International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

An Improved Intrusion Detection System Using Clustering Technique in Data Mining

Heena Batra & Gaurav Gautam

Department of Computer Science & Engineering, IITM, Murthal
Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat
heenabatra29@gmail.com

Department of Computer Science & Engineering, IITM, Murthal
Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat
gaurav.gautam200@gmail.com

Abstract— *An intrusion detection system is an application that provides protection from malicious activities or policy violations and generates various rules to defend computer security and this system is relevant for intrusion detection. With popularization of internet, internet attack belongings are also increasing, thus information safety has became a important issue all over the world, hence Nowadays, it is an critical need to detect, identify and hold up such attacks effectively. In this modern world intrusion occur in a fraction of seconds and Intruders expertly use the modified version of command and thereby erase their footprints in audit and log files. Most of the existing systems have security breaches that make them purely vulnerable and could not be solved. Successful Intrusion Detection Systems academically differentiate both intrusive and nonintrusive records. In this paper, we provide an efficient Intrusion Detection system using clustering technique in Data Mining.*

Keywords— Intrusion Detection, Data Mining, Clustering, k-means

I. INTRODUCTION

Over the last decade, our society has become technology dependent. People rely on computer networks to receive news, supply prices, email and online shopping. The integrity and availability of all these systems need to be protected against a number of threats. Amateur hackers, rival corporations, terrorists and even foreign governments have the motive and capability to carry out complicated attacks against computer systems. The main focus of Intrusion detection and prevention systems (IDPS) is to identify the possible incidents, logging information about them and in report attempts. In addition, organizations use IDPS for other purposes, like identifying problems with security policies, deterring individuals and documenting existing threats from infringing security policies [1].

The security of computer network systems cannot be assured if it merely depends on conventional peripheral protection mechanisms such as firewalls and various authentication methods. Firewalls cannot ward off all outside attacks on the network and are useless to defend against inside attacks. Authentication systems cannot prevent legitimate 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 [2].

In the world of communication, Most of our crucial data is stored in a computer remote and in the most cases we exchange it over a network hence security is a big concern. But it's not just our data transmitting over the network but different types of attacks that can harm our stored data. So Monitoring computer system, its logs (administration logs, security logs, system logs, network logs) and protecting our crucial data is necessary. An intrusion detection system is an application that provides protection from malicious activities or policy violations and generates various rules to defend computer security and this system is relevant for intrusion detection. Intrusion detection system can be designed and developed on any platform but for its better functionality, we are using data mining technique [3].

II. INTRUSION DETECTION SYSTEM

In 1980, James Anderson first introduced the concept of Intrusion Detection. Since then, Intrusion detection techniques are considered as the second gate for providing networks security behind firewalls. The purpose of Intrusion Detection Systems (IDS) is designed to detect attacks against computer systems over insecure networks by this way that detects attempts by legitimate users to abuse their privileges or to exploit security vulnerabilities for comprising the computers. In fact, Intrusion detection is a process of gathering intrusion related knowledge occurring during the system monitoring, and then analyzing collected data to draw a conclusion whether the

system is intrusive or nor according the user activity, system logs, etc. Having detecting the some possible intrusion behaviors, the IDS raise the alarm to the network administrator and do some protection processing [4].

In Information Security, intrusion detection is the act of detecting actions that attempt to compromise the integrity, confidentiality, or availability of a resource. Intrusion detection system performs many functions which are vital for the system. These are as follows:

1. Monitoring and analyzing both user and system activities.

2. Analyzing system configuration and vulnerabilities.

3. Assessing system and file integrity.

4. Ability to recognize patterns typical of attack.

5. Analysis of abnormal activity patterns.

6. Tracking user policy violations

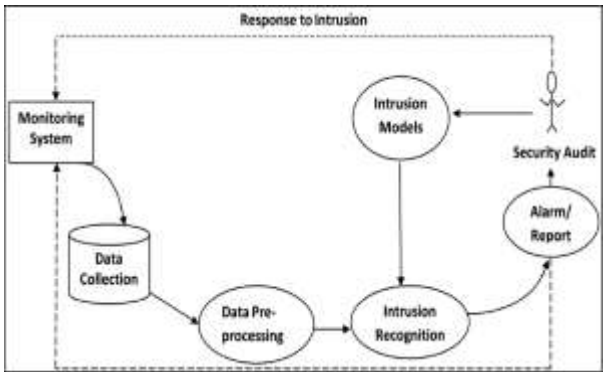

Figure 1: Overall structure of intrusion detection system

III. Overview of Work

Intrusions Detection Systems (IDS) has been developed into an important component in security infrastructures as they authorize networks administrators to identify strategy variations. These strategy violations range from outside attackers trying to increase unauthorized access to intruders abusing their access. Current IDS has a number of considerable drawbacks [11].

• **Threshold detection:** positive attributes of user and system behaviour are expressed in terms of count with some level established as permissible. Such behaviour attributes can include the number of files accessed by user in a given period of time the number of failed attempts to login to the system the amount of CPU utilized by a method. Use this method in Anomaly Based Intrusion Detection System generate a high level of false positives alarms.

• **False positives:** a false positive occurs when normal attack is incorrectly classified as malicious and treated therefore. The solution is to examine and review the IDS configuration to prevent false positive from occurring again.
• **False negatives:** A false negative occurs when an attack or an event is either not detected by the IDS or is considered kind by the analyst. Ordinarily the term false negative would only apply to the IDS not coverage an event.

• **Updates lag:** the main matter occurs to Signature-Based Intrusion Detection System is the update lag. In other words, will be always a lag between the appearances of new thread and the IDS's updates.

• **Data size:** the amount of data the analyst can efficiently analyze.

IV. Proposed Solution

In the world of communication, security is a big concern. Most of our crucial data is stored in a computer system and in most cases we exchange it over a network. But it's not just our data transmitting over the network but different types of attacks too. These attacks can harm our stored data. Monitoring computer system and its logs (administration logs, security logs, system logs, network logs) and protecting our crucial data is necessary. For these necessities we use intrusion detection system. An intrusion detection system is an application that provides protection from malicious activities or policy violations and generates various rules to defend computer security. Intrusion detection system can be designed and developed on any platform but for its better functionality we are using data mining technique [12].

Different data mining techniques such as clustering, classification and association rule finding are being used for intrusion detection. Data clustering is the method of grouping travel document into one or more categories based on their content. There are many techniques for data clustering; we are using k-mean clustering which is an unsupervised learning algorithm. This technique is relies on finding cluster centres by trying to minimize a cost function of dissimilarity measure.

Clustering is a process of labelling data and assigning that data into groups of similar objects. Each group is called as cluster. It consists of members from the same cluster that are similar and members from the different clusters that are different from each other. K-means [13] is one of the simplest unsupervised learning clustering algorithms. Its procedure follows an easy way to classify a given data set through a certain number of k clusters that are fixed a priori. First k center locations (c1,……,ck) are initialized. Then each data point xi is assigned to its nearest cluster centre ci. Each cluster centre ci is updated as the mean of all data point xi that has been assigned closest to it. The positions of the k centers are recalculated until the centers no longer move. The proposed method is based on K-means clustering, which is a typical clustering algorithm. It overcomes the drawbacks of K-means thereby employing a hybrid approach.

After applying filtering, initially the clusters are formed using K-means algorithm. The clusters that are formed by running the K-means algorithm are divided and merged again. By dividing and merging the clusters the number of k cluster centroids is calculated. The density of each point is calculated in filtered dataset to choose the appropriate initial centroids. These points are sorted as their density in descending order. Then the k points with the larger density are selected as the initial centroids. Again the clusters formation is done on the data set which is noise free using the calculated numbers of k cluster and the initial cluster centroids. Since the single clustering algorithm is difficult to get the great effective detection, the clustering ensemble is introduced by varying the value of k for the effective identification of attacks to achieve high accuracy and detection rate as well as low false alarm rate. The proposed method described aims to achieve high accuracy, high detection rate and very low or no false alarm rate.

V. IMPLEMENTATION

We have implemented our proposed work in MAATLAB 2010a. The implementation results for various parameters are performed as explained below.

Figure 3 below shows time complexity comparison between basic k-means and improved k-means clustering on Intrusion Detection Systems.

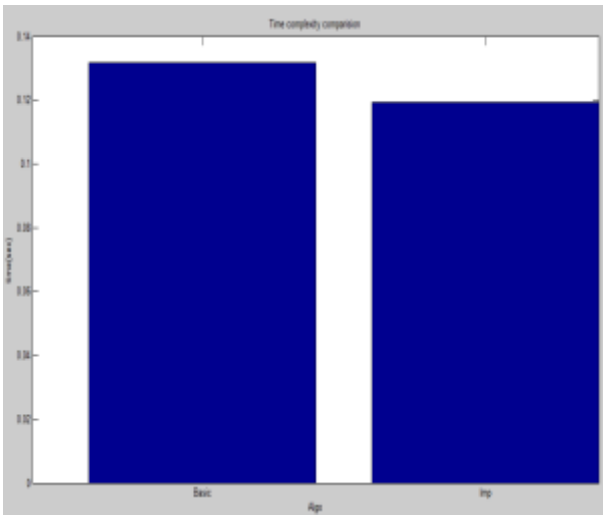

Figure 3: Time complexity comparison between basic k-means and improved k-means clustering on IDS.

Figure 4 below shows total distance comparison between basic k-means and improved k-means clustering on Intrusion Detection Systems.

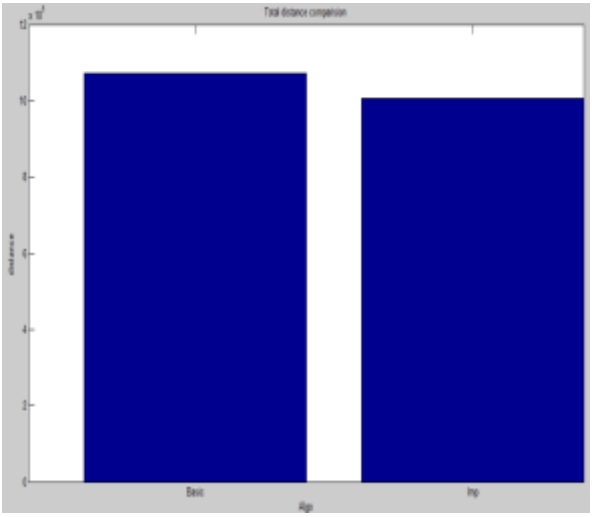

Figure 4: Total distance comparison between basic k-means and improved k-means clustering on IDS.

Figure 5 below shows precision comparison between basic k-means and improved k-means clustering on Intrusion Detection Systems.

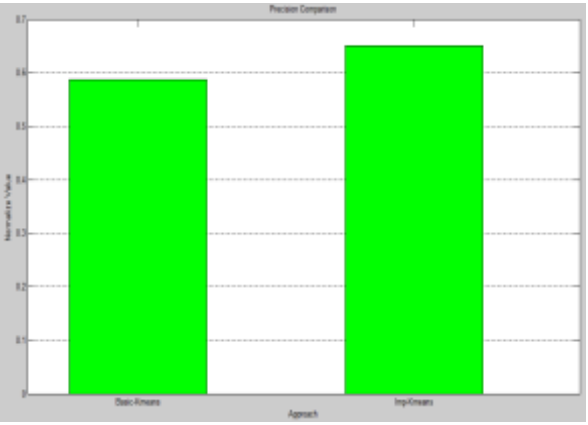

Figure 5: Precision comparison between basic k-means and improved k-means clustering on IDS.

VI. RESULTS

K-means clustering algorithm is one of the commonly used Partition based clustering algorithm. A common problem shared by current IDS is the high false positives and low detection rate. An unsupervised machine learning using k-means was used to propose a model for Intrusion Detection System (IDS) with higher efficiency rate and low false positives and false negatives.

The proposed method described aims to achieve high accuracy, high detection rate and very low or no false alarm rate.

VII. CONCLUSION

The security of computer network systems cannot be assured if it merely depends on conventional peripheral protection mechanisms such as firewalls and various authentication methods. Firewalls cannot ward off all outside attacks on the network and are useless to defend against inside attacks. Authentication systems cannot prevent legitimate 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. In this paper, we provided an efficient Intrusion Detection system using clustering technique in Data Mining.

REFERENCES

[1] Ahmed Patel, Mona Taghavi, Kaveh Bakhtiyari "An intrusion detection and prevention system in cloud computing: A systematic review "IEEE 2012.

[2] Jabez J, B. Muthukumar, "Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection approach", 2015.

[3] Su-Yun Wua, Ester Yen "Data mining-based intrusion detectors" Crown Copyright _ 2008 Published by Elsevier Ltd. All rights reserved Corresponding author"IEEE 2008.

[4] 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", 2003.

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

[6] S.V. Shirbhate, S. S. Sherkar, V. M. Thakare, "Performance Evaluation of PCA Filter In Clustered Based Intrusion DetectionSystem", 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologiessan, 2014.

[7] Hatim Mohammad Tahir, Abas Md Said, Nor Hayani Osman, Nur Haryani Zakaria, "Improving K-Means Clustering Using Discretization Technique In Network Intrusion Detection System", 2016 3rd International Conference On Computer And Information Sciences (ICCOINS), ©2016 IEEE

[8] Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P., Srivastava, J., Kumar, V."The MINDS – minnesota intrusion detection system. Next generation data mining. 2004.

[9] Bace, Rebecca G."NIST special publication on intrusion detection systems"2002.

[10] Anusha Jayasimhan and Jayant Gadge, "Anomaly detection using a clustering technique", International Journal of Applied Information Systems (IJAIS)–ISSN, pp. 2249–0868, 2012.

[11] Nadya EL Moussaid, Ahmed Toumanari Essi, "Overview of Intrusion Detection Using Data-Mining and the features selection"IEEE 2015.

[12] Ketan Sanjay Desale, Chandrakant Namdev Kumathekar, Arjun Pramod Chavan "Efficient Intrusion Detection System using Stream Data Mining Classification Technique", IEEE 2015.

[13] Z Muda, W Yassin, MN Sulaiman, and NI Udzir, "Intrusion detection based on k-means clustering and naıve bayes classification", in Information Technology in Asia (CITA 11), 2011 7th International Conference on. IEEE, 2011, pp. 1–6.