Intrusion Detection System Based on Hybrid Optimization and using Neural Network: A Review

Amanpreet Singh[1], Akhil Goyal[2]
CDAC Mohali
(E-mail: aman.bimbro@gmail.com)

***Abstract*—**An intrusion detection system is composed of the following three components: Sensors: - which sense the network traffic or system activity and generate events. Console: - to monitor events and alerts and control the sensors, Detection Engine: - that records events logged by the sensors in a database and uses a system of rules to generate alerts from the received security events. In this paper intrusion detection system is explained in details. Various types of intrusion detection system such as Network Intrusion Detection System, Host-based Intrusion Detection System and Hybrid intrusion detection system are discussed in this paper. For the implementation MATLAB simulator will be used. MATLAB is a stage named as matrix laboratory which give the numerical registering in multi-design. It is likewise called as the fourth era dialect of programming. In this paper matlab and its working environment is also discussed. The author in this paper has given the idea of hybridization by using two algorithms named as Flower pollination Algorithm and Particle Swarm Optimization.

Keywords*—IDS; PSO; FPA;Matlab; API*

I. INTRODUCTION

An intrusion detection system (IDS) is composed of hardware and software elements that work together to find unexpected events that may indicate an attack will happen, is happening, or has happened. Note that we must think in all three tenses; some products warn in advance that an attack may take place, some warn as they notice an attack in progress, and some warn when they notice the aftereffects of the attack. An IDS is composed of the following three components: Sensors: - which sense the network traffic or system activity and generate events. Console: - to monitor events and alerts and control the sensors, Detection Engine: - that records events logged by the sensors in a database and uses a system of rules to generate alerts from the received security events. There are several ways to categorize an IDS depending on the type and location of the sensors and the methodology used by the engine to generate alerts. In many simple IDS implementations all three components are combined in a single device or appliance

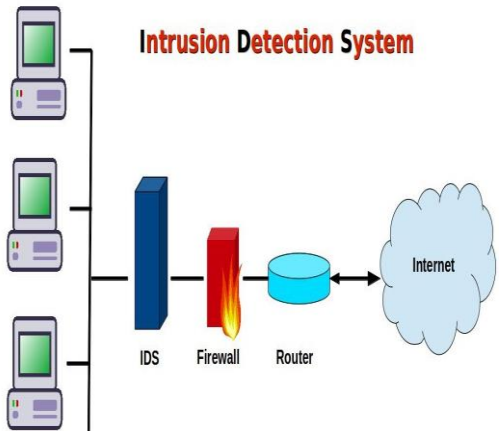

Figure 1: Intrusion detection system

A. *Types of Intrusion Detection systems*

Network Intrusion Detection System: - identifies intrusions by examining network traffic and monitors multiple hosts. Network Intrusion Detection Systems gain access to network traffic by connecting to a hub, network switch configured for port mirroring, or network tap. An example of a NIDS is Snort.

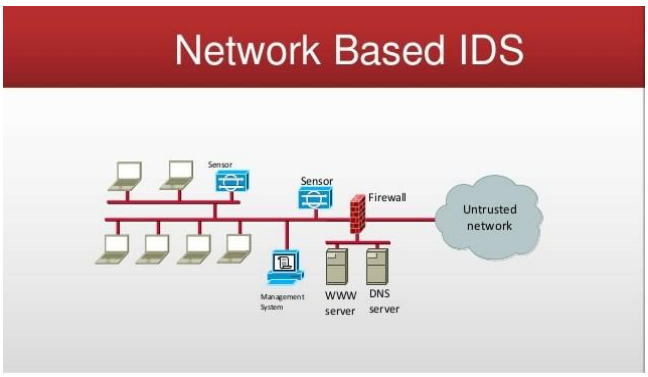

Figure 2: Network based IDS

Host-based Intrusion Detection System: - consists of an agent on a host which identifies intrusions by analyzing system calls, application logs, file-system modifications

(binaries, password files, capability/acl databases) and other host activities and state.

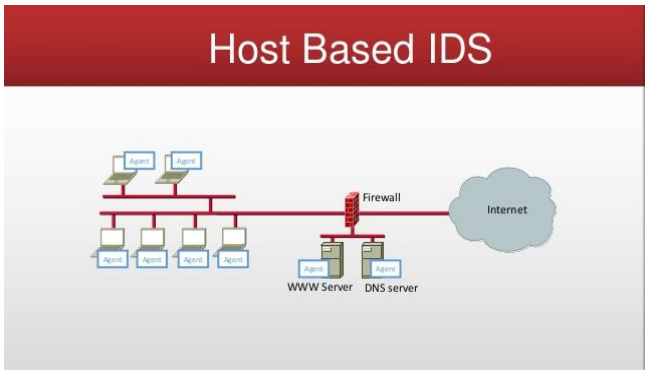

Figure 3: Host based IDS

Hybrid Intrusion Detection System: - combines one or more approaches. Host agent data is combined with network information to form a comprehensive view of the network. An example of a Hybrid IDS is Prelude.

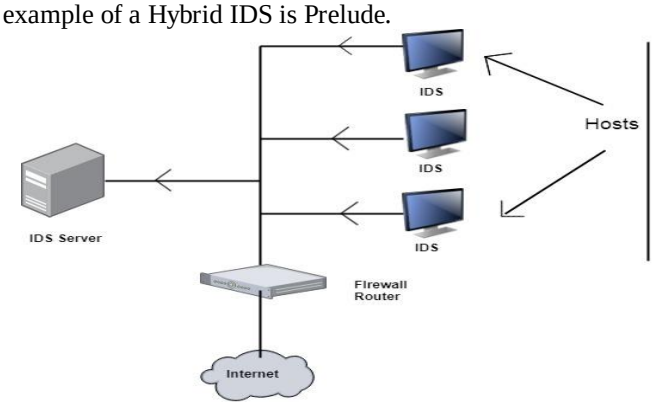

Figure 4: Hybrid IDS

B. Advantages of IDS

- Allows administrator to tune, organize and comprehend often incomprehensible operating system audit trials and other logs.
- Can make the security management of the system by non-expert staff possible by providing user friendly interface.
- Can recognize and report alteration to data files.
- IDS generate alarm and report to administrator that security is breaches and also react to intruders by blocking them or blocking server.
- It provides time to time information, it recognize attacker and report alteration to data file.

II. LITERATURE REVIEW

Ravale, Ujwala et al. [1] The proposed hybrid technique combines data mining approaches like K Means clustering algorithm and RBF kernel function of Support Vector Machine as a classification module. The main purpose of proposed technique is to decrease the number of attributes associated with each data point. So, the proposed technique can perform better in terms of Detection Rate and Accuracy when applied to KDDCUP'99 Data Set.

Parvat, Thaksen J. et al. [2] Intrusion detection and prevention is a fundamental piece of packet inspection. There are numerous techniques and calculations to distinguish marks; every ha its own particular benefits and constraints. The measure of this time, exactness and space necessity. All strategies and calculations are upgraded with innovation unrests. The advancement in irregularity and abuse discovery in this decade is significant as web administrations become huge. Overseeing secure system is a test today. The goals change as per the foundation administration and security strategy. There are different approaches to distinguish payload movement utilizing DPI, Network security, protection and QoS. The elements of DPI are convention discovery, against infection, hostile to malware and IDS/IPS. The identification motor may support by a marks or heuristics. The greater part of the calculations do preparing and testing independently, it takes around double time.

Aggarwal, Preeti et al. [3] This paper introduces the examination of KDD informational index regarding four classes which are Basic, Content, Traffic and Host in which all information properties can be arranged. The investigation is finished as for two noticeable assessment measurements, Detection Rate (DR) and False Alarm Rate (FAR) for an Intrusion Detection System (IDS). Because of this exact examination on the informational index, the commitment of each of four classes of properties on DR and FAR is indicated which can help improve the appropriateness of informational index to accomplish greatest DR with least FAR.

Pawar, Sunil Nilkanth et al. [4] In this paper, it is proposed to utilize variable length chromosomes (VLCs) in a GA-based system interruption recognition framework. Less chromosomes with applicable highlights are utilized for administer era. A successful wellness work is utilized to characterize the wellness of each run the show. Every chromosome will have at least one principles in it. As every chromosome is an entire answer for the issue, less chromosomes are adequate for powerful interruption recognition. This decreases the computational time. The proposed approach is tried utilizing Defense Advanced Research Project Agency (DARPA) 1998 information. The test comes about demonstrate that the proposed approach is effective in arrange interruption location.

Karkouch, Aimad, et al. [5] In the Internet of Things (IoT), information accumulated from a worldwide scale organization of brilliant things, are the base for settling on astute choices and giving administrations. On the off chance that information are of low quality, choices are probably going to be unsound. Information quality (DQ) is critical to pick up client engagement and acknowledgment of the IoT worldview and administrations. This paper goes for upgrading DQ in IoT by giving a review of its best in class. Information properties and their new lifecycle in IoT are overviewed. The idea of DQ is characterized and an arrangement of non specific and space particular DQ measurements, fit for use in evaluating IoT's DQ, are chosen. IoT-related variables imperiling the DQ and their effect on different DQ measurements and on the general

DQ are thoroughly dissected. DQ issues indications are talked about and their side effects recognized. Information anomalies, as a noteworthy DQ issue appearance, their basic learning and their effect with regards to IoT and its applications are considered. Procedures for upgrading DQ are given an extraordinary concentrate on information cleaning strategies which are inspected and contrasted utilizing a broadened scientific categorization with layout their attributes and their wellness for use for IoT.

Guo, Chun, et al. [6] In this exploration, we propose another and simple to-execute half breed learning technique, named distance sum-based support vector machine (DSSVM), which can be utilized as a viable interruption recognition display. In DSSVM, we present the distance sum, a connection between's every datum test and group focuses. Consider an informational index spoke to by n-dimensional element vectors, each distance sum for an information test in the informational collection is gotten from the distances between this information test and k−1 of kcluster focuses found by a grouping calculation. Another informational collection speaking to the highlights of these distance sums is shaped and used to prepare a support vector machine classifier. By applying DSSVM to the KDD'99 informational index, our test comes about demonstrate that the proposed crossover technique performs well in both location execution and computational cost, which recommends it is a focused possibility for intrusion detection.

Bamakan, Seyed Mojtaba Hosseini, et al. [7] In this paper they show another technique in view of numerous criteria direct programming and molecule swarm advancement to upgrade the precision of assaults discovery. Different criteria direct writing computer programs is an arrangement technique in light of scientific programming which has been demonstrated a potential capacity to take care of genuine information mining issues. In any case, tuning its parameters is a basic strides in preparing stage. Particle swarm optimization (PSO) is a hearty and easy to actualize streamlining system has been utilized as a part of request to enhance the execution of MCLP classifier. KDD CUP 99 dataset used to assess the execution of proposed strategy. The outcome showed the proposed demonstrate has tantamount execution in light of location rate, false alert rate and running time contrast with two other benchmark classifiers.

Ghanem, Tamer F. et al. [8] This paper proposes a hybrid approach for irregularity recognition in huge scale datasets utilizing identifiers created in view of multi-begin metaheuristic technique and genetic algorithms. The proposed approach has taken some motivation of negative selection-based detector generation. The assessment of this approach is performed utilizing NSL-KDD dataset which is an altered form of the generally utilized KDD CUP 99 dataset. The outcomes demonstrate its viability in creating an appropriate number of indicators with a precision of 96.1% contrasted with different contenders of machine learning calculations.

Lee, Gangin et al. [9] They propose a correct, proficient calculation for mining unverifiable continuous examples in light of novel information structures and mining procedures, which can likewise ensure the rightness of the mining comes about with no false positives. The recently proposed list-based information structures and pruning strategies permit a total arrangement of questionable incessant examples to be mined all the more productively without design misfortunes. They likewise show that the proposed calculation outflanks past cutting edge approaches in both hypothetical and observational angles. Particularly, they give investigative aftereffects of execution assessment for different sorts of datasets to indicate effectiveness of runtime, memory utilization, and versatility in their strategy.

Aljawarneh, Shadi A et al. [10] In this paper, they likewise introduce the design and assessment of every strategy, and the execution utilized as examples by SRE component to extricate precise marks. Such usage was refined through utilization of the Needleman– Wunsch calculation, which was insufficient to deal with the invariant parts and separations confinements of the polymorphic worm. Thus, an Enhanced Contiguous Substring Rewarded (ECSR) calculation is produced to enhance the outcome extraction from the Needleman– Wunsch calculation and create exact marks. The mark era by SRE is observed to be more precise and proficient as it saves all the vital highlights of polymorphic worms. The assessment comes about demonstrate that the mark contains conjunctions of tokens, or token subsequence can deliver lost essential data, for example, overlooking one byte token or ignoring the limitation separations. Besides, the Simplified Regular Expression should be refreshed and precise when contrasted and signature and polygraph techniques.

Ahmad, Iftikhar, et al. [11] The support vector machine (SVM) is utilized for characterization reason. This examination work utilized the learning revelation and information digging glass dataset for experimentation. The execution of this approach was broke down and contrasted and existing methodologies. The outcomes demonstrate that proposed technique improves SVM execution in interruption recognition that outflanks the current methodologies and has the ability to limit the quantity of highlights and expand the location rates.

Airehrour, David et al. [12] The Internet of Things (IoT) could be depicted as the unavoidable and worldwide system which helps and gives a framework to the checking and control of the physical world through the gathering, preparing and investigation of produced information by IoT sensor gadgets. It is anticipated that by 2020 the quantity of associated gadgets is evaluated to develop exponentially to 50 billion. The fundamental drivers for this development are our regular gadgets, e.g: autos, fridges, fans, lights, cell phones and other operational advancements including the assembling frameworks which are currently getting to be noticeably associated frameworks over the world. It is evident that security will represent a key empowering factor for the effective organization and utilization of most IoT applications and specifically secure routing among IoT sensor hubs along these lines, components should be intended to give secure routing correspondences to gadgets empowered by the IoT innovation. This review investigates existing routing conventions and systems to secure routing interchanges in IoT,. confirm that you have the correct template for your paper size. This template has been tailored for output on the

A4 paper size. If you are using US letter-sized paper, please close this file and download the file "MSW_USltr_format".

III. PROBLEM STATEMENT

In previous work, data set used consists of 41 features, divided into category of 5 classes.

All these features come into consideration while calculating the performance metrics of algorithms. Large number of features not only increases the processing time of intrusion detection but also reduces the accuracy of algorithm

So our aim to improve the efficiency of intrusion detection by optimizing the feature set i.e. reducing the number of features to be considered by using the hybrid approach of Flower Pollination algorithm and Particle Swarm Optimization algorithm.

A. Objectives to be fulfilled

To reduce the feature set using the hybrid approach of Flower Pollination algorithm and Particle Swarm Optimization algorithm.

b) To implement linear classifier on given feature of dataset.

c) To analyze the classifier models on the basis of accuracy, precision, recall and detection rate.

d) We will also compare our proposed work with existing work on various performance parameters.

IV. METHODOLOGY USED

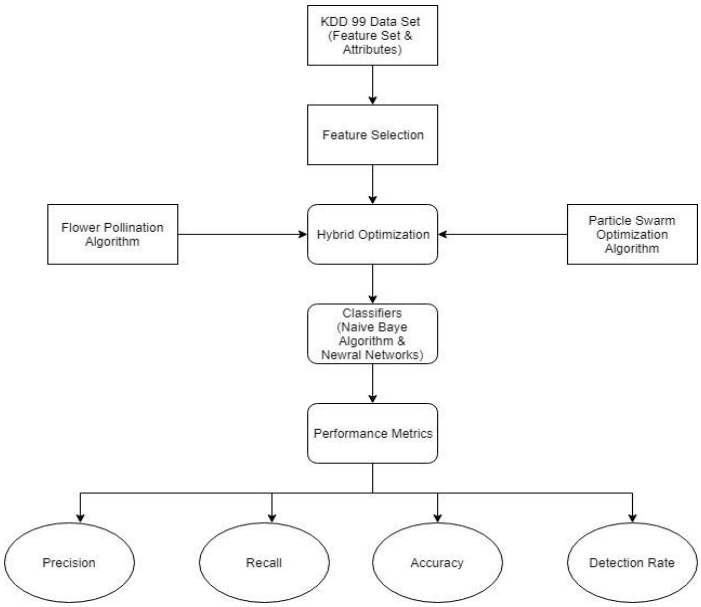

Figure 5: Methodology used

A. Flower pollination Algorithm (FPA):

The Flower pollination optimization algorithm (FPOA) is a recently invented optimization algorithm. It is inherited from the natural inspiration of pollination process. It mimics the process of flowering planets reproduction via pollination. As pollinators are mainly responsible for transferring pollens among flowers, pollination may occur in either local or global flow [4]. Pollination process can fall into two form categorizes; biotic and abiotic based on the pollens transferring mechanism. For biotic pollinations, flowers always depend on insects and/or animals as pollinators to transfer the flowering pollens. However for abiotic, flowers do not need any pollinators for the pollens transferring process [5, 6]. Naturally most of flowers considered to follow the biotic pollination form. This indicates that pollination or crosspollination process can take place by pollinators' movements or travelling long distances causing a global pollination. Travelling pollinators are usually follows the L'evy's flight behavior. Their flying steps are also follows the L'evy's flight distribution [7]. For each kind of pollinators, there is a specific type of flowers that it is responsible for, this called flower consistency. Flower consistency helps to minimize the cost of investigation of each pollinator. Evolutionary wise, it increase the transferring time of pollens and hence optimize and maximize the reproduction process. With the limited available memory of pollinators, flower consistency eliminates the learning, investigation and switching [8]. Furthermore, it can be considered as an incremental step based on the similarity/difference of any two flowers. The biological objective of flower pollination is to optimally reproduce a new enormous generations of the flower kind with the fittest features that ensure the kind`s survival. In order to ideally formalize the flower pollination algorithm, characteristics of pollination process, flower constancy and pollinator behavior should be approximated based on the following essential rules:

i. Global pollination achieved by L'evy's flights` travelling pollinators for both biotic and cross-pollination.

ii. Local pollination achieved abiotic and self-pollination.

iii. The new generation reproduction probability depends on the flower consistency and proportional to flowers` similarities/differences.

iv. The switch probability $p \in [0, 1]$ controls the shift between local and global pollination.

The simple flower pollination model assume that each plant has only one flower, and each flower only produce one pollen gamete. Thus, there is no need to distinguish a pollen gamete, a flower, a plant or solution to a problem [9].

B. Particle Swarm Optimization

Particle Swarm Optimization belongs to the field of Swarm Intelligence and Collective Intelligence and is a sub-field of Computational Intelligence. Particle Swarm Optimization is related to other Swarm Intelligence algorithms such as Ant Colony Optimization and it is a baseline algorithm for many variations, too numerous to list. Particle Swarm Optimization is inspired by the social foraging behavior of some animals such as flocking behavior of birds and the schooling behavior of fish. The Particle Swarm Optimization (PSO) algorithm is already proved efficient in the rule extraction in intrusion detection. But in practice the most intrusion detection systems

often have a high false alarm rate. To solve it, a new PSO-based algorithm which has a special fitness function to extract better rules set with lower false alarm rate to detect the attacks.

C. Classifiers Used For Research Work

1) Naive Bayes Classifier: The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods. Naive Bayes classifier is a straightforward and powerful algorithm for the classification task.

2) Neural networks: A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand.

3) Performance metrices: To The List below shows important performance parameters chosen to analyze the results.

- Precision
- Recall
- Accuracy
- Detection Rate

Precision: It is the number of instances correctly classified as class X among those classified as class X.

$$\text{Precision} = TP/(TP+FN)$$

Recall: It is equivalent to the true positive rate (TPR). It measures performance of the algorithm.

$$\text{Recall} = TP/(TP+FN)$$

Accuracy: It shows how accurate the system can detect attacks.

$$\text{Accuracy} = (TP+TN)/(TP+FP+TN+FN)$$

Detection Rate: It shows percentage of successfully classified incidents

$$DR = TP/(TP+TN)\ X100\%$$

Where TP, FP, FN & TN are True, Positives, False Positives, False negatives & True Negatives numbers

4) Machine learning tool: MATLAB is a stage named as matrix laboratory which give the numerical registering in multi-design. It is likewise called as the fourth era dialect of programming. The analyst of Math Works Inc. Dr. Cleve Molar had gathered the primary type of MATLAB in 1970's. It is created for the understudies so they can get to the LINPACK and EISPACK ventures with no need of taking in the FORTRAN dialect. It allows matrix manipulation, data implementation of algorithms, interfacing with programs written in some other language. MATLAB application is built around the scripting language.

By using MATLAB, abuse of matrix turn out to be simple, the in grouping and capacities can be plotted effectively, figuring can be executed, a calculation can be executed, uncommon customer interfaces can be made also interfacing ought to be plausible with the undertakings which are realized in changed programming dialects like JAVA, C++, C and Python. It is utilized for matrix hypothesis, direct polynomial math and numerical examination. MATLAB application is manufactured around the MATLAB scripting dialects.

Ordinary uses include: Modeling advancement algorithm, reenactment, and Data prototyping inspect, investigation, and Scientific representation and illustrations designing Application improvement, including Graphical User Interface constructing MATLAB is an intelligent framework that essential information constituent is a cluster which does not need dimensioning.

The MATLAB name remains for matrix laboratory. MATLAB was at first written for making accessible simple access for matrix programming created with the EISPACK and LINPACK ventures that commonly speaks for the programming cutting edge for matrix design.

MATLAB has developed over a time of years to contribute from different clients. In college conditions, its the instructional apparatus standard to early on and propelled courses in designing, science and arithmetic. MATLAB in industry is the decision instrument for high-efficiency research, improvement, and look at.

a) The MATLAB System: The MATLAB system includes of three main parts:

The MATLAB language: This is a state matrix/cluster language with control articulation stream, information structures, capacities, input yield, and protest programming situated highlights. Both are permitted in it "programming in the little" for rapidly making discard no-nonsense projects, and "programming in the substantial" for making total vast and application complex programs.

The MATLAB working environment:This is the instruments arrangement and comforts working with as the MATLAB client or software engineer. It integrates pleasantries to deal with your workspace factors and information brings in and sends out. It includes likewise tools for creation, overseeing, troubleshooting, and M-documents profiling, MATLAB's applications.

The MATLAB Application Program Interface (API):

This is a library which enables you for composing Fortran and C programs participating with MATLAB. It incorporate comforts for schedules calling from MATLAB (dynamic connecting), MATLAB called as a computational motor, and to peruse and MAT-records composition.

V. CONCLUSION

An intrusion detection system is composed of the Sensors, Detection Engine and Console. Sensor is the one which sense the network traffic or system activity and generate events. Console: - to monitor events and alerts and control the sensors, Detection Engine: - that records events logged by the sensors in a database and uses a system of rules to generate alerts from the received security events. In this paper intrusion detection system is explained in details and literature review of intrusion detection system has been discussed. Various types of intrusion detection system such as Network Intrusion Detection System, Host-based Intrusion Detection System and Hybrid intrusion detection system are discussed in this paper. For the implementation MATLAB simulator will be used. MATLAB is a stage named as matrix laboratory which give the numerical registering in multi-design. It is likewise called as the fourth era dialect of programming. In this paper matlab and its working environment is also discussed. The author in this paper has given the idea of hybridization by using two algorithms named as Flower pollination Algorithm and Particle Swarm Optimization. The Flower pollination optimization algorithm (FPOA) is a recently invented optimization algorithm. It is inherited from the natural inspiration of pollination process. It mimics the process of flowering planets reproduction via pollination and Particle Swarm Optimization belongs to the field of Swarm Intelligence and Collective Intelligence and is a sub-field of Computational Intelligence. Particle Swarm Optimization is related to other Swarm Intelligence algorithms such as Ant Colony Optimization and it is a baseline algorithm for many variations, too numerous to list**.** In future, hybridization will be done using these two algorithms.

VI. REFERENCES

[1] Ravale, Ujwala, NileshMarathe, and Puja Padiya. "Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function." Procedia Computer Science 45 (2015): 428-435.

[2] Parvat, Thaksen J., and Pravin Chandra. "A Novel approach to deep packet inspection for intrusion detection." Procedia Computer Science 45 (2015): 506-513.

[3] Aggarwal, Preeti, and Sudhir Kumar Sharma. "Analysis of KDD dataset attributes-class wise for intrusion detection." Procedia Computer Science 57 (2015): 842-851.

[4] Pawar, Sunil Nilkanth, and RajankumarSadashivraoBichkar. "Genetic algorithm with variable length chromosomes for network intrusion detection." International Journal of Automation and Computing 12.3 (2015): 337-342.

[5] Karkouch, Aimad, et al. "Data quality in internet of things: A state-of-the-art survey." Journal of Network and Computer Applications 73 (2016): 57-81.

[6] Guo, Chun, et al. "A distance sum-based hybrid method for intrusion detection." Applied intelligence 40.1 (2014): 178-188.

[7] Bamakan, SeyedMojtabaHosseini, et al. "A new intrusion detection approach using PSO based multiple criteria linear programming." Procedia Computer Science 55 (2015): 231-237.

[8] Ghanem, Tamer F., Wail S. Elkilani, and Hatem M. Abdul-Kader. "A hybrid approach for efficient anomaly detection using metaheuristic methods." Journal of advanced research 6.4 (2015): 609-619.

[9] Lee, Gangin, and Unil Yun. "A new efficient approach for mining uncertain frequent patterns using minimum data structure without false positives." Future Generation Computer Systems 68 (2017): 89-110.

[10] Aljawarneh, Shadi A., Raja A. Moftah, and Abdelsalam M. Maatuk. "Investigations of automatic methods for detecting the polymorphic worms signatures." Future Generation Computer Systems 60 (2016): 67-77.

[11] Ahmad, Iftikhar, et al. "Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components." Neural Computing and Applications 24.7-8 (2014): 1671-1682.

[12] Airehrour, David, Jairo Gutierrez, and Sayan Kumar Ray. "Secure routing for internet of things: A survey." Journal of Network and Computer Applications66 (2016): 198-213.

[13] Yüksel, Buket, AlptekinKüpçü, and ÖznurÖzkasap. "Research issues for privacy and security of electronic health services." Future Generation Computer Systems 68 (2017): 1-13.

[14] Peng, Jian, Kim-Kwang Raymond Choo, and Helen Ashman. "User profiling in intrusion detection: A review." Journal of Network and Computer Applications 72 (2016): 14-27.

[15] Erfani, Sarah M., et al. "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning." Pattern Recognition 58 (2016): 121-134.