iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VII **Month of publication:** July 2017

DOI:

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, July 2017- Available at www.ijraset.com

Efficient Intrusion Detection System Using Clustering Technique in Data Mining

Salona Ranga[1], Suman[2]
[1,2] *Department of Computer Science Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat*

Abstract: ***The application of computer system that provides safety (protection) from nasty activities is called Intrusion Detection System. The Intrusion Detection System provides multiple rules so that protections of computer systems are maintained. In the age of Internet, many Internet related attacks compromise the security of computer system. Therefore, we must provide security from these types of attacks & intrusion detection system comes in aid for this. 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. In our existing computer systems there are maximum chances for security breaches. Successful Intrusion Detection Systems protect computer systems from various types of internal as well as external attacks. In this paper work, we provide an efficient Intrusion Detection System using clustering technique of Data Mining.***
Keywords: Intrusion Detection, IDS, Data Mining, Clustering, K-means

I. INTRODUCTION

Over the last few years, our society is fully dependent on technology. Their day-to-day activities revolve around computer and other technologies. For example, people in the companies communicate with each other through emails, we can list the prices of various products from online shopping sites; we frequently search Google for various topics etc. Mostly all the technologies uses the internet which is not safe. There are different types of experts in the fields of hackers that compromise the security and integrity of computer systems and information. Not only hackers but different terrorists and rival organization can attacks our data, information and computer systems & misuse them. The Intrusion Detection System helps people and organization to detect the attacks, hackers, their logging information and report these information to the owner of the computer system. The Intrusion Detection System not only identifies the attack on the computer system, it also determines problems with current security policies [1].

We generally applied conventional security mechanism for protection of out computer peripherals. The popular conventional security mechanisms are – authentication and firewall security. The authentication protects the computer integrity and security from unauthorized person but it cannot prevent authorized (legitimate) users from performing harmful operations on a computer system. On the other hand firewall only security from some internal attacks to the computer peripherals and information, it cannot provide complete security from outside attacks on the internet. The intrusion detection system is a powerful technology that provides security from both the inside as well as outside attacks [2]. In the world of communication, we exchange our data with another users using internet. Also in the age of cloud computing our data is stored on the remote computer which can be accessed using Internet. Therefore, security of data is big concern for different users. We need not only to protect the data, which exchanged through internet but also to protect the stored data from different types of attacks. Therefore, we must provide some mechanism to monitor our computer system from various attacks, protect our data exchanged on the internet and maintain reports of system and network logs for future reference. An Intrusion Detection System does all the above activities for us.

Successful Intrusion Detection Systems protect computer systems from various types of computer system attacks. It also provide security on the data we exchanged on the internet. We can construct Intrusion Detection Systems on various platforms. One such platform is data mining [3]. In this paper work, we provide an efficient Intrusion Detection System using clustering technique of Data Mining.

II. INTRUSION DETECTION SYSTEM

Intrusion Detection term was first introduced by James Anderson in 1980s. Now it is an important part network and firewall security. The primary goal of Intrusion Detection Systems (IDS) is to identify (detect) attacks from insecure networks such as internet to our computer system.

With Intrusion Detection Systems, we can obtain all the intrusion related information that occurs during the monitoring of system. We then analyse these information to determine whether our computer system is intrusive against any attack or security breach or

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, July 2017- Available at www.ijraset.com

not. When Intrusion Detection System detects something disturbing then it gives signal to the network administrator and performs some types of acts already defined to protect the system [4]. In the field of Information and data Security, intrusion detection systems determine the security breaches that compromise the three important security terms – confidentiality, integrity and availability. The intrusion detection system provides system logs regarding security breach, on the basis of these logs the administrator can determine the person that misuses the system and then he/she minimize their rights for future.
ntrusion detection system (ids) performs many functions which are vital for the system. These are as follows:

A. The Intrusion detection system can monitoring different system and users activities. On the basis of this monitoring it analyse the risk for current types of attacks
B. Intrusion detection system can analyse the system Vulnerabilities and it also analyse the configuration of computer system.
C. With intrusion detection system, we can access and maintain the integrity of computer file and system.
D. With intrusion detection system, we can identify pattern of different types of attacks. We can analyse activity patterns, which are not normal.
E. With intrusion detection system, we can determine the violation of user's policy.

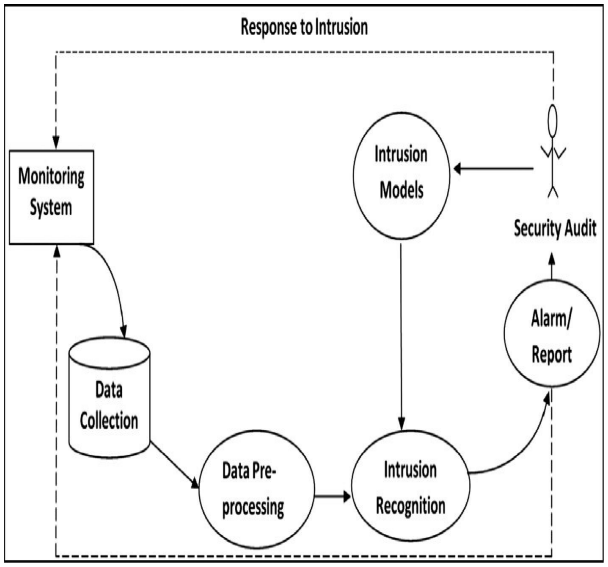


Figure 1: Overall structure of intrusion detection system

III. INTRUSION DETECTION SYSTEM (IDS) CLASSIFICATIONS

The Intrusion Detection Systems can be divided into following two categories:

A. According to Based on Data Sources

According to data sources used for IDS, the intrusion detection system can be categorized as host-based IDS, network-based IDS, Distributed IDS, Mobile agent based IDS, Cluster based IDS, Cryptography
based IDS, Neighhood watch IDS, Cross-feature analysis IDS, Collaborative IDS [5][6].

1) *Host-based Intrusion Detection Systems*: The data for host based Intrusion Detection Systems collected from host records of various activities such as operation system's audit record, classification logs, application programs information, etc. For example, the event logs mechanism of Windows NT operation system finds and gathers following three system events patterns – "Operation system event, safety event and application event".

The advantages and disadvantages of Host-based Intrusion Detection Systems in detailed.

Advantage:	Disadvantage:
Main advantages of this method are:	Main disadvantages of this method are:
1. "It can evaluator	1. "Higher cost":

whether or not the host is intruded more accurately": The data of Host-based Intrusion Detection Systems comes form system audit records and system logs of hosts. Therefore, it can determine various network attacks on hosts quickly and accurately.	The cost required for monitoring the host for intrusion detection is high. Because hosts are heterogeneous thereby for each individual host different intrusion detection system is required
2. "It can detect attack under encrypted network environment": The data from the system comes from files, which are encrypted therefore data is not affected by network attacks.	2. "It may affect system competence of monitored hosts": The Intrusion detection system occupies some resources of hosts while monitoring the host. Therefore, it affects the performance of host.

2) *Network-based Intrusion Detection Systems :* The intrusion detection of this type collect its data network stream with the help of different segments of data for example Internet packets. This method removes the burden from the hosts because it uses network traffic for checking the data source. Thereby hosts perform their normal operation of computing. It detect different types of network attacks such as signature based, anomaly based etc. Its main limitation is that there are large numbers of false alerts.[7]

Advantage:	Disadvantage:
Main advantages of this method are:	Main disadvantages of this method are:
1. "Low cost": It detects all types of network attacks & still require low cost of installing intrusion detection device.	1. "The flux is large, and some packet may be misplaced, and it cannot detect all packets in network".
2. It can detect (sense) attacks – DOS, DDOS etc, which are not detected by host based intrusion detection system.	2. "In important network, it requires more rapid CPU and additional memory space, to analyse bulk data".

3) *Distributed Intrusion Detection Systems:* The distributed intrusion detection system collects the data for auditing from various hosts. It also obtains the audit data from the network connecting different hists. It generally detects attacks, which involve multiple hosts.

4) *Mobile Agent Based Intrusion Detection Systems:* In this type of intrusion detection technique mobile nodes examine the different activities of nodes & gives report for intrusion. Based on intrusion report, the intrusion detection process starts. The limitation of this approach is that it involve large communication overhead. Its main advantage is that it reduces the energy

consumption rate of various sensor nodes because in this method mobile agents take the burden of data or information collection about intrusion.

5) *Cluster based Intrusion Detection Systems:* As the name suggests in this method we divide nodes into multiple clusters. Each cluster group has one cluster head & the cluster head monitors the nodes. The information collected by one cluster head is transferred to all the clusters through gateway. There are number of factors that help to construct the cluster head such as – average load, faithfulness etc.
6) *Cryptography based Intrusion Detection Systems:* Another method for intrusion detection is cryptography based Intrusion Detection Systems. It detect false route using route discovery technique. The network control traffic is need not to validate the route.
7) *Neighbourhood Watch Intrusion Detection Systems:* In this method of Intrusion Detection Systems, we check the number of nodes received from neighbours and number packets forwarded by it. If number mismatches then intrusion is detected and reports send to neighbour nodes.
8) *Collaborative Intrusion Detection Systems*: In this method of Intrusion Detection Systems, the decision for an intruder node is collaboratively taken. But in this method there is vast amount of communication overhead.

B. According to based on Different Analysis Methods

According to analysis method, intrusion detection system can be classified as "Misuse Detection and Anomaly Detection".

1) *Misuse Detection :* The misuse detection scheme of intruder detection is also termed as signature- based detection. It stores attack related information in the signature database. Now when an attack occurs then it is first compared with database of attacks signature. After confirmation from signature database the attack related data is termed as actual attack.[9]

In misuse detection approach, "it defines abnormal system behaviour at first, and then defines any other behaviour, as normal behaviour. It assumes that abnormal behaviour and activity has a simple to define model. It advances in the rapid of detection and low percentage of false alarm. It fails in discovering the non-pre-elected attacks in the feature library, so it cannot detect the abundant new attacks".

The framework of system model consists of following components

a) Data Collection Module
b) Pre processing Modules
c) Associative rule mining modules
d) Detection & analysis Modules

Advantages	Disadvantages
Main advantages of misuse detection are:	Main disadvantages of misuse detection are:
1. Intrusion Detection rate is high.	1. For unknown attacks, its intrusion detection rate is low.
2. For known attacks, False alarm rate is low	2. Database that stores attacks information must be frequently updated

2) *Anomaly Detection :* This method of intrusion detection predicts in advance the expected behavior of the network. If in future expected behavior is not reported then there must be some attacks on the network.

"The main advantage of this approach is that it can examine unknown and more complicated intrusions. The shortcoming of this approach is its low detection rate and high false alarm rate".

From above discussion, we can conclude that to detect unknown attacks we use anomaly detection method. While misuse detection is capable of detecting the known attacks[10].

IV. LITERATURE REVIEW

The Intrusion Detection System provides multiple rules so that protections of computer systems are maintained. Intrusions Detection Systems (IDS) has been developed into an important part in security infrastructures because they authorize networks administrators to identify possible variations. These deviation is due to outside attackers which increases unauthorized access for the intruders. There are multiple works in this method was performed earlier which are explained below.

A. Suman, Parvinder Singh, R.B.Patel,

"For seamless connectivity in Heterogeneous wireless networks (HWN), decision regarding vertical handoff is very crucial. In this paper we present an adaptive network selection algorithm UIVH (User Specific Intelligent Vertical Handoff). UIVH uses Sugeno fuzzy inference system (FIS) to decide when to perform handoff. ANFIS is used to rank different wireless networks for VHO based on set of parameters along with user preferences on a mobile device. UIVH fulfills specific needs of users and simultaneously balance overall load of HWN. Simulation results demonstrate that UIVH enhances quality of service (QoS) by reducing handoff latency".[11]

B. Suman, Parvinder Singh, R.B.Patel

"Selecting most optimal network to satisfy user requirements is very important for seamless mobility of users across heterogeneous wireless networks (HWNs). For overall network stability decision regarding when to perform vertical handoff and to which network is very crucial. In this paper we present an intelligent adaptive and user-centric network selection algorithm which uses Sugeno fuzzy inference system (FIS) to decide when to perform handoff. ANFIS is used to rank different wireless networks for VHO based on set of parameters along with user preferences on a mobile device. Our algorithm fulfills specific needs of users and simultaneously balance overall load of HWN. Simulation results demonstrate that our algorithm gives high network throughput and reduce packet drop rate and handoff latency."[12]

C. Sandeep Kumar, Dr. Suman Sangwan,

Wireless Sensor Networks (WSN) is a trending technology now-a-days and has a wide range of applications such as battlefield surveillance, traffic surveillance, forest fire detection, flood detection etc. But wireless sensor networks are susceptible to a variety of potential attacks which obstructs the normal operation of the network. Black hole attack is one of severe security threat that affects the network from its normal functioning by maliciously advertising itself having shortest route to the destination and then drops all receiving packets. There are lots of mechanisms have been proposed to defend network from black hole attack, but none of the solution looks most promising to defend against black hole attack. So in this paper, we have surveyed and compared the existing solutions to black hole attacks on AODV protocol. Tabular representation of comparison depicts clear picture of these solutions".[13]

D. Koumal Kaushik, Suman,

a video steganography method is proposed using hash based round Least Significant Bit technique. Video steganography is a more secure than any other steganography technique in hiding information because of its complex structure it disables the intruder to attack. In this paper secret text message is embedded in the video file using proposed hash based round Least Significant Bit technique. This work will improve the information security and embedding capacity. The proposed technique hash based round Least Significant Bit will be compared with the hash based Least Significant Bit. The technique proposed in this paper is analysed in term of Peak Signal to Noise Ratio, Mean Square Error and Embedding capacity".[14]

V. PROPOSED SOLUTION

In the world of communication, we exchange our data with another users using internet. Also in the age of cloud computing our data is stored on the remote computer which can be accessed using Internet. Therefore, security of data is big concern for different users. We need not only to protect the data, which exchanged through internet but also to protect the stored data from different types of attacks. Therefore, we must provide some mechanism to monitor our computer system from various attacks, protect our data exchanged on the internet and maintain reports of system and network logs for future reference. An Intrusion Detection System does all the above activities for us.

The Intrusion Detection Systems protect computer systems from various types of computer system attacks. It also provides security on the data we exchanged on the internet. We can construct Intrusion Detection Systems on various platforms. One such platform is data mining [15]. 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. The clustering process gives label to each data and then labelled data into groups of similar objects. These group of similar data form a cluster. The members of same group (cluster) are alike and members of different group are different from each other. One of the simpler and popular unsupervised clustering algorithm is K-means [16]

Figure 2 depicts "the system architecture for intrusion detection. It consists of feature selection, filtering, clustering, divide and merge, clustering ensemble and normal and intrusion detection. Feature selection is important if the data set consist of large number of attributes. It consists of selecting features using an information gain feature selection method which selects the important attributes from the data set. A filter method is proposed to reduce the noise and isolated points on the data set. It calculates the sum of the distance of each point from every other point and also calculates the sum of the average distance. For any point, if sum of the distance is greater than the average distance then that point is considered as an outlier and it is removed from the data set".

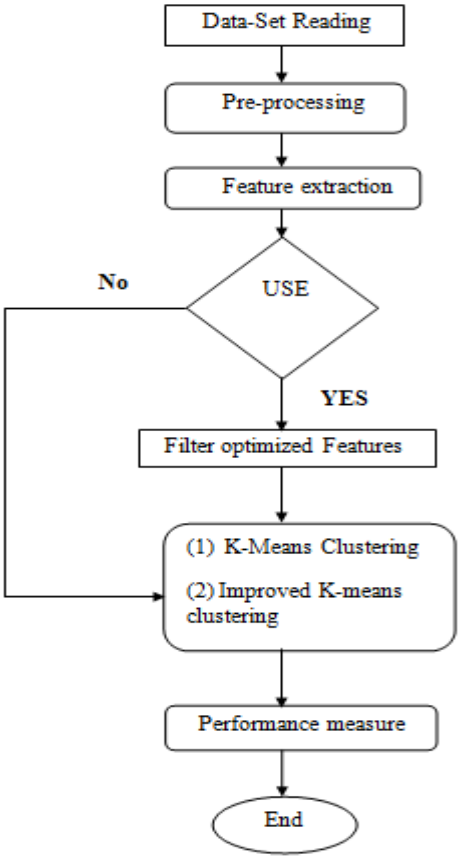


Figure 2: Structure of proposed work

From the figure we find that "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".

VI. 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.

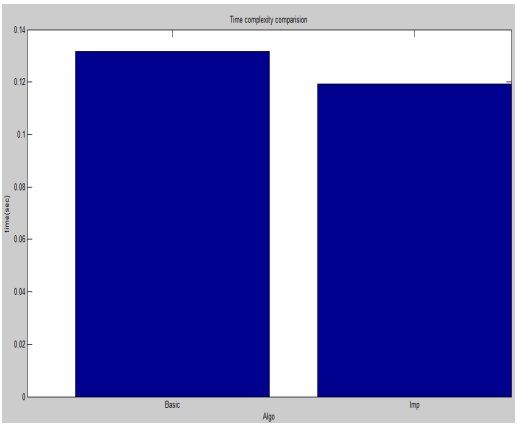


Figure 3: Time complexity comparison between basic k-means and improved k-means clustering on IDS.

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

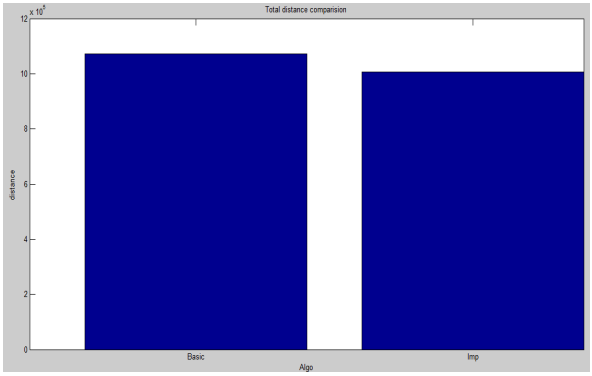


Figure 4: Total distance comparison between basic k-means and improved k-means clustering on IDS.

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

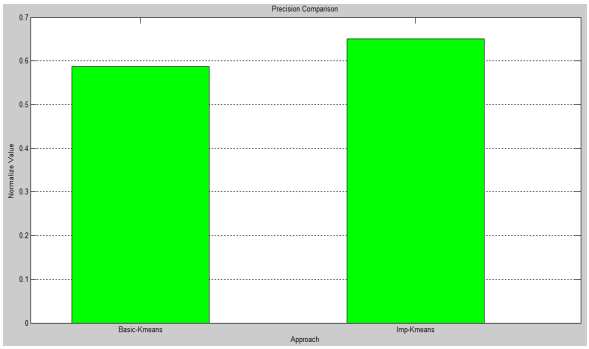


Figure 5: Precision comparison between basic k-means and improved k-means clustering on IDS.

Figure 6 below shows recall comparison between basic k-means and improved k-means clustering on Intrusion Detection Systems.

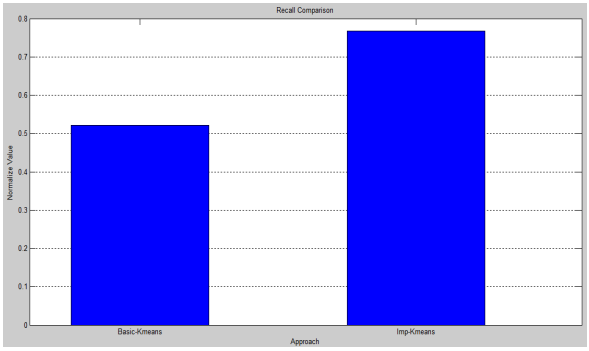


Figure 6: Recall comparison between basic k-means and improved k-means clustering on IDS.

Figure 7 below shows F-measure comparison between basic k-means and improved k-means clustering on Intrusion Detection Systems.

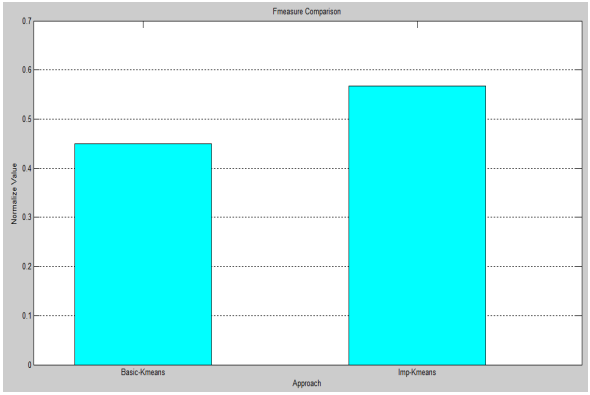


Figure 7: F-measure comparison between basic k-means and improved k-means clustering on IDS.

VII. RESULTS

One of the commonly used clustering algorithm which is based on partition method is K-means clustering algorithm. The existing Intrusion Detection Systems face two basic problems for intrusion detection- low detection rate and false positive is high. The unsupervised learning algorithm k-means proposes a new intrusion detection model that provides high rate of detection and low false positives values.

Our proposed algorithm produces result of intrusion detection with high accuracy, no or little false alarm rate and high detection rate. The analyses of result of our work with the work of previous existing methods are shown below.

- A Y-means clustering algorithm has better detection rate and low false alarm rate. But it cannot solve real time anomaly detection, since it cannot update the date set dynamically during the process.

- The major advantages of K-means are that it is a lightweight, fast iterative algorithm, which is easy to understand and implement. However, the major drawbacks are its sensitivity to initial conditions such as the number of partitions and the initial centroid, and it is sensitive to outliers and noise.

- A parallel clustering ensemble algorithm forms the clusters more speedily to mass data. It also achieves high detection rate but its false alarm rate is low.

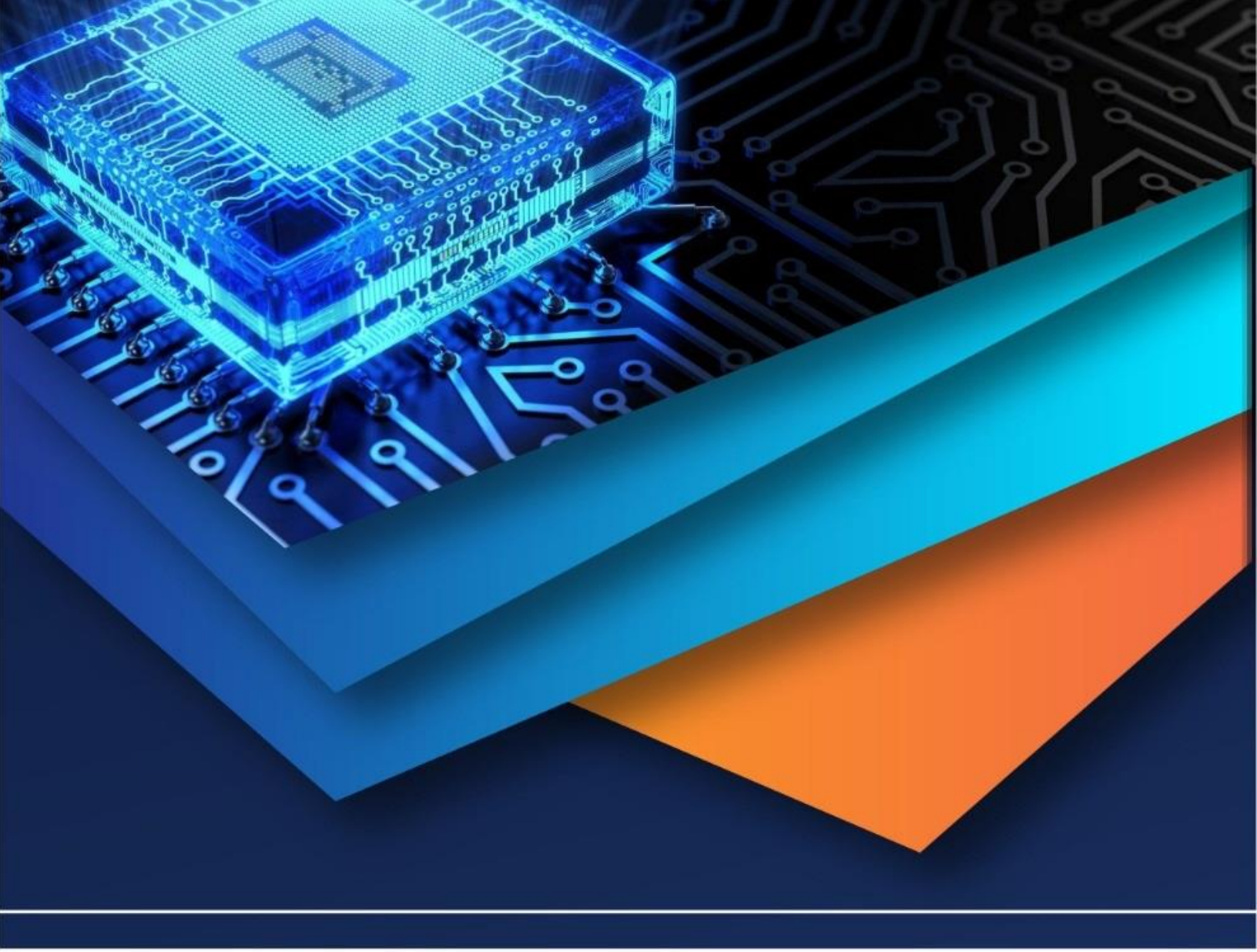
Our work of hybrid approach using improved classification technique and unsupervised learning algorithm k-means eliminate the shortcoming of existing methods such as low rate of detection and high rate of false alarm.

VIII. CONCLUSION

The Intrusion Detection System helps people and organization to detect the attacks, hackers, their logging information and report these i formation to the owner of the computer system. The Intrusion Detection System not only identifies the attack on the computer system, it also determines problems with current security policies. In the age of Internet, many Internet related attacks compromise the security of computer system. Therefore, we must provide security from these types of attacks & intrusion detection system comes in aid for this. We can construct Intrusion Detection Systems on various platforms. One such platform is data mining. In this paper work, we provide an efficient Intrusion Detection System using clustering technique of 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 Detection System", 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] Suman, Parvinder Singh, R.B.Patel, "Adaptive Vertical Handoff in Heterogeneous Wireless Networks", International Journal of Data & Network Security Volume1 No.3, Dec10, 2012
[12] Suman, Parvinder Singh, R.B.Patel, "User Specific Algorithm for Vertical Handoff in Heterogeneous Wireless Networks", International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013.
[13] Sandeep Kumar, Dr. Suman , "A Survey of Black Hole Detection Techniques in WSNs", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 5, May 2015
[14] Koumal Kaushik, Suman, "An Innovative Approach for Video Steganography", I. J. Computer Network and Information Security, 2015, 11, 72-79.
[15] Ketan Sanjay Desale, Chandrakant Namdev Kumathekar, Arjun Pramod Chavan "Efficient Intrusion Detection System using Stream Data Mining Classification Technique", IEEE 2015
[16] 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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY