International Journal of Research in Advent Technology (IJRAT), Special Issue, March 2019
E-ISSN: 2321-9637
ICCCMIT 2019 organised by M.O.P. Vaishnav College for Women (Autonomous)
Chennai-34, India
Available online at www.ijrat.org

An Quantum Based Efficient Algorithm for Qualitative and Quantitative Anomaly Detection In Network

J.DILLIBABU.[1] DR.K.NIRMAL[2]
Assistant Professor, Dept. of computer Science, [1]
St. Thomas College of Arts and Science, Chennai, India
Associate Professor, Dept of Computer Science, [2]
Quaid-E-Millath Government College for Women, Chennai, India
Email:jdbabumca@rediffmail.com,nimimca@gmail.com

Abstract: Networking Environment is challenging hybrid information's are transferred through some type of applications by means of different request at the time of Request processing multiple no of data's are hitting a single application or different application with processing time, here the issue is arising if the data range is extended. Due to the currently high volume of networks traffic in addition to the increased number of attacks and their dynamic properties, NIDSs have the challenge of improving their classification performance. A time quantum based immune clonal algorithm with the estimation of distribution algorithm is proposed in this paper to build a new IDS simultaneously data validation is also done for Qualitative and Quantitative data setis used for classification methodology algorithm of the new IDS where it is trained and tested using the knowledge discovery data set.

Keywords: **Distribution Algorithm, Intrusions Detection System, time Quantum , Optimization Algorithm.**

1. INTRODUCTION

Intrusion Detection Systems (IDSs) are the systems responsible to identify deviated activities by monitoring system behavior of singleusers or networked environmental. They detect the abnormal behavior or unwanted traffic and take the appropriate responsible against such activities. [27]. There are two broad types of intrusion detection approaches already existing here collectively referred as misuse and anomaly detection approaches. Misuse detection uses known patterns, signatures, login procedures,etc., of unauthorized behavior to detect intrusions. It is quite strong to detect known intrusions but it has low degree of accuracy in detecting unknown intrusions. Since it already exists on signatures extractedby authorized user experts. Anomaly detection establishes a baseline of normal usage patterns and if it find something that is widely deviated from the boundary lines as we expected, the deviations are agged possible intrusion. Although it is powerful to identify new types of intrusion as deviations from normal usage, a potential drawback is the false alarm rating is too high comparing to previously unseen system behaviors may be recognized as anomalies, and hence flagged as high potential intrusions [15]. IDSs can also be categorized to host-based and network-based environmental data's that is depending on the audit data which they took for analysis. Host based systems validate data's of users parallely how much

files were accessed and what are the applications were hitted to complete their execution. Network-based IDS (NIDS) examines data as packets of information exchange through the networks. The goal of NIDSs is to quickly and accurately recognize and distinguish the normal and abnormal activity which is affecting network connections. They always seek to have a high intrusions detection rate and a low false alarms rate, detecting normal connections as intrusions, to ensure high classification accuracy. There are some challenges facing systems which makes it difficult to achieve such a goal and keep tight security of networks. Computational Intelligence techniques, BIOs, known for their ability to adapt and to exhibit fault tolerance, high computational speed and persistence against noisy information, compensate for the limitations of these two approaches [30].

Many BIOs have been applied as classification techniques for NIDSs, like the artificial immune system (AIS), Artificial Neural Networks (ANNs), Particle Swarm Optimization (PSO) and many defined techniques. They were able to easily and automatically extract the discrimination rules of

normal or abnormal behavior from the large scaled networks logs [7]. Also the Quantum Inspired EvolutionaryAlgorithms (QIEAs) have been used to build NIDSs.

QIEAs are some sort of hybrid algorithms appeared in 1990s. They hybridize the classical BIOs with quantum computing (QC) paradigm to improve the performance of these algorithms especially in complex large problems with high dimensions.

The algorithms in this new field proved in many applications including networks security where they proved their effectiveness over the traditional BIOs. New hybrid algorithms can be introduced to enhance the performance of QIEAs and increase the quality of obtained solutions.

The aim of this paper is to build a NIDS based on a new proposed BIO named Quantum Vaccined Immune Clonal Algorithm with Estimation of Distribution Algorithm (QVICA-with EDA). This algorithm introduces the vaccine operator and the EDA sampling to be added to the classical Quantum Inspired Immune Clonal Algorithm (QICA). The algorithm is used to build a better NIDS with a higher intrusions classification accuracy. It is trained and tested Qualitative and Quantitative audit data's of the benchmark KDD dataset and is compared with another NIDS based on Particle Swarm Optimization (PSO). Results show the superiority of the proposed algorithm over the PSO. The rest of this paper is organized as follows: Section 2: Introductory part about related works in the field of IDS classification algorithms and a Technical background discussions of necessary terms. In Section 3: we are introducing the proposed algorithm where experimental results and discussion are in. Final Part is Section 4: The last section is discussion and future enhancement based this proposed techniques to conclusions and further works.

2. RELATED WORKS AND BACKGROUND

EAs and QIEAs have been used in past few years in many researches for intrusions classification to increase the detection accuracy of IDS. A sampling process about these researches are shown in this section. The immune clonal algorithm was used in many networking applications like intrusion detection for securing networks and the spam detection. A Network Intrusion DetectionSystem (NIDS) based on the immunological approach was proposed where an adaptive sampling algorithm was applied during the data collection stage. This algorithm was used according to the dynamic character detection data's(DCD's), where it can able to improve the data-processing speed and the fault tolerance of the system [8]. A collaborative intrusion detection system was proposed to detect denial of service(DOS) attacks by using the artificial immune system due to its distributional, collaborative, robust and adaptive capabilities. This system implementation used for peer-to-peer networks(P2P). where it was able to increase the precision of attack discovery and decreases false positive rate [21]. Also, An IDS based on Immune Algorithm (IA) and support vector machine (SVM) was introduced. In this two major methodologies (A).Immune Algorithm is used to preprocess the network data's (B). SVM is adopted to classify the optimization data, and recognize intruders. Results showed that the feasibility and efficiency of the system [5].

A hybrid intrusion detection system based on rough set (RS)for feature selection and simplified swarm optimization for intrusion data classification was applied. RS is proposed to select the most relevant features and the PSO with a new weighted local search(WLS) strategy was used for classification. The WLS is added to discover better solution from the neighborhood of the current solution produced by PSO. The testing results explored the proposed hybrid system can achieve highest classification accuracy [6]. Particle Swarm Optimization and its variants were combined with various Machine Learning techniques. They were used for Anomaly Detection in Network Intrusion Detection System to enhance the performance of system [22].

An improved Negative Selection Algorithm(NSA) integrates a novel further training strategies to reduce self-samples, to minimize computational cost in testing stage, was developed. This algorithm can able to get highest detection rate and the lowest false alarm rate in several cases [9].

The Support Vector Machine (SVM) was introduced as a data classifier for an Intrusion Detection System (IDS).It's a wrapper based feature selection approach using Bees algorithm (BA) as a search strategy for subset data generator. The result shows that these combined algorithms has yielded better quality IDS [2]. Also, an improved incremental SVM algorithm (ISVM) combined with a kernel function U-RBF was proposed and applied into network intrusion detection. The simulation results showed that the improved kernel function U- RBF has played some role in saving training time and test time [33]. An anomaly based network IDS using Gradient Analyzing approach was also adopted where the proposed IDS used for an adaptive GA for both learning and detecting intrusion which are Qualitatively circulating into the network while application starts its service in different remote locations. The proposed methodology was efficient with respect to good detection rate with low false positives in addition to

the lower execution time [24]. An IDS based on GA was proposed where GA uses evolution theory to information evolution in order to filter the traffic data and so reduce the complexity.

A QIEA using Eigen vectors and niching strategy was doing optimizer's into the database specially for signature based detection systems, an ID system type that is known with its poor detection performance [29], and the algorithm was able to improve the ID's ability in A quantum neural network (NN) was applied for intrusion detection application to overcome the weakness of the backpropagation method for NN which may fall into local minimum [4]. The quantum particle swarm optimization (QPSO) was used as a trainer method for NN and SVM to get better IDS performance. It was applied to train the wavelet NN to improve the detection rate for anomalies and reduce the false detection alarms in the network anomaly detection [17]. It was also applied with the Gradient Descent (GD) method to train the Radial Basis Function NN for network anomaly detection [18]. It was used with SVM for network Intrusion feature selection and detection where each particle was a selected as a subset of features and its effectiveness was defined as classification percentage by SVM [11]. It was used also for solving the linear system of equations of the least square SVM (LS-SVM

Algorithm 1 Observation Process

1: for i = 1 to m do
2: Generate a random number r between 0 and 1
3: $\leq h$
4: set the binary bit of i as 0
else
5. Repeat Step 2
6. Stop

2.1. Quantum Inspired Immune Clonal Algorithm

The artificial immune system algorithms (AIS) are a set of EAs inspiring their procedure from the Human Immune systemfunctionality. AIS include many different algorithms where the two major used algorithms are the negative selection algorithm and the immune clonal algorithm (ICA). ICA is inspired from the human immune systems clonal selection process over the B cells where the evolution process of the antibodies is a repeated cycle of matching, cloning, mutating and replacing.

The best B cells are allowed through this process to survive which increases the attacking performance against the unknown antigens. Vaccination is another immunological concept that ICA applies through the vaccine operator. This operator is used to introduce some degree of diversity between solutions and increase their fitness values by using artificial Vaccines.

detecting the unknown attacks [34]. A QIEA was also used for optimizing the features selection and kernel parameters of the support vector machine used for anomaly detection[31]. The Qualitative Gradient Analyst was used to optimize the clustering methodology and to collect the optimal number of clusters to be used for data classifications the data collected by the IDS [32].

Quantum Inspired Evolutionary Algorithms (QIEAs) were introduced in the 1990s, they integrate the quantum computing concepts with evolutionary process of EAs. They are able to improve the quality of solutions and enhance the algorithms performance as EAs suffer from bad performance in high dimensional problems. EAs, including the ICA, have to do numerous evolutionary operations and fitness evaluations in large problems which limit them from performing effectively. The hybridization between quantum properties and traditional ICA process enhanced its performance in complex problems. QICA combined quantum computing principles, like quantum bits, quantum superposition property and quantum observation process, with immune clonal selection theory. These concepts are described below in details where the quantum bit representation for antibodies and vaccines in QICA has the advantage of representing a linear superposition of states (classical solutions) in search space probabilistically. Quantum representation can guarantee less population size as a few number of antibodies and vaccines can represent a large set of solutions through the space [23]. The quantum observation process plays a great role in protecting the multi-state quantum antibodies into one of its basic states to help in the individuals evaluation.

Quantum Bit: By using quantum bit (q-bit) representation, a small population of antibodies can be created where it represents a larger set of antibodies due to the quantum superposition property. The quantum antibody population is initialized with n quantum antibodies using m q-bits for each one.

Where α and β are complex numbers and $\alpha^2$is the probability to have value 0 and $\beta^2$is the probability of having value 1 and $\alpha^2 + \beta^2 = 1$.

Observation Process: The quantum represented individuals are converted into binary representation by iterating over each q-bit in the individual and change it to a binary bit. The process is described in algorithm 1

Algorithm 1 Observation Process

1: for i = 1 to m do
2: Generate a random number r between 0 and 1
3: $\leq h$
4: set the binary bit of i as 0

5: else
6: set the binary bit as 1
7: end if
8: end for

Here the identifier 'h' represents a boundary value for the number generation Step 2.

2.2. Estimation of Distribution Algorithm

The most of evolutionary algorithms (EA) use sampling during their evolution process for generating new solutions. Some of these algorithms use it implicitly, like Genetic Algorithm (GA), as new individuals are sampled through the genetic operators of the crossover and mutation of the parents. Other algorithms apply an explicit sampling procedure through using probabilistic models representing the solutions characteristics. These algorithms are called the iterated density estimation evolutionary algorithms (IDEAs) where an iterated process of probabilistic model estimation takes place to sample new individuals [10], [16]. Estimation of Distribution Algorithms (EDAs), an example of the IDEA, are population based algorithms with a theoretical foundation on probability theory.

They can extract the global statistical information about the search space from the search so far and builds a probability model of promising solutions [25]. Unlike GAs, the new individuals in the next population are generated without crossover or mutation operators. They are randomly reproduced by a probability distribution estimated from the selected individuals in the previous generation [16].

EDA has some advantages:

1. Identifying the interrelations.
2. Inter dependencies between the Quantitative and Qualitative variables.
3. It doesn't require any additional parameters

They can extract the global and statistical information's about the searching space from the search location. The promising probability model [25]. Unlike GAs, the new individuals in the next population are generated without any crossover or mutation operators. They are randomly reproduced by a probability distribution estimated from the selected individuals in the previous generation [16]. The general EDA procedure is shown in

Algorithm 2.Estimation of Distribution Algorithm

1: Initialize the initial population.
2: while termination condition is not satisfied do
3: Select a certain number of excellent individuals.
4: Construct probabilistic model by analyzing information of the selected individuals.
5: Create new population by sampling new individuals from the constructed probabilistic model.
6: End of while

The EDA relies on the construction and maintenance of a probability model that generates satisfactory solutions for the problem solved. An estimated probabilistic model, to capture the joint probabilities between variables, is constructed from selecting the current best solutions and then it is simulated for producing samples to guide the search process and update the induced model. Estimating the joint probability distribution associated with the data constitutes the bottleneck of EDA. Based on the complexity of the model used, EDAs are classified into different categories, without interdependencies, pair wise dependencies and multiply dependencies algorithms as below [10].

Null Interdependencies EDAs:

These models are used when there is no dependency assumed between the variables of the problem.The jointprobability distribution is factorized to n independent univariate probability distributions p(xi). Univariate Marginal Distribution Algorithm (UMDA), an example for this category, estimates the p(xi) from the relative marginal frequencies of the Xi of the selected data. Other examples of EDAs under this category are the Population Based Incremental Learning (PBIL) and compact genetic algorithm (cGA) [14]. UMDA, as an example.

Clustered Dependencies EDAs:

This type of EDAs assumes dependency between pairs of the variables. The joint probability distribution of the variables is factorized as the product of a univariate density function and (n - 1) pairwise conditional density functions given a permutation = (, … ,) between variables. Examples are the Bivariate Marginal Distribution Algorithm (BMDA), Mutual Information Maximization for Input Clustering (MIMIC) and Combining Optimizers with Mutual Information Trees (COMIT) algorithms.

Multi-Interdependencies EDAs:

Dependencies are assumed between multiple variables where probabilistic graphical models based on either directed or undirected graphs are widely used. Structural and parametric learning are done to learn the topology of the networks and estimate the conditional probabilities. Bayesian network algorithm (BOA), the Markov network EDA and factorized Distribution Algorithm (FDA) are some examples. (For more details, see [14] and [28]).

3. PROPOSED NETWORK INTRUSION DETECTION SYSTEM.

This work builds a IDS using a new classification algorithm to improve the detection performance. It proposes a new BIO named Quantum Vaccinated Immune Clonal Algorithm(QVICA)

with Estimation of Distribution Algorithm. It is based on the quantum computing concepts, immune clonal selection principles and the vaccine operator with EDA sampling. The IDS is then compared with another system base on PSO[6]. A general schema of the proposed IDS is shown in figure1.

Figure1. Shows that the system has three main stages. First stage is about data preprocessing; the second is the training phase of the proposed classification algorithm and the last one is the testing phase where a detailed description is shown below.

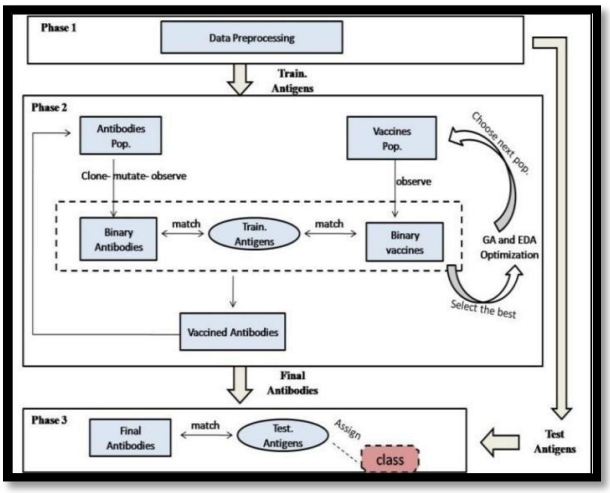

Fig 1: Time quantum multivariate optimization algorithm.

3.1 Data Preprocessing Phase:

This phase is concerned with the preprocessing of data represented in the dataset records (network connections). The data are divided into two sets, the training set and the test set. There is a class assigned to each record in the training set, to indicate either it is a normal connection or an attack one, where test records are without class labels. Each record does not have the total 41 features of the KDD, instead it has only six selected features. These six features are those selected and used in the PSO work which are the Service, src_bytes, dst_bytes, Rerror_rate, dst_host_srv_count and dst_host_diff _srv_rate. The proposed NIDS uses the same selected features for fair comparisons where the symbolic conversion of the symbolic features are done followed by Equal frequency discretization (EFD) for the continuous features. The processed data is the two outputs of this phase where the processed training set, called training Antigens (AGs), is used as an input for the training phase.

3.2 Training Phase

In this phase, the QVICA-V with EDA is trained using the training AGs to learn how to classify the data into normal and intrusions. as mentioned before, the algorithm integrates the quantum computing, for representation, and immune clonal selection principles with thevaccine operator for diversifying solutions. The algorithm also utilizes the EDA probabilistic models and sampling to improve the fitness of antibodies (solutions), increase the degree of diversity and shorten the execution time of the whole algorithm. The main steps of the proposed algorithm are described in Algorithm 3 [23]. As shown in algorithm 3, the algorithm starts by initializing both the quantum antibody population Q(t) and the quantum vaccine population V (t) followed by cloning and mutating antibodies to be then decoded for evaluation. Additional steps to the simple QICA, like vaccine decoding and sampling will be described in details.

Initialization: Quantum antibodies and vaccines populations are created where V (t) is initialized with n quantum vaccines where n is the number of grids.

Testing Phase: In this phase, the QVICA-with EDA is tested to evaluate the training process. It is tested using the test antigens, with no class label, produced by the first phase. The final trained antibodies, from the training phase, are matched with the test antigens. Each testantigen is matched with the whole set of these antibodies to be assigned to a class. The higher the number of matched ABs with the antigen, the more probability that this antigen follows their class. At the end of the phase, the class labels of all the test antigens are detected either as normal or attack.

4. EXPERIMENTS AND RESULTS

The proposed algorithm is implemented as the classification algorithm for a NIDS and compared with another classification algorithm based on PSO [6]. The experiments were implementedover the KDD-Cup 99 (Knowledge Discovery and Data Mining Tools Conference), a benchmark dataset for the network intrusion detection systems [1]. Each record in the KDD represents a TCP/IP connection that is composed of 41 features that are both qualitative and quantitative in nature. There are 39 types of distinct attacks in KDD, grouped into four classes of attack and one class of non-attack (normal connections). The main attack types are Denial of Service (DoS), Probe, Remote-to-Local (R2L) and User-to-Root (U2R) where detailed description of each and its sub types is below [19], [13] and [3].

Denial of Service (DoS) attacks: where an attacker makes some computing or memory resource too busy or too full to handle legitimate requests, thus denying legitimate users access to a machine. Sub attacks of

DoS are, Back, Land, Neptune, Pod, Smurf and teardrop.

Remote-to-Local (R2L) attacks: where an attacker sends packets to a machine over a network, then exploits machines vulnerability to illegally gain local access as a user. Sub attacks of R2L that can be found in the sets' records are guess-passed, ftp-write, Imap, Phf, multihop, warez master, warez client and spy.

User-to-Root (U2R) attacks: where an attacker starts out with access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system. U2R sub attacks are Buffer- overow, load module, perl and rootkit.

Table 1: Accuracy For Different Experiments Using Qvica-With Eda.

QVICA-with EDA				
	Iterations			
Pop Size	**15**	**20**	**35**	**45**
15	89.0	86.5	87.7	89.9
20	89.1	87.1	92.5	89.1
35	86.6	90.8	93.1	93.4
45	93.8	93.2	94.7	94.5
50	89.3	93.4	94.8	94.2

The same parameters settings of the PSO-WLS algorithm are used for our algorithm for fair comparison. A set of 4000 records is selected from the KDD based on the selected features of the PSO-WLS work to evaluate the performance of the QVICA- with EDA. The 10-fold cross validation method is applied where the data are distributed as 10 for testing and the remaining 90 for training.

Table2: Accuracy For Different Experiments Using Pso-Wls

PSO-WLS				
	Iterations			
Pop Size	**08**	**15**	**20**	**35**
8	85.6	87.0	90.4	91.8
15	87.7	89.0	90.6	93.2
20	88.6	89.5	93.5	93.5
35	88.6	90.0	93.5	93.6
45	88.6	91.5	93.5	93.6

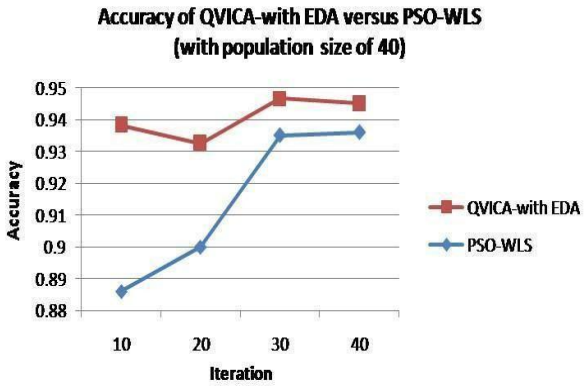

Fig 3: Classification Accuracy at different iterations with population size of 40

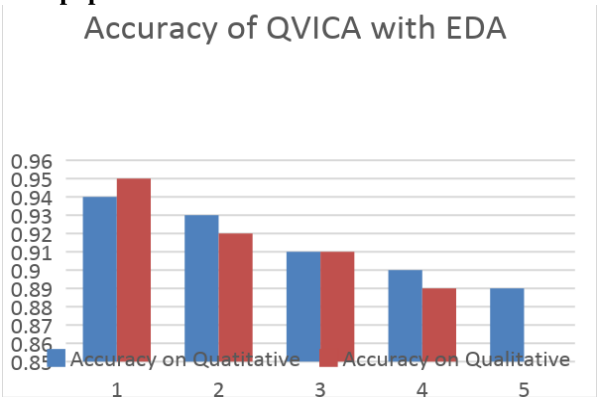

Fig 3.1: classification accuracy based Qualitative and Quantitative size

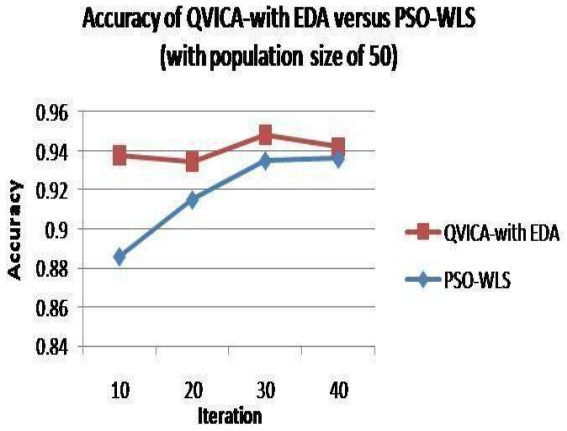

Fig 4: Classification Accuracy at different iterations with population size of 50

For more investigation, the accuracy of each fold in one of the best experiments is recorded in table 3 and compared with the folds results of the PSO-WLS algorithm where higher values are obtained by our algorithm. The detailed accuracy values of the 20 runs of the best experiment is in table 4 and are compared

with the best experiment of the other algorithm. The proposed algorithm is able to get higher classification accuracy than the best value of the PSO based algorithm using the same parameters.

Table 3: accuracy of the 10 fold cross validation of one of the best runs.

Fold	Accuracy Value	
	QVICA-with	**PSO-WLS EDA**
1	0.9425	0.943
2	0.9776	0.925
3	0.9626	0.925
4	0.8803	0.938
5	0.9476	0.933
6	0.9476	0.925
7	0.9027	0.943
8	0.9651	0.945
9	0.9651	0.945
10	0.9451	0.933

Table 4: accuracy of the two algorithms over 20 runs (for the best experiment)

Accuracy Value		
Run	**QVICA- with EDA**	**PSO-WLS**
1	94.1	93.4
2	92.9	93.5
3	93.6	93.3
4	94.7	93.5
5	94.2	93.2
6	92.6	93.3
7	88.0	93.2
8	90.2	93.4
9	91.2	93.5
10	94.0	93.4
11	92.1	92.5
12	92.3	93.3
13	93.9	93.2
14	91.8	93.4
15	93.3	93.3
16	93.8	93.0
17	92.0	93.4
18	91.1	93.5
19	94.8	93.0
20	89.6	93.4
Mean	**94.81**	**93.3**

5. CONCLUSIONS

A quantum vaccinated immune clonal algorithm with the estimation of distribution algorithm (QVICA-with EDA) was proposed in this paper as a classification algorithm for the NIDS. It was compared with another classification algorithm based on particle swarm optimization (PSO) on the KDD data sets based on Qualitative and Quantitative intruders in fast networked environment. Classification accuracy values obtained at the different experiments showed the ability of the algorithm of achieving high classification accuracy.for feature selection of ovarian cancer data. In Education Technology and Training, 2008. and 2008 International Workshop on Geoscience and Remote Sensing. ETT and GRS 2008. International Workshop on, volume 1, pages 681–685. IEEE, 2008.

REFERENCES

[1] http://kdd.ics.uci.edu/databases/kddcup99.
[2] Osama Alomari and Z Othman. Bees algorithm for feature selection in network anomaly detection. Journal of Applied Sciences Research, 8(3):1748–1756, 2012.
[3] Ver´onicaBol´on-Canedo, Noelia S´anchez-Maro˜no, and Amparo Alonso-Betanzos. Feature selection and classification in multiple class datasets: An application to kdd cup 99 dataset. Expert Systems with Applications, 38(5):5947–5957, 2011.
[4] FENG Jian-li GONG Chang-qing. Study of an intrusion detection based on quantum neural networks technology [j]. Journal of Shenyang Institute of Aeronautical Engineering, 1:016, 2010.
[5] Yu Sheng Chen, Yu Sheng Qin, Yu Gui Xiang, Jing Xi Zhong, and Xu Long Jiao. Intrusion detection system based on immune algorithm and support vector machine in wireless sensor network. In Information and Automation, pages 372– 376. Springer, 2011.
[6] Yuk Ying Chung and Noorhaniza Wahid. A hybrid network intrusion detection system using simplified swarm optimization (sso). Applied Soft Computing, 2012.
[7] Tian Fang, Dongmei Fu, and Yunfeng Zhao. A hybrid artificial immune algorithm
[8] Qinghua Zhang; Yuzhen Fu. Research of adaptive immune network intrusion detection model. International Journal of Systems, Control and Communications, 3(3):280–286, 2011.
[9] Maoguo Gong, Jian Zhang, Jingjing Ma, and Licheng Jiao. An efficient negative selection algorithm with further training for anomaly detection. Knowledge-Based Systems, 30:185–191, 2012.
[10] Xiaojuan He, Jianchao Zeng, SongdongXue, and Lifang Wang. An new estimation of distribution algorithm based edge histogram model for flexible job-shop problem. In Computer Science for Environmental Engineering and EcoInformatics, pages 315–320. Springer, 2011.
[11] ZhangHongmei, GaoHaihua, and Wang Xingyu. Quantum particle swarm optimization based network intrusion feature selection and detection. 2007.

[12] MohammadSazzadulHoque, MdMukit, MdBikas, Abu Naser, et al. An implementation of intrusion detection system using genetic algorithm. arXiv preprint arXiv:1204.1336, 2012.