iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VIII **Month of publication:** August 2017

DOI: http://doi.org/10.22214/ijraset.2017.8322

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

Intrusion Detection Systems in Clustering: A Review

Parmod Kumar[1], Sunil Kumar[2]
[1,2] *Department of Computer Science Engineering Guru Jambheshwar University of Science & Technology, Hisar*

Abstract: With popularization of internet, internet attack belongings are also increasing, thus information safety has became a important issue all over the world, Nowadays, it is an critical need to detect, identify and hold up such attacks effectively. In this modern world, intrusion occurs in a fraction of seconds and Intruders expertly use the modified version of command and thereby erase their footprints in audit and log files. Successful IDS academically make different both intrusive and nonintrusive records. Most of the existing systems have security breaches that make them purely vulnerable and could not be solved. In this paper, we provide review of various Intrusion Detection Systems in the literature.
Keywords: Intrusion Detection, Data Mining, Clustering, k-means, improved k-means

I. INTRODUCTION

In the world of communication, Most of our essential data is stored in a computer remote and in the most cases, we exchange it over a network hence security is a big concern. It not just our data transmitting over the network but different types of attacks that can harm our stored data. Therefore, monitoring computer system, its logs (administration, security, system and 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. On any platform, intrusion detection system can be designed and developed but for its better functionality, we are using data mining technique [1]. IDS can also classified in terms of detection behaviour into misuse–based and anomaly–based detections. The first approach, based on signature matching while the second is to detect the anomaly behaviour from the network. Each has its own strength in order to cope with both known and unknown attacks in an efficient way with high detection precision and speed. In Information Security, intrusion detection is the act of detecting actions that attempt to compromise the integrity, confidentiality, or availability of a resource. In 1980, James Anderson first introduced the concept of Intrusion Detection. Since then, Intrusion detection techniques are considering as the second gate for providing networks security behind firewalls. The purpose of Intrusion Detection Systems (IDS) is to design and detect attacks against computer systems over insecure networks by the 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 analysing 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 behaviours, the IDS raise the alarm to the network administrator and do some protection processing [2]. In this paper, we provide review of various Intrusion Detection Systems in the literature

II. TYPES OF INTRUSION ATTACKS

With the increased use of computers and ease of access to internet, the ways to attack and deceive a system has also increased. The types of computer attacks detected by IDS are categorized into selfish behaviour attacks and malicious attacks [3].

A. Selfish Behaviours

Selfish node does not cooperate in any network functions and exploits the services of the network for its advantage, in order to save its own resources such as battery life. While such misbehaviour may not be launched with explicitly bad intentions, it can lead to serious disruptions in network communications such as high route discovery delays and dropped data packets. To render its selfish behaviour undetectable, selfish node can use multiple identities (e.g. Sybil attack) .

1) *Routing Messages Dropping:* Selfish node intentionally drops routing packets that are not destined for it, or forwards those packets but with a time-to live (TTL) of 0 to prevent the creation of routes. There by the selfish node can avoid forwarding many subsequent data packets. This attack may reduce the network performance and prevent end-to-end communications between nodes, if the dropping node is at a critical point in the network.

2) *Routing Metric Inflation:* Selfish node may make routes through itself appear longer than they actually are, by increasing hop counts so the source nodes are more probably to select other routes that seem to be shorter. Selfish node can also generate packets that advertise arbitrarily high distance to a given destination, or stops announcing updates that contain better routes that pass through it. In routing protocols that suppress duplicate packets, such as DSR and AODV , where each node forwards only the first received packet and deletes any later copies of the same packet. Selfish node can break this rule by waiting to receive several duplicates and then forwards the packet with the highest routing metric, decreasing the probability of being selected in the discovered route.

B. *Malicious Attack*

1) *Modification:* Modification is the most common attack; in which malicious node modifies the content fields of routing packets that transit through it. A malicious node could modify packets before rebroadcasting them, so that they include less attractive metrics, false addresses, and fake hop count in order to redirect network traffic. This attack can cause severe routing disruptions such as; conflicted and suboptimal routes, erroneous routing table, network partition and lose of connectivity.
2) *Fabrication :* This attack refers to the generation of faked routing messages, in order to disrupt network operation or to deplete other nodes' resources. Such attack is difficult to detect.
3) *Impersonation:* Also called spoofing attack, it usually constitutes the first step in the majority of attacks. The malicious node hides its real identity and takes legitimate node's identity, thus it can receive all the messages destined to this node and gain access to the network. This attack can also be used for creating loops in order to isolate a target node from the rest of the network.
4) *Black Hole:* This attack exploits the vulnerabilities of routing protocols and it is carried out in two steps. First, the malicious node attracts traffic through itself by advertising better routes to the requested destinations. Afterward, the malicious node drops all the data or control packets passing through it without any forwarding.
5) *Gray Hole:* This attack is a refined form of black hole attack, in which a malicious node drops only selected packets and forwards the others, depends on the source or the destination of Packets. Another kind of gray hole may behave maliciously for a given period by dropping all packets then switch to normal behaviour later. This attack defeats trust-based mechanisms and makes the detection of malicious node more difficult to achieve.
6) *Wormhole:* Also called tunnelling attack, it is one of the most sophisticated attacks in MANETs. In this attack, a malicious node captures packets from one location in a network and tunnels them through an out-of-band channel to another malicious node located several hops away, which replays them to its neigh boring nodes, . The tunnel between the malicious nodes is actually faster than links between legitimate nodes, so the tunnelled packets arrive sooner than packets through other routes. Therefore, the malicious nodes are more likely to be included in the route and take an advantage for future attack. Detection of wormhole attack is generally difficult, and requires the use of an unalterable and independent physical metric, such as time delay or geographical location.
7) *Rushing:* This attack can be carried out against on-demand routing protocols that use duplicate suppression in their operations. In order to reduce the route discovery overhead, each intermediate node processes only the first received route request packets and rejects any duplicate packets that arrive later. Rushing node exploits this mechanism by disseminating route request packets in order to be included in the discovered routes. Rushing attack can be performed in many ways: by transmitting at a higher wireless transmission power level, by ignoring delays at MAC or routing layers, by keeping other nodes' transmission queues full or by using a wormhole tunnel.
8) *Byzantine:* A malicious node or a group of malicious nodes create or modify control packets with false routing information in order to disrupt or degrade the routing operation. This attack is not easy to detect because it has not a specific form; the malicious node can create routing loops, drops or diverts packets to non-optimal routes .
9) *Location Disclosure:* Location disclosure is an attack that aims the privacy attribute of an ad hoc network. A malicious node can reveal important information such as location of nodes, or even the structure of the entire network, by the use of traffic analysis techniques, or with simpler probing and monitoring approaches. Colluding nodes may gather information regarding the identities of communication parties; analyze traffic to learn the network traffic pattern, track changes in the traffic pattern, and then plans further attack scenarios. The leakage of such information is devastating in security sensitive scenarios.
10) *Blackmail:* This attack occurs against routing protocols that employs malicious detection mechanisms like Watchdog and bathwater. Malicious node may exploit these mechanisms to blackmail legitimate nodes in order to incite other legitimate nodes to put those legitimate nodes in their blacklists.

III. LITERATURE REVIEW

Several works related to our work, which presents the intrusion detection system in computer system as follow:

A. *Sang-Hyun Oh et al.(2006)*

Described in their paper "In anomaly intrusion detection, how to model the normal behavior of activities performed by a user is an important issue. To extract the normal behavior as a profile, conventional data mining techniques are widely applied to a finite audit data set. However, these approaches can only model the static behavior of a user in the audit data set. This drawback can be overcome by viewing the continuous activities of a user as an audit data stream. This paper proposes a new clustering algorithm which continuously models a data stream. A set of features is used to represent the characteristics of an activity. For each feature, the clusters of feature values corresponding to activities observed so far in an audit data stream are identified by the proposed clustering algorithm for data streams. As a result, without maintaining any historical activity of a user physically, new activities of the user can be continuously reflected to the ongoing result of clustering"[4].

B. *Su-Yun Wua et al.(2009)*

Has described , "With popularization of internet, internet attack cases are increasing, and attack methods differs each day, thus information safety problem has became a significant issue all over the world. Nowadays, it is an urgent need to detect, identify and hold up such attacks effectively. The research intends to compare efficiency of machine learning methods in intrusion detection system, including classification tree and support vector machine, with the hope of providing reference for establishing intrusion detection system in future. Compared with other related works in data mining-based intrusion detectors, we proposed to calculate the mean value via sampling different ratios of normal data for each measurement, which lead us to reach a better accuracy rate for observation data in real world. We compared the accuracy, detection rate, false alarm rate for four attack types. Moreover, it shows better performance than KDD Winner, especially for U2R type and R2L type attacks" [5].

C. *Chunfu Jia et al.(2009)*

Has described , "There are two technologies in intrusion detection systems: misuse detection and anomaly detection. Both misuse detection and anomaly detection have advantages and disadvantages. At present, the intrusion detection system is developed by using these two technologies in conjunction with one another, but there is not an effective method to evaluate the intrusion detection systems collaborative detection's performance. It is necessary to analyse it by establishing a strictly mathematical assessment equation. Considering the information theory method to analysis this problem, the intrusion detection capability can be used to analysis and evaluation. By contrast two intrusion detection systems, it turns out, the system that based on misuse and anomaly collaborative detection has the better detection effects" [6].

D. *E. Kesavulu Reddy et al.(2011)*

Mention that "Network security technology has become crucial in protecting government and industry computing infrastructure. Modern intrusion detection applications facing complex problems. These applications has to be require reliable, extensible, easy to manage, and have low maintenance cost. In recent years, data mining-based intrusion detection systems (IDSs) have demonstrated high accuracy, good generalization to novel types of intrusion, and robust behavior in a changing environment. Still, significant challenges exist in the design and implementation of production quality IDSs. Instrumenting components such as of data transformations, model deployment, cooperative distributed detection and complex engineering endeavor [7].

E. *Krishna Kant Tiwari et al.(2012)*

In their paper, "comparison made between 23 related papers of using data mining techniques for intrusion detection. Our work provide an overview on data mining and soft computing techniques such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Multivariate Adaptive Regression Spline (MARS), etc. In this paper comparison shown between IDS data mining techniques used for intrusion detection. In those 23 related papers, 7 research papers use ANN and 4 ones use SVM, because of ANN and SVM are more reliable than other models and structures. In addition, 8 re-searches use the DARPA1998 tuples and 13 researches use the KDDCup1999, because the standard tuples are much more credible than others. There is no best intrusion detection model in present time"[8].

F. *Surasit Songma et al.(2012)*

Proposed "a two-phase classification method. Specifically, in the first phase, a set of patterns (data) are clustered by the k-means algorithm. In the second phase, outliers are constructed by a distance-based technique and a class label is assigned to each pattern. The Knowledge Discovery Databases (KDD) Cup 1999 data set, which has been utilized extensively for development of intrusion detection systems, is used in our experiment. The results show that the proposed method is effective in intrusion detection [9].

G. Li Hanguang et al.(2013)

Has described that "By analysing the technology of Intrusion Detection System and Data mining in this paper, the author uses Apriori algorithm which is the classic of association rules in Web-based Intrusion Detection System and applies the rule base generated by the Apriori algorithm to identify a variety of attacks, improves the overall performance of the detection system"[10].

H. Subaira. A. S et al.(2013)

Specify that "In spite of growing information system widely, security has remained one hard-hitting area for computers as well as networks. In information protection, Intrusion Detection System (IDS) is used to safeguard the data confidentiality, integrity and system availability from various types of attacks. Data mining is an efficient artifice that can be applied to intrusion detection to ascertain a new outline from the massive network data as well as it use to reduce the strain of the manual compilations of the normal and abnormal behavior patterns. This work reviews the present state of data mining techniques and compares various data mining techniques used to implement an intrusion detection system such as, Support Vector Machine, Genetic Algorithm, Neural network, Fuzzy Logic, Bayesian Classifier, K- Nearest Neighbor and decision tree Algorithms by highlighting the advantages and disadvantages of each of the techniques" [11].

I. Mohit Sharma et al.(2014)

Has described that "Low rate Denial of Service attack is an advanced form of DoS attack which has high cover up characteristics due to its behaviour like normal traffic. Due to this characteristics of LDoS, at present available tools and IDS have very low helpfulness in detection of LDoS attacks and hence these attacks are successful in uncomfortable legitimate users from access the network resources. Most of the researchers proposition to make changes in protocols, router, network arrangement, node functionalities and network topography in their proposed explanation which is highly unworkable. There are also certain types of attacks such as denial of service which eats up all the resources of a server to render its services useless for its legitimate clients"[12].

J. Chakchai et al.(2014)

Mention that Due to a rapid growth of Internet, the number of network attacks has risen leading to the essentials of network intrusion detection systems (IDS) to secure the network. With heterogeneous accesses and huge traffic volumes, several pattern identification techniques have been brought into the research community. Data Mining is one of the analyses which many IDSs have adopted as an attack recognition scheme. Thus, in this paper, the classification methodology including attribute and data selections was drawn based on the well–known classification schemes, i.e., Decision Tree, Ripper Rule, Neural Networks, Naïve Bayes, k–Nearest–Neighbour, and Support Vector Machine, for intrusion detection analysis using both KDD CUP dataset and recent HTTP BOTNET attacks. Performance of the evaluation was measured using recent Weka tools with a standard cross–validation and confusion matrix [13].

Clustering plays a very vital role in exploring data, creating predictions and to overcome the anomalies in the data.

K. Preeti Arora et al.(2015)

Has described , "Clusters that contain collateral, identical characteristics in a dataset are grouped using reiterative techniques. As the data in real world is growing day by d ay so very large datasets with little or no background knowledge can be identified into interesting patterns with clustering. So, in this paper the two most popular clustering algorithms K-Means and K-Medoids are evaluated on dataset transaction10k of KEEL [14]. The input to these algorithms are randomly distributed data points and based on their similarity clusters has been generated. The comparison results show that time taken in cluster head selection and space complexity of overlapping of cluster is much better in K- Medoids than K-Means. Also K-Medoids is better in terms of execution time, non sensitive to outliers and reduces noise as compared to K-Means as it minimizes the sum of dissimilarities of data objects".

L. Ketan Sanjay Desale et al.(2015)

Define data mining as "Data mining is the method of extract valid, formerly known & comprehensive datasets for the future decision making. As the enhanced knowledge by World Wide Web the streaming data come into picture with its challenges. The data which change with time & update its cost is known as streaming data. As the most of the data is streaming in nature, there are so many challenge need to face in the sense of security point of view. Intrusion Detection System (IDS) works in the idea of detecting the intruders to protect the personal system. The research in data stream mining & Intrusion detection system gain high desirability due to the meaning of system's safety measure. An Intrusion Detection System (IDS) is important knowledge to detect such intruders who are harmful to the system. Main goal of the IDS is to save from harm the system & network from the intruders. IDS keep track of behaviour of the activities; if they are malicious to the system then it'll be automatically detected by the IDS"[15].

M. Amrit Priyadarshi et al.(2015)

Explained in their paper "Due increased growth of Internet; number of network attacks has been increased. Which emphasis needs for intrusion detection systems (IDS) for securing network? In this process network traffic is analyzed and monitored for detecting security flaws. Many researchers operational on number of data mining technique for developing an Intrusion detection system. For detecting the intrusion, the network traffic can be confidential into normal and anomalous. In this evaluated five rule base classification algorithms namely Decision Table, JRip, OneR, PART, and ZeroR. The essential requirement of any IDS is accuracy. The other requirements are extensibility and adaptability"[16].

N. Shengyi pan et al.(2015)

Have described The creation of traditional intrusion detection systems (IDSs) that use manually created rules base upon expert information is knowledge-intensive and is not suitable in the context of this big data problem. A data mining method called common path mining is used to routinely and accurately learn patterns for scenarios from a fusion of synchrophasor dimension data, and power system audit logs. As a proof of concept, an IDS prototype was implementing and validate. The Intrusions Detection Systems prototype accurately classifies disturbances, normal control operation, and cyber-attacks for the distance security method for a two-line three-bus power transmission system. Intrusion detection systems (IDSs) identify activities that violate the protection policy of a computer system or network. IDS are a necessary complement to preventive security mechanisms such as firewalls since IDS detect attacks that exploit system design flaws or bugs and IDS provide forensic evidence to inform system administrator's reaction to cyber-attacks. The manual development process results in limited scalability and updates are slow and expensive"[17].

O. According to Anna Little et al.(2016)

"The problem of differentiating regular from anomalous traffic has been studied extensively. However, the classification of anomalous traffic to different types of attacks remains a difficult and widely unexplored area. In the age of big data analytics it is becoming paramount to automate the security process, such as reaction to attacks and mitigation based on their type. This paper investigates the use of spectral clustering techniques for classifying computer network attacks. Spectral clustering is a highly robust classifier for big data, and is found to accurately and efficiently classify the attack data using a minimal number of select features. Extensive investigation into feature selection and weighting is discussed. Our classification results can easily be adapted by an Intrusion Detection System (IDS) for real-time attack detection, and the classification information used to mitigate future attacks"[18].

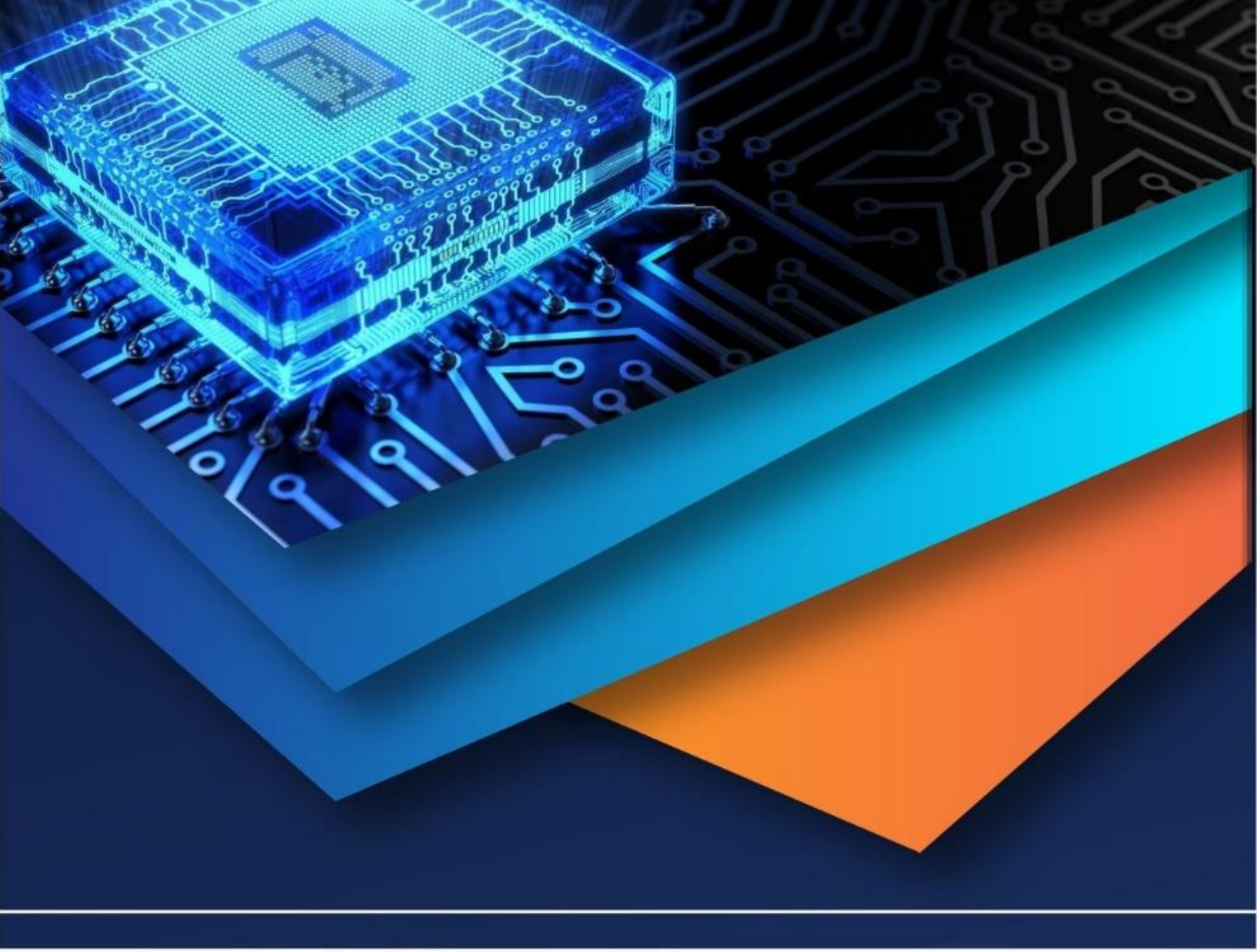
IV. CONCLUSION

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. 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 occurs in a fraction of seconds and Intruders expertly use the modified version of command and thereby erase their footprints in audit and log files. In this paper, we provided review of various Intrusion Detection Systems in the literature.

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] M. L. Lab, "Intrusion detection attacks database," 1999. [Online]. Available: http://www.ll.mit.edu/ ideval/docs/attackDB.html

[4] Sang-Hyun Oh, Jin-Suk Kang, Yung-Cheol Byun, Taikyeong T. Jeong, and Won-Suk Lee, "Anomaly Intrusion Detection Based on Clustering a Data Stream", © Springer-Verlag Berlin Heidelberg 2006
[5] Su-Yun Wua, Ester Yen, "Data mining-based intrusion detectors", 2008 Published by Elsevier Ltd
[6] Chunfu Jia, Deqiang Chen, "Performance Evaluation of a Collaborative Intrusion Detection System", 2009 Fifth International Conference on Natural Computation, © 2009 IEEE
[7] E. Kesavulu Reddy, V. Naveen Reddy, P. Govinda Rajulu, "A Study of Intrusion Detection in Data Mining, Proceedings of the World Congress on Engineering 2011 Vol III WCE 2011, July 6 - 8, 2011
[8] Krishna Kant Tiwari, Susheel Tiwari, Sriram Yadav, "Intrusion Detection Using Data Mining Techniques, International Journal of Advanced Computer Technology (IJACT)", 2012.
[9] Surasit Songma, Witcha Chimphlee, Kiattisak Maichalernnukul, Parinya Sanguansat, "Classification via k-Means Clustering and Distance-Based Outlier Detection", 2012 Tenth International Conference on ICT and Knowledge Engineering, ©2012 IEEE.
[10] Li Hanguang, Ni Yu, "Intrusion Detection Technology Research Based on Apriori Algorithm", International Conference on Applied Physics and Industrial Engineering, Published by Elsevier, 2013.
[11] Subaira. A. S, Anitha. P, "A Survey: Network Intrusion Detection System based on Data Mining Techniques", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 10, October 2013, pg.145 – 153
[12] Mohit Sharma, Nimish Unde, Ketan Borude, "A Data Mining Based Approach towards Detection of Low Rate DoS Attack", 2014.
[13] Chakchai So, Nutakarn Mongkonchai, Phet Aimtongkham, "An Evaluation of Data Mining Classification Models for Network Intrusion Detection", ©2014 IEEE.
[14] Preeti Arora, Deepali, Shipra Varshney, "Analysis of K-Means and K-Medoids Algorithm For Big Data", International Conference on Information Security & Privacy (ICISP2015), Published by Elsevier, 11-12 December 2015.
[15] Ketan Sanjay Desale, Chandrakant Namdev Kumathekar, Arjun Pramod Chavan, "Efficient Intrusion Detection System using Stream Data Mining Classification Technique", IEEE 2015.
[16] Amrit Priyadarshi M. M. Waghmare, "Use of rule base data mining algorithm for Intrusion Detection", 2015 IEEE.
[17] Shengyi Pan, Thaomas marris, "Developing a Hybrid Intrusion Detection System Using Data Mining for Power Syetem", IEEE 2015
[18] Anna Little, Xenia Mountrouidou, "Spectral Clustering Technique for Classifying Network Attacks", IEEE 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY