An Intelligent Feature Selection with Optimal Neural Network Based Network Intrusion Detection System for Cloud Environment

A. Thirumalairaj, M. Jeyakarthic

Abstract*: At present times, Cloud Computing (CC) becomes more familiar in several domains such as education, media, industries, government, and so on. On the other hand, uploading sensitive data to public cloud storage services involves diverse security issues, specifically integrity, availability and confidentiality to organizations/companies. Besides, the open and distributed (decentralized) structure of the cloud is highly prone to cyber attackers and intruders. Therefore, it is needed to design an intrusion detection system (IDS) for cloud environment to achieve high detection rate with low false alarm rate. The proposed model involves a binary grasshopper optimization algorithm with mutation (BGOA-M) as a feature selector to choose the optimal features. For classification, improved particle swarm optimization (IPSO) based NN model, called IPSO-NN has been derived. The significance of the IPSO-NN model is assessed using a set of two benchmark IDS dataset. The experimental results stated that the IPSO-NN model has achieved maximum accuracy values of 99.36% and 97.80% on the applied NSL-KDD 2015 and CICIDS 2017 dataset. The obtained experimental outcome clearly pointed out the extraordinary detection performance of the IPSO-NN model over the compared methods.

***Keywords*: Cloud computing, Intrusion, Detection, Feature Selection, Neural Network.**

I. INTRODUCTION

Nowadays, cloud computing (CC) has become a major revolution in the field of information technology (IT) with a rapid development of computing networks. The engaging attributes of CC follows various applications like government sectors, companies, academics, entertainment, and so on [1]. In general, CC is described by the National Institute of Standards and Technology (NIST) as a computing method which provides a satisfied, on-demand and network access for the distributed pool of network services by Internet to convince the computing demand of customers. Hence, the provided resources could be obtained rapidly and exhausted with lower handling effort communications [2]. NIST has deployed CC by assuming five major parameters like bandwidth, fast reliability, adaptable, on-demand service, resource pooling and several other factors. Besides, a set of three service delivering methods are also available such as software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) [3]. Followed by, the pay-as you-go and on-demand elastic process of cloud features are often modified, transforming on-premises structures to off premises data centers, applied through online and maintained with the help of cloud hosting providers [4].

Modeling CC is assumed to be various techniques along with ability for supporting a wider spectrum of applications. Actually, it has been aroused as a breakaway in applying the Internet. It is considered to be the investigation of offering different advantageous facilities with the help of Internet [5]. An alternative merit of CC is the use of reduced hardware units as the customers does not use highly powered hardware sources, frequent as well as quickly upgrading services, maximum potential for memory, global access of documents where the clients can retrieve the essential files by linking to online which might be parallel processing, resource sharing, simulation as well as time saving. But, few of the most challenging issues of CC are effectiveness, trust, privacy and accessibility, fault tolerance, fault recovery, and the expenditure of connecting bandwidth. As CC services are provided through Internet, data securities as well as privacy are assumed to be the main barriers in attaining effective CC and wider application firms and companies. Furthermore, open access as well as decentralized behavior of CC leads to a class of computation, vulnerable for cyber attacks and intrusions. NIST refers that, intrusion is an attempt to threaten the security criteria like privacy, integrity and availability or network security techniques. A major security problem involved in CC is to predict and avoid the network attacks as network is considered to be the main backbone of Cloud, where the chance of network getting affected has been improved. Subsequently, the risk factors become enhanced, that tends to emerge different threatening aspects, along with hackers to put more efforts for finding novel types of attacks. The significant inclusion of data security methods in past decades, intrusions as well as attacks are repeated to decompose the present intrusion detection systems (IDS) in Cloud platforms [6]. Hence, hackers deployed many fresh technologies with the potential to obtain complete Cloud environment. In recent days, a harmful distributed denial of service (DDoS) attack has reduced about 70 crucial facilities of Internet with Github, Twitter, Amazon, Paypal, and so forth. Several hackers applied the merits of CC as well as Internet of Things (IoT) to produce a massive number of attack traffic that is greater than 665 Gb/s [7, 8]. Furthermore,

Revised Manuscript Received on February 05, 2020.
***** Correspondence Author

A. Thirumalairaj*, Assistant Professor, Department of Computer Science, Kunthavai Naacchiyaar Govt Arts College for Women, Thanjavur. Email: a.thirumalairaj@gmail.com.

Dr. M. Jeyakarthic, Assistant Director (Academic), Tamil Virtual Academy, Chennai. Email: jeya_karthic@Yahoo.Com

Epsilon released millions of names as well as email addresses from the respective database, and from Stratfor in US, 75,000 credit card numbers and 860,000 usernames and passwords have been attacked [9]. In 2017, ransom ware hacking attacked several other banks, National Health Service hospitals in UK, bugger telecom firms, as well as natural gas organizations, whereas in 74 countries many types of security systems have been attacked [10]. Therefore, intrusion and attacking devices has become a promising issue of previous system. It is a real fact that a network Cloud IDS must investigate higher volumes of network traffic data to predict the novel attacks to meet greater accuracy measure. But, preprocessing, examining and forecasting intrusions in Cloud platforms with the application of conventional models are assumed to be more expense with respect to determination, duration as well as cost. Thus, effective IDS in Cloud platform need smarter methods like Machine Learning (ML) and computational intelligence techniques.

[11] projected a hybrid network IDS that merges signature relied prediction model and anomalous-dependent detecting model to find effective inner as well as physical attacks from cloud platforms. Snort has been applied in the form of signature-based IDS for predefined attacks with the help of rules database and obtained attacks database. [12] proposed a network IDS (NIDS) for Cloud atmosphere under the application of Multilayer Perceptron (MLP) as well as Particle Swarm Optimization Algorithm (PSO) to forecast intrusions and attacks. Here, PSO method is applied to identify the optimal weights and biases of neural network (NN) that undergoes training by equipped data and derived best weights. [13] projected a NIDS in CC platform based on Fuzzy C Means (FCM) technique for detecting intrusion actions from common nature, and, oppose network usage in Cloud resources as well as services from diverse types of threats and attacks. In order to ensure the cloud integrity, [14] deployed an anomalous-based ID with the integration of Cuckoo Optimization Feature Selection (COFS) and Naïve Bayes Algorithm (NBA). In the developed model, COFS is applied as q feature selection technique whereas NBA is employed as a classification framework.

This paper presents a new feature selection (FS) with optimal neural network (NN) based IDS model for cloud environment. The proposed model involves a binary grasshopper optimization algorithm with mutation (BGOA-M) as a feature selector to choose the optimal features. For classification, improved particle swarm optimization (IPSO) based NN model, called IPSO-NN has been derived. The significance of the IPSO-NN model is assessed using a set of two benchmark IDS dataset. The obtained experimental outcome clearly pointed out the extraordinary detection performance of the IPSO-NN model over the compared methods.

The remaining sections are arranged as follows. Section 2 elaborates the IPSO-NN model, validates it in Section 3, and concludes in Section 4.

II. THE PROPOSED IPSO-NN MODEL FOR IDS

The working procedure of IPSO-NN model for IDS is depicted clearly in Fig. 1. A set of four main processes are involved namely, FS, preprocessing, detection and classification, and alarm system. At the earlier stage, BGOA-M model is applied to extract the useful features from the input data. Then, the extracted features undergo preprocessing which takes place at two levels namely categorical value conversion and data normalization. Afterwards, the IPSO-NN model is executed to train the dataset for classifying the presence of intrusion in the preprocessed data. Upon providing testing data into the presented IDS, the FS process and preprocessing step takes place. Then, testing of data is carried out to identify the presence of intrusion. Finally, an alarm will be generated to indicate the presence of intrusion. These processes will be clearly discussed in the following subsections.

A. Feature selection (FS) using BGOA-M

The operation of FS in IDS can be considered as a combinatorial optimization issue. BGOA method has been used to obtain efficient FS work to be processed in IDS. GOA shows the best outcome to resolve the optimization issue. It is a competitive one in terms of exploitation, exploration, avoid from local optimum as well as converging measure. The dynamic behaviour of comfort area coefficient tends to powerful trade-off between exploitation as well as exploration. In wrapper-based models, identifying optimized features from FS is considered to be a most significant task. This happens as the selected subset needs a validation process by applying learning methodologies at each optimization step. Hence, an adaptive optimizing technique should reduce the verification strategy. There are few advantages in GOA that makes the FS process as a suitable one. Based on the behaviour of FS problem, the searching space can be found as binary measures [0, 1] and binary operator is simpler when compared with alternate models. Therefore, a binary variant of GOA is named as BGOA have been projected to solve the problem existed in FS. While in continuous variant of GOA, all solutions are based on the current position, location of best grasshoppers or target, and position of alternate grasshoppers.

The currently presented solutions are formulated by contributing the step for position vector. Simultaneously, from binary space, add operator could be applied in the form of position vector with 0’s and 1’s. Finally, the presented model applies 2 methods which simulate to convert the frequent variations of GOA to binary metrics. Followed by, each grasshopper changes the corresponding location by considering the position of grasshoppers on the basis of optimized solution attained. The motivation from current grasshoppers and final target can be explained with the help of stochastic mutation. The mutation value is differed according to the step vector. Therefore, positive components assume the measures of same components of targets; otherwise, grasshoppers are identified from the search space as well as from random assumption of component relied on likelihood of maximum value in the following.

$$x_{t+1}^{i} = \begin{cases} Target_t^i & \Delta x_t^i \geq 0 \\ \begin{cases} 1 & r_3 \geq 0.5 \\ 0 & r_3 < 0.5 \end{cases} & \Delta x_t^i < 0 \end{cases} \quad (1)$$

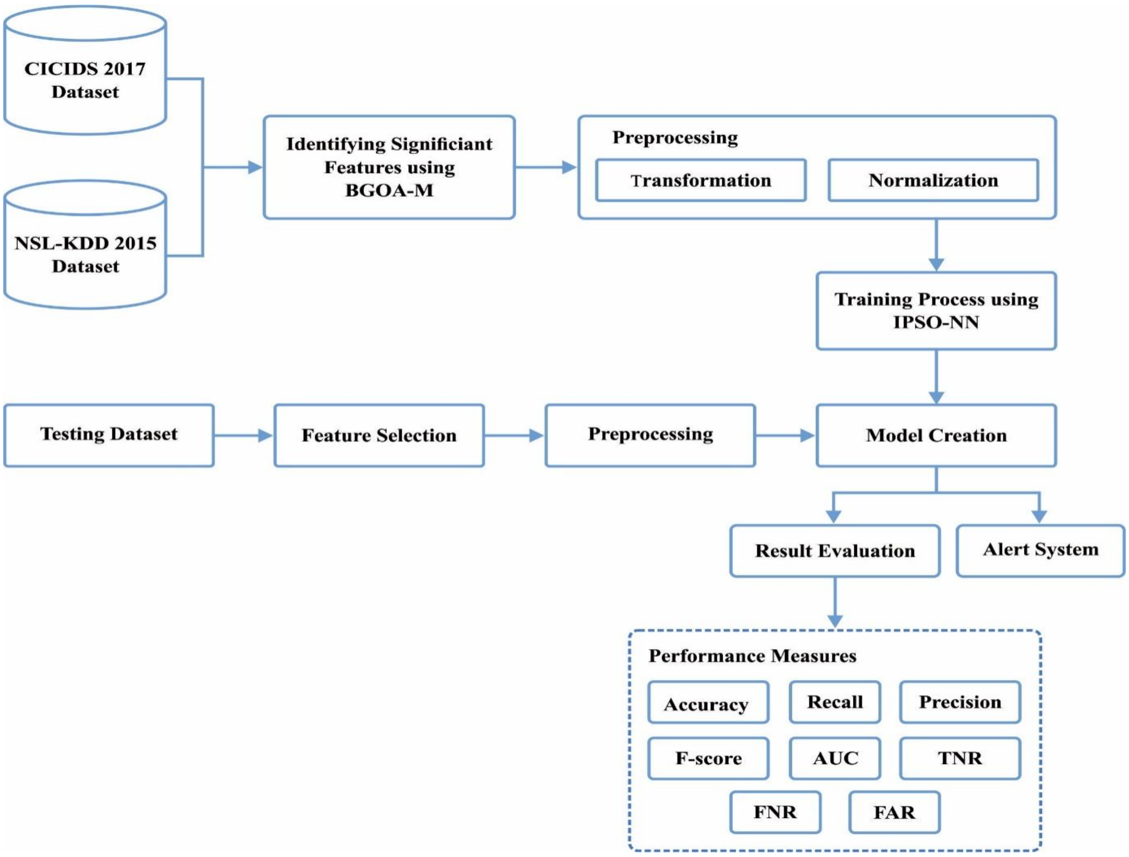


Fig. 1. Overall process of IPSO-NN model

iterations, $\varDelta K_t^i$ implies the measure of i th dimension from step vector and r_3 which is the random value exists from (0, 1). The initial constraint signifies exploitation and consecutive 2 conditions operate the unwanted alterations from solution which indicates the exploration of different areas of search space. Then, the produced solution has been upgraded by processing the mutation as local search with novel positions along with proper mutation value.

Here, the solution has to be migrated to best solutions. To eliminate the initial convergence, a mutation operator by proper mutation value can be applied to improvise the diversity which is termed as BGOA-M. The mutation values are critical feature that handles mutation operator as higher mutation rate increases the viability of identifying massive areas in search space, however, removes the convergence along with optimal solution. Simultaneously, the application of minimum mutation value tends to process predefined convergence. The mutation value *r* applied in this work is given in Eq. (2). The variable r is reduced linearly from 0.9 to 0 on the basis of iteration value n..

$$r = 0.9 + \frac{-0.9 * (n-1)}{Max\_Round\ -1} \tag{2}$$

B. Data pre-processing

Once the features were selection, preprocessing of data takes place, which generally involves two levels namely, data conversion or categorical encoding and Normalization. The “Categorical encoding” is defined a task of declaring arithmetic values for non-numeric attributes to build easier process, since numeric data can be handled easily. “Normalization” is the scaling process of feature measures to minimum range which attempts to get optimal prediction outcome and removes mathematical complexities while calculation is carried out. The data pre-processing system applies Min-Max and statistical normalization techniques. The key objective of this model is to create records for training as well as testing subsets obtained from 2 IDS dataset.

1. **Categorical encoding**: It defines the task of declaring arithmetic values for non-numeric features that tends to create an elegant computation and to handle the process easily.
2. **Normalization:** Any parameter with maximum values could dominate the outcome of the attributes with minimum values. This dominance might be decreased by using normalization process that is scaling of values inside an assertive range. It refers the task of enveloping parameters to a particular range to reduce the complications in managing data spread than accurate range as well as type of measures. Therefore, it has to be reduced into definite range for appropriate computation and effective data examination. In order to normalize data, the mean-range [0, 1] and statistical normalization techniques are applied [15].
3. **Mean range [0, 1] (Min-Max normalization):** As showcased in Eq. (3), the mean range method helps to normalize a parameter metric by decreasing the lower value of corresponding attribute from present value.

4. It is again classified by the differences among higher and lower values of respective features.

$$X' = \frac{x - MinA}{MaxA - MinA} \qquad (3)$$

where x and x' are the values that should undergo normalization. $MinA$ and $MaxA$ are assumed to be lower as well as higher viable values of attribute A in prior to conducting normalization process. Then, Statistical normalization or Z-score normalization could be expressed as:

$$X' = \frac{x - \mu}{\alpha} \qquad (4)$$

The value x of parameter A is converted in X', μ denotes the mean and α is SD of provided attribute.

C. Detection and classification model

1. NN

It is a soft computation approach mainly used for processing data. It is inspired from the working of human nervous system. It can be defined based on weighted directed graphs to acts as artificial neurons and directed boundaries between the neurons weights. It undergoes classification into two kinds namely Feedforward and Recurrent networks. The former one is stationary and the latter one is non-stationary. It generates a set of resultant values with the sequence of values from a given input. Feedforward networks are memory-less and self-determining on the state of the proceeding network. An effective input design is developed for calculating the output of the neurons. At the earlier stage, the input to each neuron enters into a novel state. It is noticed that many layers of NN has been efficiently implemented in decision support systems. At each input layer, every individual neuron in other layers acts as a processing element using non-linear activation function. The criteria of NN is that, when data gets accessible at the input layer, then NN determines the value at the subsequent layer until the resultant value is obtained at every resultant neuron. The outcome of NN holds appropriate class label for the provided input data. Every neuron in both the input as well as hidden layers is linked to one another of the subsequent layer with few weight values.

The neurons exist in the hidden layer is accountable for determining the cumulative weights of all inputs and include a threshold. The input layer represents the number of attributes present in the provided dataset. The operation of hidden layer indicates the features in the dataset are non-linearly partitioned and the output layer offers the needed outcome. The threshold node is besides additional in input layer which offers the weight function. The total outcome is applied for achieving the performance of the neurons with the execution of the sigmoid activation function. It can be computed using Eq. (5).

$$p_j = \sum_{i=1}^{n} w_{j,i} x_i + \theta_j, m_j = f_j(p_j) \qquad (5)$$

where p_j is the linear arrangement of inputs $x_1, x_2, \ldots, x_n$, and threshold θ_j, $w_{j,i}$ is the linked weighs among the input x_i, and the neurons j, and f_j is the activation function of the jth neurons, and m_j is result. The sigmoid function is applied as an activation function as given in Eq. (6).

$$f(t) = \frac{1}{1 + e^{-t}} \qquad (6)$$

For training a NN, backpropagation models are applied. Each weight vector (w) is determined using small random values by the use of pseudorandom sequence generator. But, these methods get into several phases to train the network, and it changed the weights as calculated at every phase. To defeat the aforementioned issues, the IPSO algorithm is used to calculate the best value of the weight and threshold functions as it has the ability to define the equivalent weights and identifying the better results.

2. IPSO algorithm

Assume that the particle transformed by 1D and δ. Then, the related location (x) can be measured with the given functions:

$$\mathrm{x} = p \pm \frac{L}{2} \ln\left(\frac{1}{u}\right) \qquad (7)$$

where p implies the particle motion centre. In IPSO technique, it is named as attractor of a particle, L denotes the feature length and δ is the corresponding measure has been directly compared to convergence speed as well as searching potential of this model. u denotes the random value with an even distribution function from the range $(0,1)$. The parameter L has to be properly computed by IPSO model while this attribute can be measured from the given expressions:

$$L_{\mathrm{i,j}} = 2\beta \cdot \left\| mbest_j - x_{i,j} \right\| \qquad (8)$$

where

$$mbest = \frac{1}{N} \sum_{i=1}^{N} pbest_i \qquad (9)$$

where $pbest_i$ represents the best location of an individual from search history of a particle x_i and β is a Contraction-Expansion (CE) factor [16]. These parameters have to be minimized at the time of implementing this method.

In IPSO algorithm, every particle consumes the weighted maximum location of single historical best location as well as good position of group history as the respective attraction point. This estimation can result in particle motion trajectory outcomes. Though these calculations are elegant, it is composed with 2 disadvantages: First, every particle location is based on the historical optimized position of a group. It results in accelerated declination in the diversity of higher groups that minimizes the methods potential to solve difficult multi-peak optimizing issues. Secondly, feasible distribution space of all particles attraction point is minimized at the time of evolution task of any model. Therefore, the particles are restricted for a rectangle along with vertices $pbest_{i,t}$ and$gbest_{i,t}$. The $attractor_{i,t}$ seeks for $best_{i,t}$. As a result, this model does not shift from local optima present in the target level. Then, it is concluded with

$$attractor_{i,t} = u_{\mathrm{i,t}}\, pbest_{i,t} + \left(1 - u_{\mathrm{i,t}}\right) pbest_{b,t} + \Delta_{\mathrm{i},t} \qquad (10)$$

where $u_{i,t}$ represents an arbitrary value along with even distribution function within the range $[0, 1]$. The subscript i is said to be the value of arbitrarily chosen particle including optimal fitness measure. Furthermore, the range of a particle is chosen as $m \in (0,1]$.

$$\Delta_{\mathrm{i},t} = \left\{ \Delta_{i,t}^{1}, \Delta_{i,t}^{2}, \ldots, \Delta_{i,t}^{D} \right\}$$

is said to be a perturbation

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

vector expressed as

$$\Delta_{i,t} = \frac{pbest_{a,t} - pbest_{c,t}}{2} \qquad (11)$$

where subscripts b and c are termed as randomly chosen particles from the group as well as a ≠ b ≠ c ≠ i. The update function for particle location in IPSO model can be presented. Assume that

$$x_{i,t} = attractor_{i,t} \pm 2\beta\|mbest_j - x_{i,t}\| \qquad (12)$$

From the evolved strategy, only few data are considered to be more applicable regarding single particles as well as global best location might be lost by the model. Also, the motion of few attractors of ineffective direction tends in worst fitness in the consecutive process. Hence, to enhance the function of this model, efficient data on single as well as global best positions of particle might be applied by proper technique. In order to enhance the optimizing capability of a model with the application of defined data, the cross-over methodology and local searching has been combined into cross-sequential quadratic programming (SQP) approach. This model comes under the local optima in last stage. It refers that individual as well as global optimal locations of a particles from population are nearby with one another at the same time. By assuming the defined problem, Gaussian chaotic mutation operators have been projected to enhance the population diversity and shift out of local optima.

3. IPSO-NN

There are few major steps in IPSO based parameter optimizing task that has been consolidate in the following steps:

Step 1: Initialization

In this step, diverse attributes of IPSO has been mentioned with a population of randomly selected particles as well as velocities.

Step 2: Execute NN model and compute the fitness function (FF)

Here, NN method undergoes training along with the attributes c and r that has been included in the current location. The ten-fold cross-validation model is used to estimate the FF value. In this method, training data sets are classified into 10 mutually unique subsets of same size, where 9 subsets are applied for training the data and final subset is employed for test data. The predefined strategy is followed more than 5 times as every subset is utilized for testing. In addition, FF is described as 1-$CA_{\text{validation}}$ of the ten-fold cross-validation technology in the training data set that is depicted in Eqs. (13) and (14). Furthermore, any solution with higher $CA_{\text{validation}}$ is comprised with minimum FF measure.

$$Fitness = 1 - CA_{\text{validation}} \qquad (13)$$

$$CA_{\text{validation}} = 1 - \frac{1}{10}\sum_{i=1}^{10}\left|\frac{y_c}{y_c + y_f}\right| \times 100 \qquad (14)$$

Where, $y_c$and $y_f$are the number of true and false classifiers.

Step 3: Upgrade the global as well as local optimal positions

In the above-mentioned step, the global best as well as personal best locations of particles are upgraded on the basis of FF metrics.

Step 4: Update velocity and position

The position and velocity of every particle is upgraded with the application and attained a novel position for all particles to proceed with upcoming computation.

Step 5: Termination Condition

Follow the steps 2-4 till the termination condition has been attained.

4. Alert System

Once the model is passed through the optimization phase and exploring the best attributes to create effective IDS, the proposed IPSO-NN functions in operation mode, to detect the class of provided packets obtained from the dataset. For ordinary sample, it has the permission to apply for Cloud infrastructure, otherwise it has been remained without saving and alert system is pointed. The proposed method produces alerts regarding the attacks which has been computed with the help of optimized prediction technique on the basis of IPSO-NN, and forwards to authorized security personnel in order to provide warning and future analysis.

5. Deployment of IPSO-NNN model in cloud environment

The intention of the presented IPSO-NN model is to identify the intruders as well as doubtful actions present in and around the CC platform by observing the network traffic, and managing confidentiality, availability, integrity and efficiency of cloud resources and services. It enables to detect and halt attacks in real time impairing the security of the Cloud Data Center. The presented model can be deployed in two different places and is depicted in Fig. 2.

- **Cloud Front-end:** Deploying IPSO-NN on front end of Cloud assist to identify the intrusions and attacks in the network from external resources, executed by zombie hosts or by attacker linked to Internet who tried to bypass the firewall for accessing the internal cloud, which could be a private one. Consequently, IPSO-NN acts as a second line of defence at the back of firewall for resolving its constraints, and plays as an extra defensive layer of security.
- **Cloud Back-end:** Deploying IPSO-NN model on processing servers placed at back end of Cloud assist to identify intrusions appearing on its internal network. In a virtual platform, we several virtual machines (VMs) exist on the identical physical server, and they can inter-communicate using virtual switches without going away from the physical server. Consequently, network security gadgets on local areanetwork (LAN) could not observe it in the network traffic; when the traffic do not require passing into security appliances mainly a firewall, consequently, a ambiguity for every type of security attack will be opened. Hence, the origin of an attacker/hacker is negotiating a single VM and utilizing it as a springboard for controlling other VMs in the identical hypervisor. It is usually carried out with no monitoring, providing a hacker a massive hacking area. Besides, the virtual platform is vulnerable to diverse threats and risks,

- focused mainly on the hypervisor. The IPSO-NN has been developed for observing the virtual traffic, and also the flow of traffic from or to the processing server on the physical network.

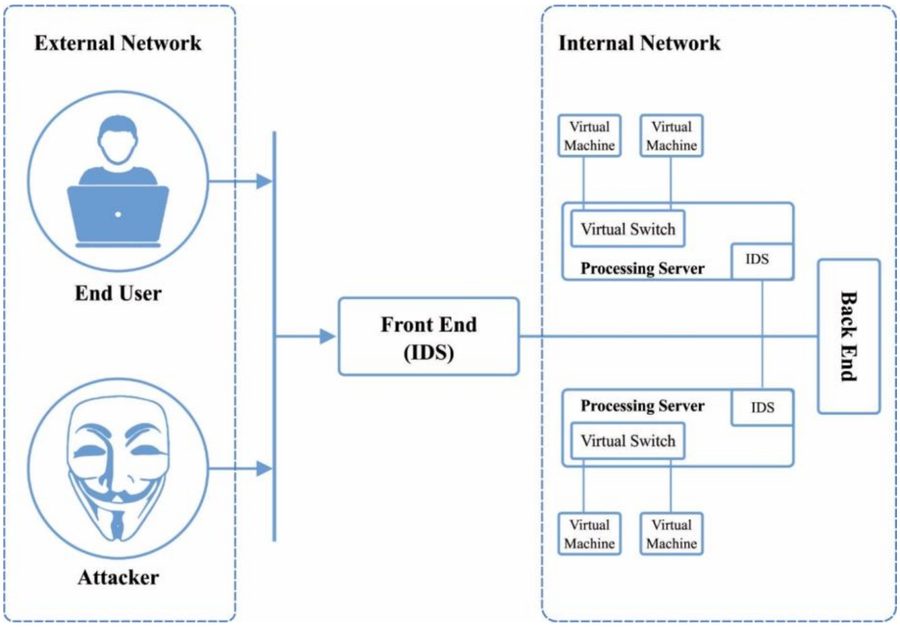


Fig. 2. Framework of IPSO-NN model in Cloud Environment

III. PERFORMANCE VALIDATION

A. Dataset description

The performance of the proposed model can be assessed by the use of two benchmark dataset namely NSL-KDD 2015 [17] and CICIDS 2017 [18]. The former NSL-KDD 2015 dataset holds a total of 125973 instances including a set of 41 attributes. It holds a set of two classes namely normal and anomaly, where a set of 67343 instances falls into the category of Normal and the remaining 58630 instances falls into the category of Anomaly. The attribute details of the NSL-KDD 2015 dataset is depicted in Fig. 3. The latter CICIDS 2017 dataset holds a total of 2830743 instances including a set of 80 attributes. It holds a set of two classes namely normal and anomaly, where a set of 2273097 instances falls into the category of Normal and the remaining 557646 instances falls into the category of Anomaly. The attribute details of the CICIDS 2017 dataset is depicted in Fig. 4.

Table 1 Dataset Description

Dataset	No. of instances	No. of attributes	No. of classes	Normal/Anomaly
NSL-KDD 2015	125973	41	2	67343/58630
CICIDS 2017	2830743	80	2	2273097/557646

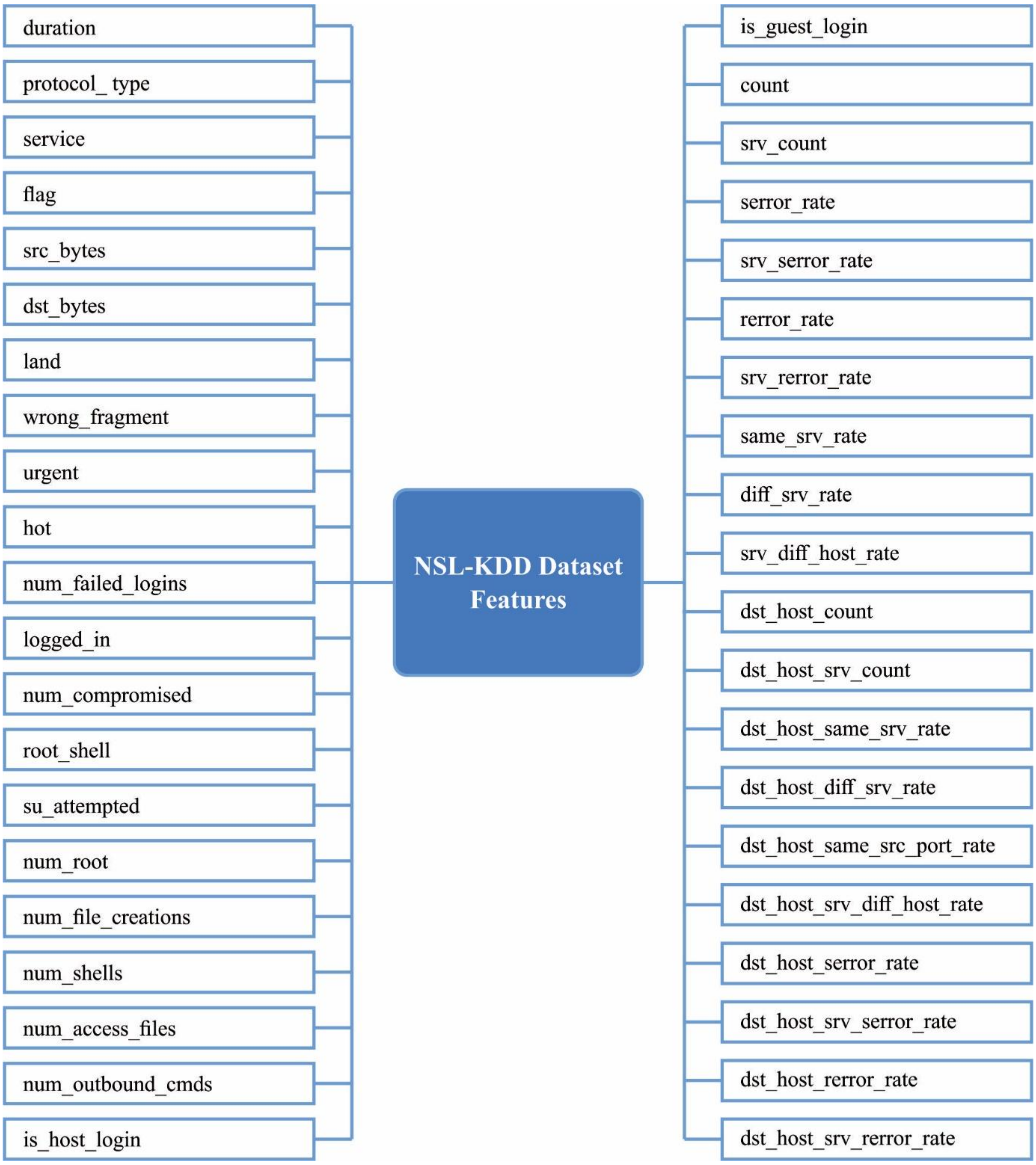
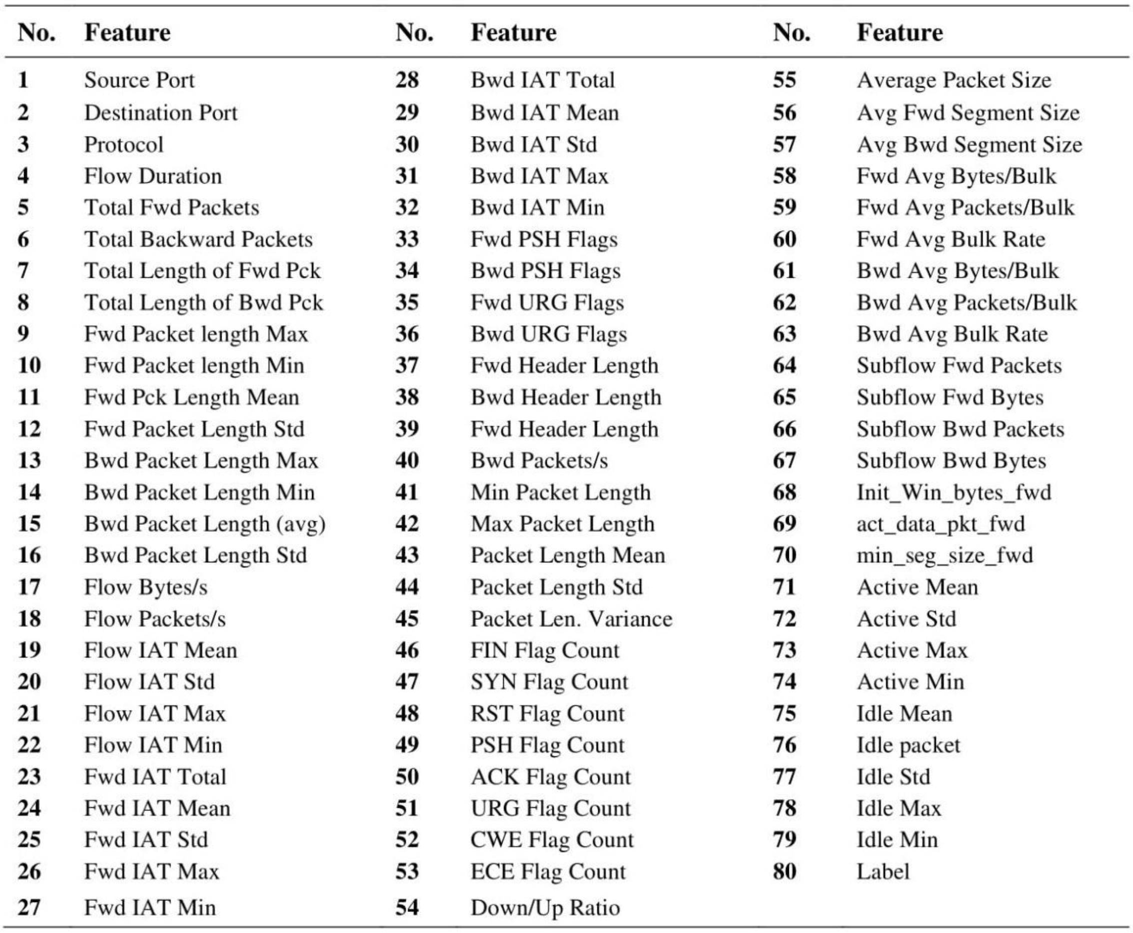


Fig. 3. Features in NSL-KDD 2015 Dataset



No.	Feature	No.	Feature	No.	Feature
1	Source Port	28	Bwd IAT Total	55	Average Packet Size
2	Destination Port	29	Bwd IAT Mean	56	Avg Fwd Segment Size
3	Protocol	30	Bwd IAT Std	57	Avg Bwd Segment Size
4	Flow Duration	31	Bwd IAT Max	58	Fwd Avg Bytes/Bulk
5	Total Fwd Packets	32	Bwd IAT Min	59	Fwd Avg Packets/Bulk
6	Total Backward Packets	33	Fwd PSH Flags	60	Fwd Avg Bulk Rate
7	Total Length of Fwd Pck	34	Bwd PSH Flags	61	Bwd Avg Bytes/Bulk
8	Total Length of Bwd Pck	35	Fwd URG Flags	62	Bwd Avg Packets/Bulk
9	Fwd Packet length Max	36	Bwd URG Flags	63	Bwd Avg Bulk Rate
10	Fwd Packet length Min	37	Fwd Header Length	64	Subflow Fwd Packets
11	Fwd Pck Length Mean	38	Bwd Header Length	65	Subflow Fwd Bytes
12	Fwd Packet Length Std	39	Fwd Header Length	66	Subflow Bwd Packets
13	Bwd Packet Length Max	40	Bwd Packets/s	67	Subflow Bwd Bytes
14	Bwd Packet Length Min	41	Min Packet Length	68	Init_Win_bytes_fwd
15	Bwd Packet Length (avg)	42	Max Packet Length	69	act_data_pkt_fwd
16	Bwd Packet Length Std	43	Packet Length Mean	70	min_seg_size_fwd
17	Flow Bytes/s	44	Packet Length Std	71	Active Mean
18	Flow Packets/s	45	Packet Len. Variance	72	Active Std
19	Flow IAT Mean	46	FIN Flag Count	73	Active Max
20	Flow IAT Std	47	SYN Flag Count	74	Active Min
21	Flow IAT Max	48	RST Flag Count	75	Idle Mean
22	Flow IAT Min	49	PSH Flag Count	76	Idle packet
23	Fwd IAT Total	50	ACK Flag Count	77	Idle Std
24	Fwd IAT Mean	51	URG Flag Count	78	Idle Max
25	Fwd IAT Std	52	CWE Flag Count	79	Idle Min
26	Fwd IAT Max	53	ECE Flag Count	80	Label
27	Fwd IAT Min	54	Down/Up Ratio		

Fig. 4. Features in CICIDS 2017 Dataset

B. FS results

Table 2 shows the FS results offered by the BGOA-M model on the applied two dataset. the table values indicated that the BGOA-M model has chosen a set of 18 features out of 41 features with the minimal best cost of 0.003289 under the NSL-KDD 2015 dataset and has selected a collection of 26 features out of 80 features with the least cost of 0.005632 under the CICIDS 2017 dataset.

Table 2 Selected features of BGOA-M for applied dataset

Methods	Best Cost	Selected Features
NSL-KDD 2015	0.003289	1, 24, 36, 40, 22, 37, 32, 28,25, 34,3, 23, 20,7,26,30,12,11
CICIDS 2017	0.005632	2,7,9,10,11,13,14,15,16,18,22,26,29,31,33,36,39,40,45,48,53,58,62,64,72,78

C. Results analysis

Table 3 provides the confusion matrix generated by the presented IPSO-NN model on the applied two dataset. On the applied NSL-KDD 2015 dataset, it is noted that the IPSO-NN model has classified the instances with the True Positive rate of 67044 instances, True Negative rate of 58133 instances, False Positive rate of 299 instances and False Negative rate of 497 instances. Similarly, on the applied CICIDS 2017 dataset, it is noted that the IPSO-NN model has classified the instances with the True Positive rate of 2217574 instances, True Negative rate of 551099 instances, False Positive rate of 55523 instances and False Negative rate of 6547 instances.

Table 3 Confusion Matrix of Applied Dataset using Proposed IPSO-NN

Dataset	True Positive	True Negative	False Positive	False Negative
NSL-KDD 2015	67044	58133	299	497
CICIDS 2017	2217574	551099	55523	6547

For experimentation, a set of different measures namely false acceptance rate (FAR), true negative rate (TNR), false negative rate (FNR), area under curve (AUC), precision, recall, accuracy and F-score are employed. Table 4 and Fig. 5 show the results offered by the proposed IPSO-NN model on the applied two dataset. On measuring the detection performance on the NSL-KDD 2015 dataset, it is noticed that the IPSO-NN model has achieved optimal detection performance by offering a minimum FAR and FNR values of

0.005 and 0.008. Along with that, a maximum TNR of 99.48%, AUC of 99.37%, precision of 99.56%, recall of 99.26%, accuracy of 99.36% and F-score of 99.40% has been offered by the IPSO-NN model on the NSL-KDD 2015 dataset. Similarly, under the applied CICIDS 2017 dataset, it is noticed that the IPSO-NN model has achieved optimal detection performance by offering a minimum FAR and FNR values of 0.091 and 0.012. Along with that, a maximum TNR of 90.84%, AUC of 95.27%, precision of 97.55%, recall of 99.70%, accuracy of 97.80% and F-score of 98.61% has been offered by the IPSO-NN model on the NSL-KDD 2015 dataset. These values ensured the optimal classification performance of the proposed model on all the applied dataset.

Table 4 Intrusion Detection performance of IPSO-NN model

Measures	NSL-KDD 2015	CICIDS 2017
FAR	0.005	0.091
TNR	99.48	90.84
FNR	0.008	0.012
AUC	99.37	95.27
Precision	99.56	97.55
Recall	99.26	99.70
Accuracy	99.36	97.80
F-score	99.40	98.61

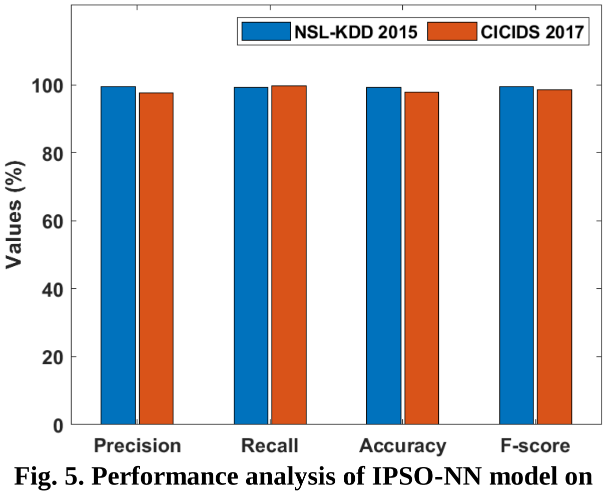


Fig. 5. Performance analysis of IPSO-NN model on applied two IDS dataset

For further validating the effective intrusion detection performance of the IPSO-NN model, a detailed comparative analysis is made with the recently presented models namely Cuckoo optimization, cuckoo search with PSO (CS-PSO), PSO-SVM, Behaviour Based IDS, Gaussian Process, Deep Neural Network with SVM, GA+Fuzzy, Fuzzy C-means and Gradient Boosting models interms of accuracy. The resultant values are tabulated in Table 5 and also shown in Fig. 6. By looking into the table, it can be easily noticed that the CS-PSO model has offered least detection performance by offering a minimum accuracy of 75.51%. Then, it is observable that the Gradient Boosting model has attained an accuracy of 84.25%, which is higher than the accuracy offered by the CS-PSO algorithm. However, the Gaussian Process and the DNN+SVM model outperforms the earlier methods by offering high as well as closer accuracy values of 91.06% and 92.03% respectively. Next to that, even higher detection performance is exhibited by Fuzzy C-means model by offering an accuracy value of 95.30%. Simultaneously, the GA+ Fuzzy and Cuckoo Optimization algorithms have accomplished manageable and identical detection results by offering accuracy values of 96.53% and 96.888% respectively. In line with, even higher detection outcome is achieved by Behaviour Based IDS model which can be noticed from the accuracy value of 98.89% whereas competitive results of 99.10% accuracy is provided by the existing PSO-SVM model. But, it is interesting that the IPSO-SVM model has outperformed all the existing methods and achieved a maximum accuracy of 99.36% on the applied dataset.

Table 5 Accuracy analysis of IPSO-NN with recently proposed models

Methods	Accuracy
Proposed (IPSO-NN)	**99.36**
Cuckoo Optimization	96.88
CS-PSO (2019)	75.51
PSO-SVM (2019)	99.10
Behaviour Based IDS (2019)	98.89
Gaussian Process (2015)	91.06
DNN+SVM (2018)	92.03
GA+ Fuzzy (2018)	96.53
Fuzzy C-Means (2018)	95.30
Gradient Boosting (2018)	84.25

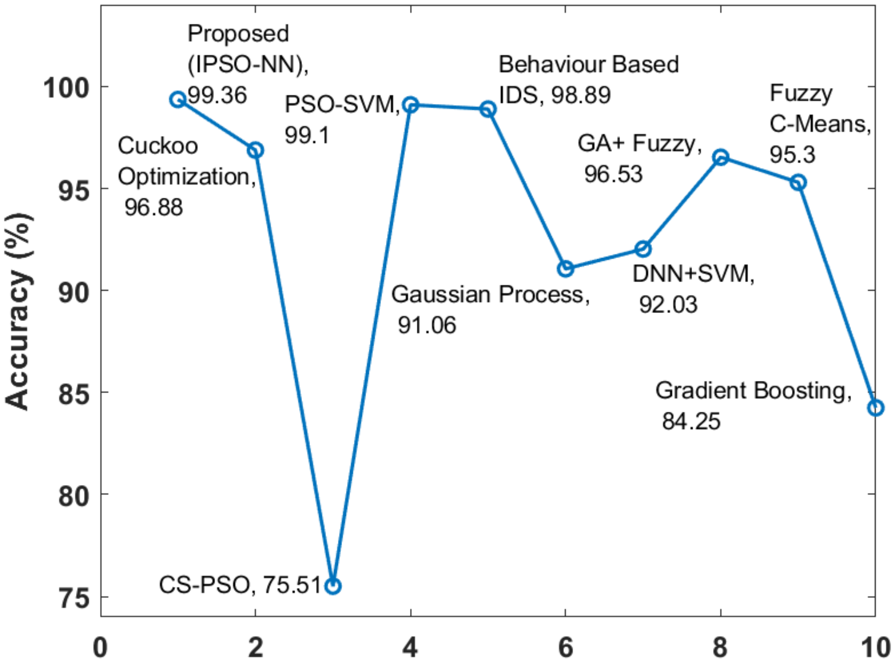


Fig. 6. Comparative analysis of IPSO-NN with recently proposed models

IV. CONCLUSION

This paper has introduced a new FS with optimal NN based IDS model for cloud environment to achieve high detection rate with low false alarm rate. The presented IPSO-NN model has operated on four basic steps namely BGOA-M based feature extraction, preprocessing, IPSO-NN based detection and classification and alarm generation. The effectiveness of the presented IPSO-NN model has been validated using a set of benchmark dataset. By looking into the tables and figures, it is verified that the IPSO-NN model has achieved maximum accuracy values of 99.36% and 97.80% on the applied NSL-KDD 2015 and CICIDS 2017 dataset. The simulation outcome strongly pointed out that the IPSO-NN model has offered extremely high intrusion detection performance over the compared models. In future, the performance of the proposed model can be improved by the use of deep learning models.

REFERENCES

1. Fernandes DA, Soares LF, Gomes JV, Freire MM, Inácio PR. Security issues in cloud environments: a survey. Int J Inf Secur 2014;13(2):113–70 https://doi.org/10.1007/s10207-013-0208-7.
2. Mell P, Grance T. (2011). The NIST definition of cloud computing. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf. Mera C, Branch JW. A survey on class imbalance learning on automatic visual inspection. IEEE Lat Am Trans 2014;12(4):657–67.
3. Brunette G, Mogull R, et al. Security guidance for critical areas of focus in cloud computing v2.1. Cloud Secure Alliance. 2009:1–76.
4. Idhammad M, Afdel K, Belouch M. Distributed intrusion detection system for cloud environments based on data mining techniques. Procedia Comput Sci 2018;127:35–41 https://doi.org/10.1016/j.procs.2018.01.095.
5. Ghosh P, Jha S, Dutta R, Phadikar S. Intrusion detection system based on BCS-GA in cloud environment. In: Proceedings of international conference on emerging research in computing, information, communication and applications. Singapore: Springer; 2016. p. 393–403. https://doi.org/10.1007/978-981-10-4741-1_35.
6. Iqbal S, Kiah MLM, Dhaghighi B, Hussain M, Khan S, Khan MK, Choo KKR. On cloud security attacks: a taxonomy and intrusion detection and prevention as a service. J Netw Comput Appl 2016;74:98–120 https://doi.org/10.1016/j.jnca.2016.08.016
7. DDoS. (2019) "DDoS attack that disrupted internet was largest of its kind in history, experts say". https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet.
8. Wikipedia, 2016dyn cyberattack, https://en.wikipedia.org/wiki/2016_Dyn_cyberattack.
9. Chou TS. Security threats on cloud computing vulnerabilities. Int J Comput Sci Inf Technol 2013;5(3):79.
10. Ismael Valenzuela. (2019) GSE #132 – Global Director, Foundstone Consulting Services. "Targeted ransomware attacks in the cloud". [Online]. Available:https://files.sans.org/summit/healthcare2016/PDFs/Prediction-2017-I-Survived-a-Ransomware-Attack-in-myCloud-Ismael-Valenzuela.pdf. Accessed 20 January 2019.
11. Hatef MA, Shaker V, Jabbarpour MR, Jung J, Zarrabi H. HIDCC: a hybrid intrusion detection approach in cloud computing. Concurr Comput: Pract Exp 2018;30(3):e4171 https://doi.org/10.1002/cpe.4171.
12. Saljoughi AS, Mehrvarz M, Mirvaziri H. Attacks and intrusion detection in cloud computing using neural networks and particle swarm optimization algorithms. Emerg Sci J 2018;1(4):179–91.
13. Mehibs SM, Hashim SH. Proposed network intrusion detection system in cloud environment based on back propagation neural network. J Univ Babylon Pure Appl Sci 2018;26(1):29–40.
14. Singh DAAG, Priyadharshini R, Leavline EJ. Cuckoo optimisation based intrusion detection system for cloud computing. Int J Comput Netw Inf Secur 2018;10(11):42. doi:10.5815/ijcnis.2018.11.05.
15. Chiba, Z., Abghour, N., Moussaid, K. and Rida, M., 2019. Intelligent approach to build a Deep Neural Network based IDS for cloud environment using combination of machine learning algorithms. Computers & Security, 86, pp.291-317.
16. Liu, G., Chen, W., Chen, H. and Xie, J., 2019. A Quantum Particle Swarm Optimization Algorithm with Teamwork Evolutionary Strategy. Mathematical Problems in Engineering, 2019.
17. CICIDS2017 data set. (2019), ttps://www.unb.ca/cic/datasets/ids-2017.html
18. NSL-KDD. (2019) Dataset of NSL-KDD University of new Brunswick. http://www.unb.ca/research/iscx/dataset/iscx-NSL-KDD-dataset.html.