Clustering Based Optimized Network Intrusion Detection System: A Survey

Devendra Baghel[1], Dr. Nirupma Tiwari[2]
[1]Dept of computer Science & Engineering
[2]Asst. Professor, Dept of computer Science & Engineering
[1, 2] ShriRam College of Engineering & Management,Banmore

Abstract- *Network intrusion is called the unwanted activity in the network and network intrusion detection system (NIDS) monitors network parameters to intrusion detect. Network defence is one of computer network management's most important challenges and intrusion presents the most popular security threats. Intrusion prevention has been a major area for network defence in recent years. When each attack class is used as a different problem and is administered by advanced algorithms, IDSs achieve better results. There are now different forms and techniques of intrusion prevention that can be seen in days. Many cluster-based detection systems use K-means basic technique to predict the type of attack. Clustering is a popular unsupervised technique of detection for anomalies that are commonly employed in the Intrusion Detection System (IDS). Optimization is now attracting major interest in the testing community, as it can fulfil the increasing need for reliable and intelligent IDS. While different optimization approaches have in recent years been proposed, some more popular optimization strategies are presented, including Ant colony method, Simulated Annealing & Tabu, Honey Bee Algorithm, Genetic Search and Search.*

Keywords- Clustering, Optimization Algorithms, (IDS) Intrusion Detection System, (NIDS) Network Intrusion Detection System.

I. INTRODUCTION

Nowadays the growth of the internet and use of computer systems has led to massive electronic data transformation with multiple problems like security, data security and confidentiality. Major strides have been made in improving the reliability of computing systems. However, mobile systems can have significant issues with stability, data protection and confidentiality. No device in the world is truly 100% secure at the moment. Moreover, there are still major attack scenarios. When a new signature is detected on a database of signatures, therefore the behaviour will be viewed as an attack [1].

Intrusion is a formal term that describes the compromise system act. And either failed intrusion detection is called intrusion detection, successful attempts to compromise the system. Easy. right? In brief, intrusion sensing systems or IDS detect intrusions that can be conceived by name. IDS applications have been developed to detect and unauthorized access to computer attacks and alert outraged people of identification and infringement of safety. The IDS can be viewed in a house as a burglar detector. They both display an intruder/attacker/snub, and both give a kind of alert and a warning. They both have separate mechanisms. [2]. The intrusion detection system (IDSs) is a monitoring device used on the protection wall to prevent malicious system intrusion. The main focus of this research is on the NIDS network because it can track the widest range of attacks about other forms of IDSs. Network IDS analyses traffic to detect persistent and incoming network attacks.

Network Related IDS observe packet which traverses thru the LAN segment or analyses network activity to detect attacks. When a network-based IDS is heard on a LAN segment network will monitor the traffic of many hosts connected to network area, so that these hosts can be protected. Network-based IDS consist mostly of hosts or sequences of one-purpose sensors placed in several LAN locations. Many of these sensors have been designed such that the presence or position of attackers is difficult in stealth mode. It is most often achieved on a network boundary, like those in virtual private servers, wireless networks and remote servers. [3].

The clustering of patterns (data items, observations, vectors) in groups (clusters) is unsupervised. Clustering In many contexts, researchers in several disciplines have addressed the clustering problem; it reflects its broad appeal and usefulness as one step in analysing data from exploration[4]. Traditional clustering techniques form a unique cluster of data set attributes. Only one cluster is assigned to each data object in the sample space. The most frequently used methods are the K-means algorithm or its various variations. The value k is the initial cluster number for the algorithm. The algorithm takes m objects into clusters k and partitions. The technique works to incorporate items into one cluster by calculating the distance between the data item

and cluster centre so intra-cluster similarities are high but cluster similarities are low. [5].

In all engineering disciplines optimization is a common mathematical concept. This just means the best/desired solution to be found. Optimization problems are broad and different methods should be an active field of research to resolve these problems. Algorithms of optimization may be either stochastic or deterministic. Former approaches to optimization require immense analytical efforts, which appear to be failing with the increasing size of the problem. This is the stimulation to use genetically coded stochastic optimization algorithms as computationally effective alternatives to detergents. Meta-heuristics are based on the iterative improvement in the population of both solutions, as in Evolutionary Algorithms, Swarm-based Algorithms, or on a single solution (e.g. tabu search). [6].

II. INTRUSION DETECTION SYSTEM

Intrusion detection is the behavioural detection of intruders against IT systems. This so-called intrusion is intended to achieve unauthorised access to the computer system. External or internal intruders may be. Intruders inside the network are users with some legal access to increase their privileges of access to abuse with unauthorised privileges. Outside the target network, external intruders are users attempting to get unauthorised system information access [7].

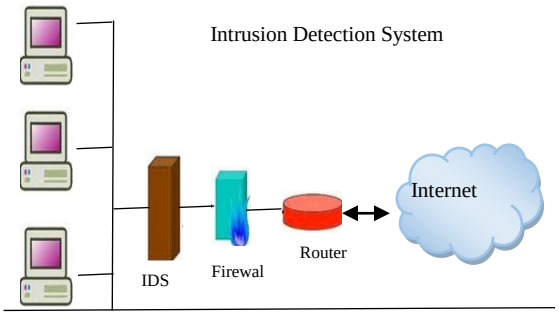

Figure 1: Intrusion Detection System

There are multiple and potentially devastating risks to networks. Back in the time, studies have discovered IDS capable of detecting attacks in many available environments [8].

Monitor, identify, and respond to any illegal access are adages of IDS. The analysis of network traffic will detect network attacks like DoS attacks. The host and the network-based are two basic forms of intrusion detection. There are common benefits and drawbacks of each monitoring and data security approach.

- ***Host-based intrusion detection systems (HIDS)*** They are IDSs running on a single workstation. Through using resources of its host to detect attacks, HIDS track traffic on its host system [9].

- ***Network-based intrusion detection systems (NIDS)*** This IDSs are stand-alone devices that operate on a network. NIDS monitor network traffic in order, by monitoring network traffic, for attacks such as a denial of service, port scans and even computer cracking efforts. [9].

TABLE 1: NIDS VS HIDS [10]

Network-Based Intrusion Detection Systems	Host-Based Intrusion Detection Systems
• Computer/application attached to a portion of the organisational network and monitors network trafficator for continuous attack or success in that segment. • NIDS types include Snort, Cisco NIDS and Netprowler • When placed next to a network system, NIDS uses a monitoring port, switch, like the hub. The port tests any traffic through the device. • Works on the rule that attack patterns match with known signatures on their database Compare. • Due to its volume of data and resources, NIDS are suitable for medium to large organisations. Many smaller companies also are hesitant to use IDS.	• The host and only monitor activities in this strategy to enforce malicious programmes reside in this computer and server. • HIDS types include Tripwire, Cisco HIDS, & ESM from Symantec. • It can monitor system databases like Windows registries and store configuration files like.ini,.cfg and.dat. • Work on the configuration rule and change management. If file attribute changes are deleted, files that are new and existing, a notification is triggered. • Most HIDS usually have common architectures, which means that most system works as organising agents that report to a central console.
Advantages	**Advantages**
• Several devices with a good network design can monitor large networks. • The deployment of NIDS should not disrupt existing network operations since they are passive devices. • The NIDS cannot be directly attacked and cannot be detected by attackers.	• HIDS can detect NIDS and Local Events attacks. • The host system HIDS functions where encrypted traffic is decrypted and processable. • HIDS does not affect the use of a switched network. • HIDS can identify inconsistencies in the application.
Disadvantages	**Disadvantages**
• When network volume becomes overwhelming NIDS can not be aware of the attack. • Since several switches have limited port monitoring capability or no port monitoring capability, some networks cannot provide all data for NIDS review. • NIDS is not able to analyse encrypted packets, invisible some traffic and reducing NIDS effectiveness. • Unable to easily detect an attack with the fragmented or malformed packet.	• More management efforts to set up and maintain HIDS are required. • The host operating system will both be specifically targeted and attacked and the HIDS functionality is compromised and/or lost. • Certain DoS attacks are susceptible to HIDS. • A large amount of disk and disc space has to be applied to the audit logs for the host OS, so that system performance can be reduced. • Multi-host and non-host network computers cannot be scanned/identified by HIDS.

A. INTRUSION DETECTION TECHNIQUES

The intrusion detection system can, according to analytical methods, be divided into two categories, one abnormal and other an abnormal detection of signature detection.

1. Anomaly Detection

Anomaly systems follow the opposite, namely identification of normal behaviour and assessment of differences from normal conduct. These differences are perceived as differences or possible intrusions. Anomaly detection systems rely on the awareness of normal behaviour to detect threats. This detects even new attacks as long as the behaviour of attacks is sufficiently different from normal behaviour. However, if the behaviour is identical to normal behaviour, this may be detected. In contrast, it is difficult to

associate variations with specific attacks. But as users change one's behaviour, normal behaviour should be reinterpreted.

2. Misuse detection

A prior attack knowledge is used to search for signs of attack by misuse detection systems. In other words, they detect intrusion by knowing what violence is. Signature (regular) systems are the most popular forms of malicious detection systems. In the signature-based identification, the attack signatures in the tracked resource are pursued. In the case of documented attacks, the signature-based method is necessarily highly particular. Furthermore, the type of attack can readily be detected since the signatures are related to that misuse activity. However, from its detection capabilities, they can only be found in the signature database. A signature database has to be continually modified to include signatures for new attacks. [11].

B. TOOLS IN INTRUSION DETECTION

That said, so many tools for a variety of farmers in diverse security attacks are being designed and implemented. IDS, which can control many computer systems: information system, network or cloud, is included in the tools. The IDS detects and describes intrusion detection systems as efforts to break strategic goals, such as integrity & availability and nonrepudiation.

1. SNORT

Snort is a lightweight or open-source software. Snort uses flexible rule-based language for traffic classification. It can read the packet in human form from an IP address. Snort detects thousands of worms, port scans and vulnerability exploit attempts by reviewing the protocol, scanning the content, and multiple preprocessors.

2. OSSEC-HIDS

the (OSSEC) open source security is free open-source software. This operates on the main operating system which uses the client/server interface. For analysis and storage, OSSEC can send OS logs to the server. It is fitted with an efficient log analyser, ISPs and universities. Track and evaluate authentication logs, firewalls and HIDS.

3. OpenWIPS-NG

OpenWIPS-NG is freely available wireless IDS/IPS based on servers, sensors and interfaces. It operates on hardware for products. This system was created by the author of Aircrack-NG and uses much of Aircrack-current NG's functions and resources for the screening, identification and intrusion prevention. OpenWIPS-NG is modular and allows an administrator with additional features to download plug-ins. The documentation is not as comprehensive as some programmes but helps companies to run WIPS on a small budget.

4. FRAGROUTE

This is considered a fragmenting router. Here the IP packet is sent from attacker to fragrouter and fragmented into the party.

5. KISMET

That's a guideline for WIDS. WIDS compromise packet load and WIDS intrusion detection systems. The burglar access point is found here.

6. HONEYD

Honeyd is a tool for creating network virtual hosts. The host Honeyd uses services for which single host will query for several network simulation addresses on LAN. The virtual machines may be knocked down or tracked across. Any type of service can be simulated on a virtual machine in simple settings file [12].

III. NETWORK INTRUSION DETECTION SYSTEMS

Intrusion is defined as any computer activity to access information illegally, to listen to the network without authorization or dangerous device [13]. Attack prevention is one of the most important cybersecurity elements [14]. NIDS for IT service administrators is critical tools for detection of company network security policy violations. The NIDS monitor and analyse the traffic of network circulating in network and security devices of organisations, which causes the alarm to notify detected intrusion administrators. [15].

There are two separate NIDs: signature-based NIDS, which also confirms patterns of network operational identification (reference profile) that align captured traffic with predefined signatures of a safety professionals database and anomaly detection NIDS (ADNIDS), alert when a deviation from a reference profile has been detected. [16].

To evaluate NIDS, scientists working on developing & improving IDS are releasing so several data sets on the Internet. In particular, NSL-KDD is the basis for KDD Cup 99, established for the Machine Learning Competition in 1999 and intended to divide ties into 5 different categories. This

categorises network flows as regular or irregular (not intrusion) and can be categorised as attacks [17]:

- **DOS:** Denial services is an aggression category that uses the victim's tools to deter legitimate needs.
- **Probe:** The objective of this form of attack is to extract information on Ex, the victim of port scanning.
- **U2R:** Unauthorized root access is a type of assault that uses an attacher to sign in regular account and seeks to obtain the root/manager rights on the vulnerability, for example of a victim's system, buffer overflow attacks.
- **R2L:** Unauthorized remote machine improvements where the attacker wants to access the remote machine and locally accesses.

The following are the advantages of using IDS dependent on the network:

- Many attackers can ignore network-based IDSs to ensure attack security.
- Certain network-based IDSs will monitor the wide network. Networked IDSs are normal passive devices that listen to the network wire without interfering with normal network operation. Network IDSs with much less effort can easily be integrated into the current network.

The disadvantages of network-based IDS are:

- Network IDSs cannot evaluate encrypted data when different organizations are using proprietary virtual networks.
- Network-based IDS would not benefit most for the narrow network segment i.e. network-based switching.
- The network-based IDS control set is limited to a single host when monitoring the range of switches, they are not universal.
- Several network-based IDS deal with attacks on the network. The IDS depending on the network requires fragmentation of the packet. The IDS is unreliable and crashed due to these anomalous packets [18].

IV. DATA SETS FOR IDS

There are various datasets accessible, most have few other limitations. DARPA 98/99 or KDD99 are most known but slow and criticised, for example. Mahoney and Chan. Mahoney and Chan. However, these data sets are still used today but are helpful findings and analyses. As already stated, several changes were made by NSL-KDD; Qian et al. proposed yet another DARPA redesign.

1. KDD Cup 99

The 1999 intrusion protection contest of KDD aims to create a standard survey dataset and analyse intrusion detection analysis prepared and controlled under the DARPA Intrusion Detection Programme's MIT Lincoln Laboratories. Intrusion Detection Contest After nine weeks of raw TCP dump data for LAN, which imitates the standard United States. They operated the LAN as an area of actual airpower but peppered the LAN with various attacks. Air Force LAB.

2. GureKDDcup

The dataset GureKDDcup contains kddcup99 connections (UCI repository database) but adds their payload (network packet content) to any link. It allows information to be extracted from each connection's payload directly for the process of machine learning.

3. NSL-KDD

This is a compilation of the data proposed in the KDD'99 data set to address some basic problems. Some of McHugh's remaining issues with KDD details. Given the lack of public data sets for network-based IDSs, the collection of data cannot represent actual specific networks optimally. However, since it is not an effective intrusion detection set, we also use 14 years of age datasets for testing intrusion detection models. Also, there are fair numbers of trains and test sets in the NSL-KDD train (125973 samples) as well as in 22544. This advantage allows experiments to be carried out in detail without having to select a small variety randomly. As a result, assessment findings will be accurate and comparable to a range of research projects. The data set contains no redundancy study. The test set contains certain attacks that are not included in the training set [19].

V. OVERVIEW OF CLUSTERING

In classes with similar objects, clustering procedures for different physical and abstract objects are called. The cluster contains the same objects and separates them from all objects in the same cluster. With clustering, crowded and sparse areas, patterns recognition and interesting links between data attributes can be detected insignificant share. It can also be used for matching patterns and data mining. These methods are also valuable by identifying unknown patterns.

A. CLUSTERING IN INTRUSION DETECTION

Clustering is used for group training instances in clusters using the basic distance metric. Such methods may be used as exceptions or normal as standard instances where the data is clustered. The framework is ready to accept network data instances after clustering has expired and the clusters are named accordingly, to equate them with the clusters already established and classify them as possible threats or instances of safety to detect possible invading events. [20].

The clustering means the division into subsets of associated objects into separate classes of data (clusters), such that the data share frequent proximity in any subset (ideally) according to a given size. Clustering is an analysis field that can be utilised as an independent method for providing insight into the allocation of data, the enforcement of each cluster's characteristics and the focus for further study on a detailed set of clusters.

B. IDS USING CLUSTERING ALGORITHM

The Cluster technique is used primarily to gather and detect outliers of data with similar characteristics. Similarly, classification is a procedure used to predict a data object type mark based on data objects previously encountered [21].

1. K-Means Clustering algorithms

K-means is the most simple and widely used clustering algorithm technique K-means. In this algorithm, several clusters K is classified in a predefined number of clusters by way of user means. The first stage of clustering K-Means is to select k instances as a cluster core. First, assign the closest cluster to each dataset instance. For eg, calculate the distance between the centroid and each specific instance using Euclidean distance, and assign each data point to cluster by a minimum distance. K – Means, when used on a small dataset, the algorithm requires a less runtime. When the data point is raised to the limit, the execution time is required. It is easily iterative but susceptible to surface and noise. This is used by the central point instance to find perpetrators of the wireless network. Then, each attacker is assigned to the nearest point or node, the attacker is determined by the least distance and the attacker is submitted to servers.

2. K-Medoids clustering algorithms

K-Medoids cluster algorithm as a K-means algorithm by partitioning it. The central instance of the cluster is known as central instead of taking the K - means the value of objects in the K-means cluster. This entity is fundamental and is known as the reference point and medoid. It decreases the distance between the middle and data points to a minimal. When data points are increasing to limit, KMedoids algorithm functions better than K-Means. It is high with noise and surface because medoids are less affected by external materials, but processing is more expensive. This is used to find the attackers in the wireless network by the instance of the centre point. It acts like the K-Means cluster algorithm and then it assigns each attacker to the nearest point or node with the help of IDS and also it chooses the references to object to find the intruder [22].

VI. OPTIMIZATION ALGORITHMS

The Optimization techniques are a way to achieve decision-making that approaches goals set for this problem. During World War 2 development of improved optimization started. There were too many techniques designed to resolve many types of concerns like single objective, multi-objective, linear, non-linear, etc. Optimization (O) method discovers the alternative under a certain constraint, with mainly cost-effective and maximum possible performance. Therefore, optimizing implies trying to achieve the best or highest results of cost. O is constrained by a lack of complete info & time to decide which data is required to improve the O process. The O method is used to achieve them. The best can be a single organization and objective process which models a similar entity. The process of O shall be applied to factors that determine the best varies with the situation. Several examples are optimized costs used raw materials and time. O can be done for local and global optimal achievement. Several major optimization algos are available in each field. Traditional techniques find this kind of difficulties in some problems. Modern optimization techniques.

A. MODERN OPTIMIZATION TECHNIQUES

1. Genetic Algorithm

Genetic algorithm is an evolutionary class of stochastic research strategies. This is a popular strategy to optimise a huge number of variables in non-linear systems. The aim function for randomly chosen domain description points is assessed by genetic algorithms. A new collection of points (a new population) is generated with this information in mind. Local maximum and minimum function steadily reaches points in the population. The function must not be continuous or distinct. Also in situations where the function has several local minima or maximum, genetic algorithms will still produce a good performance.

2. Ant Colony Method

Ant colony optimization method is Probabilistic technique. It searches for the perfect path in the graph based on the conduct of centuries searching for the path between their colony and their food supply. It's a meta-heuristic method of optimisation. This method's basic principle is based on the behaviour of the ant. Ants navigate from nest to food source. Ants are blind! Ants are blind! Ants are blind! They will find the shortest way along phéromone paths. Each ant moves at the random path. On the path is deposited phéromone. The possibility of more pheromones on the way increases. " The ACO is characterised by a policy analysis approach to the creation of parameters for the stochastic policy of so-called ant agent (known as pheromone variables according to biological metaphor) used to create solutions.

3. Honey Bee Algorithm

Honey bees are one of the most well studied social insects. The series of experiments focussing on the different bee activities solved numerical and combinatory problems during the early years. Beekeeping in nature Social insect colonies are a dynamic system that collects and adjusts information from the environment. As per their specifications, individual insects do not execute all activities during the processing of information and adjustment processes. Usually, all social insect colonies act according to their division of labours related to their morphology.

4. Simulated Annealing Algorithm

In 1953 Kirkpatrick suggested a simulation algorithm for annealing the original idea, which was successfully used in the Metropolis in 1983 for the combinatorial optimization problem. A simulated clothing annealing based on the solid textile theory, high-level heating, slow cooling, heating, solid internal particles in an unordered shape and an increase in temperature, increasing internal energy &gradually cooling particles, and reaching equilibrium at each temperature.

1. Teacher Learning-Based Optimization

TLBO may classify the big number of applications in the various fields, including electrical & mechanical & civil electronic, thermal and biotechnological applications, of the engineering and sciences sector. TLBO used to resolve the problem of constraints & unconstraint [23,24].

VII. LITERATURE SURVEY

J. V. Anand Sukumar et al.(2018) In this paper they know of an intrusion detection method using a genetic improved k-mean (IGKM) algorithm to detect intrusion. In comparison, k-Means++ and intrusion detection methods that use IGKM by using a small KDD-99 dataset subset with thousand-instance dataset are used as the paper compares the intrusion mechanism. The IGKM algorithm is seen to be much more precise than the k-means++ algorithm. [25].

H. Zhang et al. (2018) Suggest random forest structure to manage traffic data at high speeds. Network for intrusion detection. The framework consists of three parts: data collection part based on NetFlow, the data pre-processing part and intrusion detection part based on classification. The random forest classification algorithm is used in this article or adapted to Apache Spark's processing method for real-time detection. We implement the system or run too many comparison experiments to measured the accuracy of the structure. The results show that, compared to current systems, the system is effective and reliable and is therefore well placed for the intrusion detection into networks in real-time at high capacity and speed. [26].

P. Verma et al. (2018) In this paper the use of algorithms for detection of network intrusion was suggested with and without clustering. Their findings indicate better accuracies than existing models. Although test data from NSL-KDD are based on a different distribution than training data, we have been able to build a solid model. [27].

Chen, Z.-H et al. (2018) This study introduces a new hybrid IDS classification algorithm. The new metaheuristic algorithm or k-mean cluster algorithm is the basis of the proposed algorithm. The test results reveal that it could have a high detection accuracy rate, given that the proposed algorithm can maintain search diversity and during the convergence process. [28].

C. Long et al. (2019) Propose a (Gaussian Mixture Model) GMM & (k-Nearest Neighbor) K-NN hybrid intrusion detection algorithm We apply GMM to categorise each group's spatial distribution first. The new distance and density extraction training procedure, called GMDD, is then conducted on GMM. GMDD produces more concise and high-quality characteristics with sufficient classification details in original data. K-NN is eventually educated and tested for the detection model using newly modified data sets. Experimental results on KDD'99 or NSL-KDD data sets have shown that the method proposed provides great computational efficiency and not only does benefit other recent studies into accuracy, detection rate and false alarm rate. [29].

D. Liang et al. (2019) This document presents a method that integrates clustering and supporting vector machine algorithm

to enhance IDS accuracy and recognition speeds. First the pre-processed data is clustered and grouped into different sub-sets and then an algorithm for machine learning is used to model the individual sub-sets. Compared with other state-of-the-art algorithms, they showed that our method significantly reduced the training time of the model and increased the model's performance effectively. [30].

M. O. Miah et al. (2019) The new strategy has been applied to increase the identifying rate to identify class imbalance network attacks and intruders using Research of Random Forest Cluster. The proposed method is a multi-faceted classified method that can correctly distinguish incursions by examining very unequalled big data between majorities and rare individuals. Attack or not (as normal behaviour, if data point/incoming data is currently specified in the proposed approach), then attempt to define the form of attack, sub attack. To solve the imbalanced problem in the class-imbalanced and common Random Forest classification ensemble, cluster-based under-sample sampling was used. The performance of the proposed approach was evaluated with the use of existing machine-learning algorithms, including naive Bayes and Artificial neural Network, country-wide random forest and Bagging technologies. KDD99 data set for intrusion detection is used for testing and evaluation [31].

W. Liang et al. (2020) The proposed algorithm will increase the rate of identification and real-time detection of suspicious behaviour for multipurpose data in industrial networks effectively. The new feature is twice as large as cluster centre to pick the node with high safety coefficient easily and one cluster is paired with multifunction data. Experimental results reveal that, relative to other algorithms, the proposed algorithm is very high in detection and time. The suspicious data detection in the industrial network approaches 97,8%, while the detection FP is decreased by 8,8% [32].

VIII. CONCLUSION

The issue of network intrusion detection has recently become an important concern, as computer networks, internet, computer network exponential use are used to a large extent. The safety stress monitoring system & intrusion detection in the network is a recent field of research, this is of great importance in improving disaster response capability, reducing network losses and abnormal intrusions, and enhancing fighting capacity of the system. The IDS is a key component in the security of the computer network, and clustering analysis is an uncontrolled detection method. However, the highly efficient detection using a single clustering algorithm cannot be achieved, and intrusion data are usually anomalous. The clustering algorithm for intrusion detection is used to solve the problem of lack of prior intrusion detection as a supervised learning algorithm. Various techniques of optimisation are discussed in this paper. A brief description of optimisation techniques based on clustering. Genetic algorithm, Ant colony, Honey bee algorithms, simulated Annealing and so on are provided with some optimisation techniques.

REFERENCES

[1] Dali, L., Bentajer, A., Abdelmajid, E., Abouelmehdi, K., Elsayed, H., Fatiha, E., & Abderahim, B. (2015). *A survey of the intrusion detection system. 2015 2nd World Symposium on Web Applications and Networking (WSWAN).* doi:10.1109/wswan.2015.7210351

[2] Paul Innella and Oba McMillan, Tetrad Digital Integrity, LLC "An Introduction to Intrusion Detection Systems" December 6, 2001

[3] Snehil Dahima," A Survey on Various Data Mining Technique in Intrusion Detection System", IOSR Journal of Computer Engineering (IOSR-JCE), ISSN: 2278-8727, Volume 19, Issue 1, Ver. I (Jan.-Feb. 2017), PP 65-72

[4] Raúl Sánchez," Clustering for Intrusion Detection: Network Scans as a Case of Study", Springer, 2013, pp. 33–45

[5] Richa Sampat, Shilpa Sonawani," A Survey of Fuzzy Clustering Techniques for Intrusion Detection System", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 3 Issue 1, January – 2014

[6] Binitha S, S Siva Sathya," A Survey of Bio inspired Optimization Algorithms", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-2, Issue-2, May 2012

[7] Bruno Bogaz Zarpelãoa," A survey of intrusion detection in Internet of Things", Journal of Network and Computer Applications, 2017, http://dx.doi.org/10.1016/j.jnca.2017.02.009

[8] Sabahi, F., & Movaghar, A. (2008). *Intrusion Detection: A Survey. 2008 Third International Conference on Systems and Networks Communications.* doi:10.1109/icsnc.2008.44

[9] Christos Douligeris and Dimitrios N. Serpanos "Network Security Current Status and Future Trends"

[10] Karthikeyan .K.R and A. Indra," Intrusion Detection Tools and Techniques – A Survey", International Journal of Computer Theory and Engineering, Vol.2, No.6, December, 2010, 1793-8201,pp.901-906

[11] Rachna kulhare," Survey paper on intrusion detection techniques", International Journal Of Computers & Technology, May 20 , 2013, ISSN 22773061, Vol 6, No 2,pp. 329-337

[12] Jayesh Surana," A Survey On Intrusion Detection System", International Journal of Engineering Development and Research, 2017 IJEDR | Volume 5, Issue 2 | ISSN: 2321-9939,pp.1-6

[13] OUIAZZANE, S., ADDOU, M., & BARRAMOU, F. (2019). *A Multi-Agent Model for Network Intrusion Detection. 2019 1st International Conference on Smart Systems and Data Science (ICSSD).* doi:10.1109/icssd47982.2019.9003119

[14] Rathore et al. – 2016 – Real time intrusion detection system for ultrahigh speed environments – doi : 10.1007/s11227-015-1615-5

[15] Obeidat, I., Hamadneh, N., Alkasassbeh, M., Almseidin, M. & AlZubi – 2019 – Intensive Pre-Processing of KDD Cup 99 for Network Intrusion Classification Using Machine Learning Techniques

[16] Javaid et al. – 2016 – A Deep Learning Approach for Network Intrusion Detection. doi: 10.4108/eai.3-12-2015.2262516

[17] Dhanabal, L Shantharajah, S P – 2015 – A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms – International Journal of Advanced Research in Computer and Communication Engineering.

[18] Snehil Dahima," A Survey on Various Data Mining Technique in Intrusion Detection System", IOSR Journal of Computer Engineering (IOSR-JCE), ISSN: 2278-8727, Volume 19, Issue 1, Ver. I (Jan.-Feb. 2017), PP 65-72

[19] Rashmi Ravindra Chaudhari," Intrusion Detection System: Classification, Techniques And Datasets To Implement", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 02 | Feb -2017

[20] Samarjeet Borah, Debaditya Chakravorty," Advanced Clustering Based Intrusion Detection (ACID) Algorithm", CCIS 192, pp. 35–43, 2011

[21] Nandini Rebello, Manamohan K," Network Intrusion Detection System using K-Means Clustering and Gradient Boosted Tree Classifier", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019

[22] K.Kathirvel," A Survey on IDs Using Clustering Techniques in Data Mining", International Journal of Pure and Applied Mathematics, Volume 118 No. 8 2018, 655-660,pp.1-6

[23] Pinto, Vivek D., and William M. Pottenger. "A survey of optimization techniques being used in the field" In the Proceedings of the Third International Meeting on Research in Logistics (IMRL. 2000).

[24] Sanket A. Pandya," A Review Of Modern Optimization Techniques", Journal of Emerging Technologies and Innovative Research (JETIR), July 2017, Volume 4, Issue 07

[25] J. V. Anand Sukumar, I. Pranav, M. Neetish and J. Narayanan, "Network Intrusion Detection Using Improved Genetic k-means Algorithm," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 2441-2446, doi: 10.1109/ICACCI.2018.8554710.

[26] H. Zhang, S. Dai, Y. Li and W. Zhang, "Real-time Distributed-Random-Forest-Based Network Intrusion Detection System Using Apache Spark," 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC), Orlando, FL, USA, 2018, pp. 1-7, doi: 10.1109/PCCC.2018.8711068.

[27] P. Verma, S. Anwar, S. Khan and S. B. Mane, "Network Intrusion Detection Using Clustering and Gradient Boosting," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore, 2018, pp. 1-7, doi: 10.1109/ICCCNT.2018.8494186.

[28] Chen, Z.-H., & Tsai, C.-W. (2018). *An Effective Metaheuristic Algorithm for Intrusion Detection System. 2018 IEEE International Conference on Smart Internet of Things (SmartIoT).* doi:10.1109/smartiot.2018.00036

[29] C. Long, Y. Zhang, J. Wei, W. Wan, J. Zhao and G. Du, "A Hybrid Intrusion Detection Algorithm Based on Gaussian Mixture Model and Nearest Neighbors," 2019 IEEE 44th Conference on Local Computer Networks (LCN), Osnabrueck, Germany, 2019, pp. 117-120, doi: 10.1109/LCN44214.2019.8990852.

[30] D. Liang, Q. Liu, B. Zhao, Z. Zhu and D. Liu, "A Clustering-SVM Ensemble Method for Intrusion Detection System," 2019 8th International Symposium on Next Generation Electronics (ISNE), Zhengzhou, China, 2019, pp. 1-3, doi: 10.1109/ISNE.2019.8896514.

[31] M. O. Miah, S. Shahriar Khan, S. Shatabda and D. M. Farid, "Improving Detection Accuracy for Imbalanced Network Intrusion Classification using Cluster-based Under-sampling with Random Forests," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.8934495.

[32] W. Liang, K. Li, J. Long, X. Kui and A. Y. Zomaya, "An Industrial Network Intrusion Detection Algorithm Based on Multifeature Data Clustering Optimization Model," in IEEE Transactions on Industrial Informatics, vol. 16, no. 3, pp. 2063-2071, March 2020, doi: 10.1109/TII.2019.2946791.