systems.

3.2. Intrusion Detection Types:-

- **Host based intrusion detection: -** Host based intrusion detection systems calculate the modification in the system. Such as log files [6], operating system audit, application logs [1]. Here attackers make target to particular system and try to get access on confidential data and audit them which are restricted from others [2].

- **Network based intrusion detection: -**Network based intrusion detection system observe network packets those are detected on a network [1]. This activity may be done by sending large set of network traffics by using the advantages of others like as overloading of network traffic, unknown fault detection etc. [2].

- **Wireless intrusion detection: -**wireless intrusion detection system controls traffic of wireless network and analyze its networking protocol to find any illegal activity presents in the protocol themselves [1].

- **Network behavior analysis: -**Network behavior analysis calculate network traffic and identify threads which generate different traffic flow, like as DDOS, malware , worms, backdoors, policy violations and etc..

3.3. Components of intrusion detection system:-

With the whole above discussion, intrusion detection system controls, analyses, and monitors the digital data. Therefore, so many components that are used in IDS.

a. Sensor or agent: - Sensors or agents analyze the activity and monitor it. Actually sensor is used for monitor the network activities through intrusion detection system. The term agent is mainly used in host based IDS.

b. Management server:-This is a centralized server that take the information by the sensors or agent and it also responsible to manage them. Management servers also responsible to perform analysis on event information and identify the events. The information which matches from multiple sensors like as detecting events which is triggered by someone IP address that is called as correlation.

c. Database server: - A database server maintains the repository which records the event of information through sensor agents or management servers. It maintains the database server.

d. Console: - Console is a program that facilitate an interface for the user and administration of an intrusion detection system to control all the activity of the system progresses [11]. Actually console software is installed in computer system. Consoles are mainly used for configure sensors or agent, software update where as other console are being used for monitoring and analysis the intrusion.

IV. DATA MINING TECHNIQUES ON IDS

Some data mining techniques are used to detect the pattern with the help of following algorithms. In this paper we have described three main data mining techniques to implement on intrusion detection system for detecting the attacks [20].

4.1. Classification: -

Classification is a function of data mining that separates the data in the set of class or model. It is a data mining techniques that contains an instance of the data set and put it into a selected class [6]. The classes which are defined that extract the models, these model are said to be as classifiers [3]. Classification is mainly used to predict the objects class.

The IDS which is based on classifier divide whole the network traffic into normal/intrusion. Data classification mainly contains two steps. First one is learning and second one is classification, therefore learning construct the classifier whereas classification predict the class level from existing data for the model that is being used [3].

Classification is known as supervised [14] machine learning technique that manages only labeled data. So the major drawback of this technique is, it cannot support to work on unlabeled data. So here intrusion detection system performances will be degraded. That's why it has low efficient in IDS in respect to clustering. We can proposed ensemble boosted decision tree technique in the IDS. This learning techniques permits to group various decision trees to construct a classifier [3].

The Back-propagation of the artificial neural network determines the system behaviors. So with the help of this machine learning techniques we enable the IDS design and also find newly changeable environment. Therefore, this technique facilitates to list out the unknown attacks.
Support vector machine (SVM) defines as a classification methods that works well in the forms of text classification. The SVM separates the data [17] with help of two class problems which have N number of the support vectors. Hence, support vectors are treated as subset of training data which separates the two classes with boundary lines.

4.2. Clustering: -

Clustering is the process that performs labeling and assigning the data into an object that have similar objects in each groups. So each of the groups are defined as a cluster [6]. Clustering technique defined as to classify the data related to their some behavior or hidden patterns. It differentiates similar and dissimilar objects in separate clusters. So different clustering mechanism mainly partition-based is always used to partition the data. Therefore, it rearranges the data from initial group of data to specific set of data. For example K-means cluster includes in data mining to partition a group of points into K set or into cluster to indicate each cluster which are nearer to others [12].

Clustering technique gives idea to detecting the intrusion over the network data. Clustering defined as an unsupervised machine learning technique. It detects the patterns from unlabeled data on the basis of various dimensions. Clustering mechanism helps us to work on unlabeled data [3] and also provides a way to discover complex intrusion over different specific time period [6]. Clustering grouped the patterns based on their similarities. Clustering technique is implemented in both anomaly as well as misuse detection.

4.3. Association rule: -

The association rule gives a tremendous way to analyze and predict the behavior of customers in data mining. An association rule generally works on two parts that is *if* and *then*. The association rule is a pair item of attribute/value that item is collected from network request and find the item from huge number of dataset frequently that is appear in the network. The main motto of association rule is to facilitate multi-features correlation into a database table [6]. This association rule joins the associations among various data attribute according to frequency. Like as it first checks if X is available in the event, then what is the percentage of Y in the event [14]. As for example by implementing association rule identify the interesting relationship over various item set from existing data set. This approach searches strong relationship between huge amounts of dataset which have been attacked on system. Association rule is an effortless technique to retrieve the interesting rules from the given initially data set item [19].

V. WORK FLOWS OF IDS

There are following four stages to work IDS [4, 6]:-

1) **Data collection: -** It gathers the data through various sources by implementing specific software.
2) **Feature selection: -** Collect large set of data through network traffic. That's why dataset is very large for the IDS. So, it produces feature vectors for working on huge set of data that keeps only useful data.
3) **Analysis: -** Here, the data which have been gathered, we analyze to find weather the collected data is correct or not.
4) **Action: -** When attacks are detected then IDS quickly alert the administrator [6].

VI. MEASURING THE PERFORMANCE OF IDS

There are so many primary factors which are implemented to measure the performance of IDS [4].

1) **True positive (TP):-** At the time of IDS all numbers of normal data which have been found.
2) **True negative (TN**):- The total number of abnormal data which are detected in IDS.
3) **False positive (FP):-** It is also said as false alarm, the total set of normal data which are detected but that should be actual attack.
4) **False negative (FN):-** Total number of abnormal detected instance but that should be normal data.

Hence, to calculate the performance of the IDS in the form of detection rate, false alarm and accuracy.
So, Detection Rate (DR)

= (TP/TP+FN)*100% false alarm rate (FAR)
= (FP/number of attacks accuracy)
= (TP+TN/TP+TN+FP+FN)*100% [4]

VII. CONCLUSION

In this paper we have reviewed last one decade issues related to security where user and administrator are facing security problem over networking and system. Here various data mining techniques are implemented for intrusion detection system (IDS) to prevent from unauthorized access of illegal user. Network intrusion detection (IDS) system facilitates the mechanism to control the events which occurs in a computing system or networks by analyzing them with

the help of signs of intrusion. In this paper we are discussing on IDS and its various area of intrusion detection where data mining technology are being used. In our work we are integrating the data mining concept with IDS to detect the relevant and hidden data according to user's interest. Data mining technology are mainly used to find useful and consistent patterns of system that shows the behavior of the users. It also increase the different detection rate, monitor false alarm rate, and minimize false dismissals.

VIII. REFERENCES

[1] Shirbhate, S. V., Thakare, V. M., Sherekar, S. S., & Amravati, A. A. (2011). Data mining approaches for network intrusion detection system. International Journal of Computer Technology and Electronics Engineering, 41-44.

[2] Elekar, K. S. (2015, September). Combination of data mining techniques for intrusion detection system. In Computer, Communication and Control (IC4), 2015 International Conference on (pp. 1-5). IEEE.

[3] Wankhade, K., Patka, S., & Thool, R. (2013, April). An overview of intrusion detection based on data mining techniques. In Communication Systems and Network Technologies (CSNT), 2013 International Conference on (pp. 626-629). IEEE.

[4] Abhaya, K. K., Jha, R., & Afroz, S. (2014). Data Mining Techniques for Intrusion Detection: A Review. International Journal of Advanced Research in Computer and Communication Engineering, 3(6).

[5] Nadiammai, G. V., & Hemalatha, M. (2014). Effective approach toward Intrusion Detection System using data mining techniques. Egyptian Informatics Journal, 15(1), 37-50.

[6] Denatious, D. K., & John, A. (2012, January). Survey on data mining techniques to enhance intrusion detection. In Computer Communication and Informatics (ICCCI), 2012 International Conference on (pp. 1-5). IEEE.

[7] Gudadhe, M., Prasad, P., & Wankhade, L. K. (2010, September). A new data mining based network intrusion detection model. In Computer and Communication Technology (ICCCT), 2010 International Conference on (pp. 731-735). IEEE.

[8] Baseman, E., Blanchard, S., Li, Z., & Fu, S. Relational Synthesis of Text and Numeric Data for Anomaly Detection on Computing System Logs.

[9] Wenjun, L. (2010, July). An security model: Data mining and intrusion detection. In Industrial and Information Systems (IIS), 2010 2nd International Conference on (Vol. 2, pp. 448-450). IEEE.

[10] Gupta, D., Singhal, S., Malik, S., & Singh, A. (2016, May). Network intrusion detection system using various data mining techniques. In Research Advances in Integrated Navigation Systems (RAINS), International Conference on (pp. 1-6). IEEE.

[11] Zhao, Y. (2016, May). Network intrusion detection system model based on data mining. In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2016 17th IEEE/ACIS International Conference on (pp. 155-160). IEEE.

[12] Dutt, I., & Borah, D. S. Some Studies in Intrusion Detection using Data Mining Techniques. International Journal of Innovative Research in Science, Engineering and Technolog, July 2015.

[13] Duque, S., & bin Omar, M. N. (2015). Using data mining algorithms for developing a model for intrusion detection system (IDS). Procedia Computer Science, 61, 46-51.

[14] Goeschel, K. (2016, March). Reducing false positives in intrusion detection systems using data-mining techniques utilizing support vector machines, decision trees, and naive Bayes for off-line analysis. In SoutheastCon, 2016 (pp. 1-6). IEEE.

[15] Xue, M., & Zhu, C. (2009, April). Applied research on data mining algorithm in network intrusion detection. In Artificial Intelligence, 2009. JCAI'09. International Joint Conference on (pp. 275-277). IEEE.

[16] Johri, S. (2012). Novel method for intrusion detection using data mining. International Journal of Advanced Research in Computer Science and Software Engineering, 2.

[17] Reddy, E. K., Iaeng, M., Reddy, V. N., & Rajulu, P. G. (2011, July). A study of intrusion detection in data mining. In World Congress on Engineering (Vol. 3, pp. 6-8).

[18] Bhapkar, S. P., Dhamane, S. S., Kandekar, Y. S., & Lodha, K. S. A Survey on Log Mining: A Data Mining Approach for Intrusion Detection.

[19] Lakshmi, M. N. S., & Radhika, D. Y. (2015). A complete study on intrusion detection using data mining techniques. Volume IX, IJCEA Issue VI.

[20] Julisch, K. (2002). Data mining for intrusion detection. In Applications of data mining in computer security (pp. 33-62). Springer US.