IJCTA, 9(24), 2016, pp. 25-38

Efficient and Secure Intrusion Detection System Based on Feature Subset Selection with Optimized Machine Learning for Wireless Sensor Network

P. Nirmaladevi*, and A. Tamilarasi**

ABSTRACT

Network Intrusions are grave issues in computer and network systems. Several intrusion detection approaches be present to determination these severe problems but the major problem is performance. To increase performance, it is significant to rise the detection rates and reduce false alarm rates in the area of intrusion detection. The recent approaches use Principal Component Analysis (PCA) towards project features space to principal feature space and select features corresponding to the highest eigenvalue, but the features corresponding to the highest eigenvalue may not have the optimal sensitivity for the classifier due to snubbing many sensitive features. Instead of using traditional approach of selecting features with the highest eigenvalue such as PCA, proposed method applied a Bacteria Foraging (BF) to search the principal feature space for Bacteria eigenvectors that offers a subset of features with optimal sensitivity and the highest discriminatory power. To improve the network security and easy detection of malicious attacks at the classification stage this paper proposed an effective method that uses the modified artificial bee colony (mABC) to optimize the Kernel Extreme Learning Machine (KELM). With proper hidden layer neuron number, KELM could enhance the accuracy and speed of the intrusion detection. To verify the proposed technique, experimental tests have been implemented in this work. The test result demonstrates that the proposed mABC-KELM can detect the network intrusion efficiently and its performance is superior compared to existing methods.

Keywords: Intelligent Detection, Network Intrusion, Artificial Neural Network, Principal Component Analysis, Bacteria Foraging, Kernel Extreme Learning Machine.

1. INTRODUCTION

Wireless Sensor Network (WSN) is a kind of network that have many (from dozens to thousands) minute devices sensing and collecting detailed information about the physical environment. Due to WSNs economical cost and simple propagation characteristics, they are used for many different fields of science, health, military, security to sense and gather data respecting various activities, for instance exploring the battlefield (e.g.-Boomerang Sniper Identifying System), monitoring highway traffic, identifying NBC (Nuclear, Biological, Chemical) attacks, fire alarm system, learning wildlife and oceans (Great Duck Island-GDI Project), home automation systems, agriculture, transportation, space exploration and many others. With the enormous growth of WSN-based computer services and the huge increase in the number of requests running on networked systems, the adoption of appropriate security measures to protect against computer and network intrusions is aessential issue in a computing environment. Intrusions into or attacks on a computer or network system are actions or attempts to destabilize it by compromising security in confidentiality, availability or integrity of the system. As defined in [1], an Intrusion Detection System (IDS) monitors events occurring in a network and examines them for signs of intrusions.

* Assistant Professor, Department of M.C.A, Maharaja Engineering College, Avinashi, Tamilnadu , *Email: pnirmaladeviresearcher@gmail.com*

** Professor and Head, Department of M.C.A, Kongu Engineering College, Perundurai, Tamilnadu, *Email: drtamil@kongu.ac.in*

The current internet-based information processing systems are prone to different types of threats which lead to various types of damages resulting in significant losses. Therefore, the significance of information security is evolving quickly. The most basic goal of information security is to improve defensive information systems which are protected from unauthorized access, use, disclosure, disruption, modification, or destruction. Furthermore, information security minimizes the risks related to the three main security goals namely, integrity, confidentiality and availability. Many systems have been designed in the past to identify and block the Internet-based attacks.

The furthermost important systems among them are intrusion detection systems (IDS) since they resist external attacks effectually. Furthermore, IDSs provide a wall of defense which overcomes the attack of computer systems on the Internet. IDS could be used to identify different types of attacks on network communications and computer system usage where the traditional firewall cannot execute well. Intrusion detection is based on an assumption that the behavior of intruders differ from a legal user [2]. Generally, IDSs are broadly classified into two categories namely anomaly and abuse detection systems based on their detection approaches [3]. Anomaly intrusion detection determines whether deviation from the established normal usage patterns can be labelled as intrusions.

On the other hand, misuse detection systems detect the violations of permissions effectually. Intrusion detection systems can be built by using intelligent agents and classification methods. Most IDSs work in two phases namely preprocessing phase and intrusion detection phase. The intrusions identified by the IDSs can be prevented effectually by evolving an intrusion prevention system. IDS systems can be divided into two methods: abuse detection and anomaly detection [4]. Misuse detection can detect the attacks based on famoussusceptibilities and patterns of intrusions (attacks signatures) stored in a database. It matches the current conduct against the previous knowledge of those known attack patterns [5]. Therefore, this technique may not able to aware the system administrator in case of a new attack. Conversely, Anomaly detection creates a normal conduct profile and detects the intrusions based on important deviations from this normal profile [6]. Thus, anomaly detection methods can detect new types of attack, but it suffers from a high rate of false alarms to train in highly dynamic surroundings. Many challenges need to be considering when building an IDS, such as data collection, classification accuracy anddata preprocessing. Classification is the prediction of the category labels of instances that are classically described through a set of features (attributes) in a dataset.

Numerous classification techniques have been proposed for the improvement of IDS; with Fuzzy Logic (FL) [7], Neural Networks (NN) [8], Support Vector Machines (SVM) [5, 6] and Decision Tree (DT) [9]. Another important problem for constructing IDS is dealing with data comprising large number of features. Data in high dimensional space may lead to reduction the classification accuracy of the IDS. Therefore, feature selection is mandatory as a pre-processing phase for high dimensional data before solving the classification problems. Feature selection aims to decrease the number of irrelevant and redundant features.

An anomaly network intrusion detection system is proposed in this paper using Bacteria Foraging (BF) based feature selection method and the modified artificial bee colony (mABC) to optimize the Kernel Extreme Learning Machine (KELM). The effectiveness of the proposed network IDS is assessed by conducting numerous experiments on NSL-KDD network intrusion dataset. The results show that the proposed mABC-KELM rises the accurateness and speeds up the detection time. The rest of this paper is prepared as follows: Section 2 presents a background of the used approaches; Section 3 describes feature selection method and the proposed mABC-KELM IDS system. Section 4 gives the execution results and analysis. Finally, Section 5 contains the conclusion remarks.

2. RELATED WORK

Security threats take place in wireless sensor network are different from wired network threats because of structure of WSN and constrictions which it has such as limited battery life. Hence IDS implemented in

WSN has different approaches [10]. In this section, it is described that approaches pointed out by some important studies achieved in recent years. All classifications of detection approaches made by different researcher occur from public IDS taxonomy (Misuse Detection, Anomaly Detection). Due to different features of WSNs from wired and non-energy constrained wireless networks, different classification types is pointed out in this section.

In [11], classifying is made as intrusion type, intruder type, detection methods, source of the collected data, analyzing location of the collected data, usage frequency and this categorizing is the most comprehensive in the literature. In a network, intruder type is grouped into two categories. These categories are internal intruder (selfish or malicious node) and external intruder (An outside attacker trying to reach the system). In WSN, consistent with intrusion type, intrusion can be by stealing the data, by creating false data and so altering the system, by rejecting to access the system, by influencing the energy efficient. For detection approaches it has been described above as misuse and anomaly detection but additionally some papers point out hybrid or specification based detection.

In A Game-Theoretic Framework for Robust Optimal Intrusion Detection in Wireless Sensor Networks-2014, it is demanded that instead of approaches using heuristic and adhoc solutions, there is an increase to use analytical methods for security issues in WSN. Hence authors propose a nonzerosum discounted robust stochastic game framework to analyse intrusion detection difficult in WSN. Game' s parameters are modelled by features of WSN and it's environment [12].

In Anomaly Detection and Localization in UWB Wireless Sensor Networks-2013, author has been proposed an anomaly detection solution specifically designed for the ultrawideband (UWB) technology. In the paper, it is described that UWB is a key answer to serve low power consumption while wireless connectivity. To identify intrusions, a rule based approach is accepted and performance of the proposed algorithm is studied by simulations. The algorithm projected in the paper, uses a round-based (There are particular phases.) approach towards cluster structure and rule based anomaly detection. The test outcomes shown in paper point out a successful detection accuracy [13].

In Applying Data Mining Methods to Intrusion Detection in Wireless Sensor Networks-2013, it is proposed that the application using data mining approaches for intrusion detection system in wireless sensor network and proposed system can perform both irregularity detection method and misuse detection method. The IDS consists of a Central Agent and several Local Agents, which are located on the sensors and carry out intrusion detection activities. Data mining approach is used on everyagents (Local Agents, Central Agents). The test outcomes show that high detection accuracy is obtained while keeping an acceptable, however not negligible false positives rate [14].

In Intrusion Detection in WSN Using Watchdog Based Clonal Selection Algorithm [15], the watchdog approach is used to detect whether a node has abnormal behaviour while transmitting data. All nodes in the WSN isanswerable for monitoring the neighbours and transferring the info about behaviour. Misbehaviour of nodes affect act of WSN negatively. With using watchdog based method.

The execution of genetic algorithms on top of information theory to enrich intrusion detection has been proposed by Xiao, et. al. [16]. Genetic algorithms have been used for classification of Smurf attack labels in training data set, achieving a false positive rate as low as 0.2% by Goyal and Kumar [17]. Abdullah, et. al. [18] used genetic algorithms for procurementclassification rules for intrusion detection. Ojugo, et. al. [19], have used genetic algorithms to improve rule-based intrusion detection. The suitability function has been used to evaluate the rules. In one existing research work by Liu et al [20] and his colleagues, PCA is practical for classification and neural networks are meant for online computing. They selected 22 principal modules as features subset selection to obtain the best presentation. But there is a possibility to miss many vital principal modules having sensitive information for intrusion detection during selection phase.

Horng et al [21] and his co-workers observed the important features based on the accuracy and the number of false positives of the system considered with and without the feature. In additional words, the feature selection of is “leave-one-out”, remove one feature from the original dataset, redo the test, then relate the new results with the original result, if any case of the described cases happens. The feature is regarded as significant; else it is regarded as insignificant. Since there are 41 features advised in the KDD-cup99, the test is repeated 41 times to certify that each feature is either essential or inconsequential. This process involved obstacle as well as overheads on massive dataset. One of the most important works is done by Tong et al [22] and his contacts in which they employed the Radial Basis Function (RBF) network is a real-time pattern classification and the Elman network is applied to restore the memory of earlier events. They used full featured KDD-cup dataset. This rises training and testing overheads on the system.

3. PROPOSED MODEL

The proposed model contain different parts; dataset used for experiments, feature transformation and organization, classification architectures, optimal feature subset selection, implementation, training and testing, and results comparison. The proposed method communicate with cluster based medium and watch dog mechanism is monitored and updated the neighbor nodes information’s in each node information table. The block diagram of proposed model is shown in the Figure 1.

3.1. Dataset used for Experiments

The proposed system used ADFA dataset for tests. The selection of this dataset is due to its standardization, content richness and it helps to evaluate proposed resultswith existing methods of intrusion detection system. ADFA is developed using a modern operating system and contemporary attacks, for further estimation of the semantic algorithm. This new dataset is available for public use without restriction, and can be collected fromj.hu@adfa.edu.aulink. The datasetcontains833normal traces for training the IDS, 4373 normal traces for valuing FAR and 60 different attack sets, each consisting of multiple traces.

3.2. Dataset pre-processing

First, the data are pre-processed, and then it is given to the selected classifiers. The raw dataset is pre-processed for removingsymbolic values and feature transformationusing PCA. Finally, optimal features subset selection using BF.

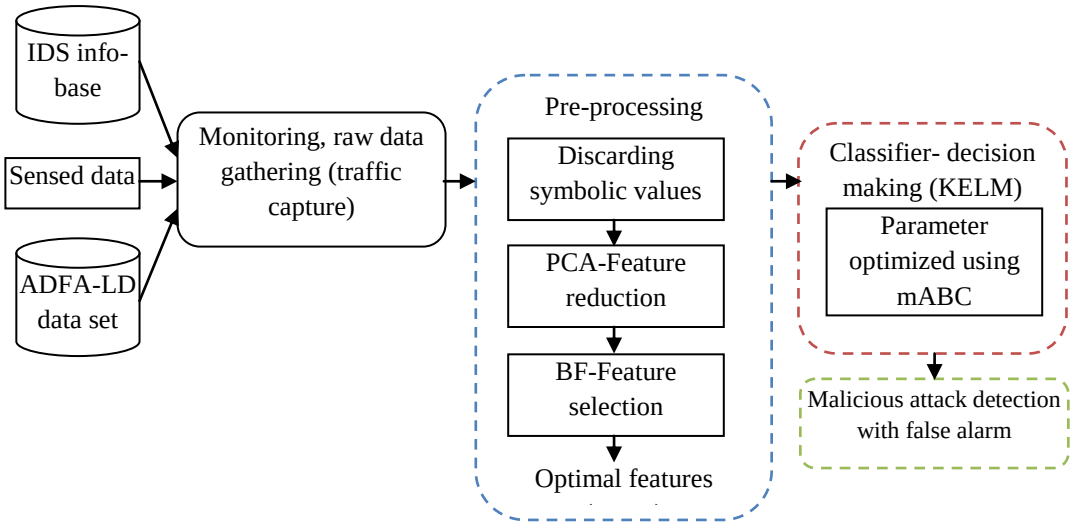

Figure 1: overall process of proposed system

3.2.1. Feature transformation and organization

In pre-processing second stage, PCA is applied for data reduction, but in this work, PCA used for feature transformation into principal components feature space and it is organized in descending order. PCA purposely is to reduce the dimension of the data while retaining as much as probable of the variation present in the original dataset. It provides a way of identifying patterns in data and state the data in such a way as to highpoint their similarities and differences [23]. However, PCA here was used to transform the input vectors to the new search space. On the other hand, the choosing of number of principal components is done by BF. PCA Algorithm is represented as follows:

Input: $x = (x_1, x_2 \ldots . x_M)\mathbf{M} -$ Maximum number of features

1. Find mean $\bar{x} = \frac{1}{M}\sum_{i=1}^{M} x_i$
2. Calculate deviation $\leftarrow$ mean $\Phi_i = (x_1 - \bar{x})$
3. Find Covariance matrix (A)// A is (N × M)matrix→ $\Phi_i (i = 1 \ldots M)$// through $C = \frac{1}{M}\sum_{N=1}^{M} \Phi_N \Phi_N \equiv AA^T$
4. Compute eigenvalue (C) as $C{:}\lambda_1 > \lambda_2 \ldots . \lambda_N$
5. Compute eigenvector (C) as $C{:}\mu_1 > \mu_2 \ldots . \mu_N$// C is symmetric form a basis, (i.e. any vector x or $x_1 - \bar{x}$ actually, can be written as a linear combination of the eigenvectors): $x_1 - \bar{x} = b_1\mu_1 + b_2\mu_2 \ldots . + b_N\mu_N = \sum_{i=1}^{N} b_i\mu_i$ //
6. Arrange (λ_i, μ_N) // descending order //

$$pc_1 > pc_2 \ldots \ldots > pc_l$$

// l is the transformed number of features//

3.2.2. Feature Subset Selection

In third step of pre-processing, Bacteria Foraging (BF) applied for optimal features subset selection from principal components search space. This is a main contribution that positively impact on the act of intrusion detection analysis engine. BF is inspired by the biological mechanisms of reproduction. In BF, E.coli bacteria searching principles are: swarming, chemotaxiselimination dispersal and reproduction.

Chemotaxis: During chemotax is in a nutrient medium, an E.coli bacterium θ_i tumbles an unit step, $C(i)$ in a random direction, given by a random vector $\Delta(i) \in R^{2D}$ whose each component is between [-1,1]. This tumbling behavior is given by (3). If this is found to be favorable then θi swims for a period of time (swim length, N_s) in that direction. The number of chemotactic steps is determined by N_c.

$$\theta_i(j+1,k,l) = \theta_i(j,k,l) + c(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \quad (1)$$

Swarming: When an E.coli cell moves up the nutrient gradient, it releases an attractant. Due to this, several cells of E.coli form stable spatio-temporal patterns of concentric rings (swarms). The cell-to-cell signaling J_{cc} by a bacteria θ to the total population P is given by (4). This is added to the fitness function $J(\theta)$(i.e. optimal feature selection)to provide a time-varying fitness.

$$\begin{aligned} J_{cc}\big(\theta, P(j,k,l)\big) &= \sum_{i=1}^{NP} J_{cc}\big(\theta, \theta_i(j,k,l)\big) \\ &= \sum_{i=1}^{NP}\left[-d_{attr}\ exp\left(-w_{attr}\sum_{m=1}^{2D}(\theta^m - \theta_i^m)^2\right)\right] \\ &+ \sum_{i=1}^{NP}\left[h_{repllent}\ exp\left(-w_{repellent}\sum_{m=1}^{2D}(\theta^m - \theta_i^m)^2\right)\right] \end{aligned} \quad (2)$$

Reproduction: From the total population, the least healthy half dies (deleted) and the healthiest half asexually splits into two (copied) to keep the population size constant. It is to be noted, here, that higher value of J_θ means lower health of the bacteria θ. This iterates for N_{re} steps. After every reproduction phase, chemotaxis and swarming repeats.

Elimination-dispersal: Sometimes, sudden change in environment kills a few bacteria and to balance this nature disperses some bacteria at a new location. For simulation, a bacteria is eliminated with probability ped. If a bacterium is eliminated, another bacteria is dispersed randomly at any location on the optimization domain. This continues for Ned steps where every step is followed by chemotaxis, swarming and reproduction.The main goal of feature subset selection is to use less features to achieve the same or better performance. Therefore, the fitness evaluation contains two terms: (i) accuracy and (ii) the number of selected features. The features are defined the traces in dataset.

3.3. Classification Architectures

3.3.1. Modified Artificial Bee Colony algorithm

Karaboga proposed Artificial Bee Colony (ABC) algorithm for real-parameter optimization [24], and is based on the behavior of a bee colony. The algorithm has three types of bees such as employed, onlooker and scout bees. In this ABC process, the Employed Bees (EB) is consider as very important role, because its only responsible for exploiting the nectar sources and giving information to Onlooker Bees (OB) (i.e waiting bees). If OB is deciding the food source, otherwise the Scout Bee (SB) randomly searches the environment and chosen a new food source based on the internal or external motivation [25]. The step by step process of ABC is given below

Step 1: initially, the bees are randomly search in environment to find food source.

Step 2: after that, the EBs exploited the discovered source and it returns to the hive with the nectar and unloads the nectar. After that, EBs can go back to their discovered source site directly or share information to OBs. If EBs source is exhausted, that bee becomes a scout and starts to randomly search for a new source.

Step 3: OBs choosing the profitable or fitness source site depending on the frequency of a dance proportional to the quality of the source.

Detection of initial food source sites

Initially, the algorithm randomly produced foot source within the range of the boundaries of the parameters in the search space.

$$p_{ij} = p_j^{min} + \alpha(p_j^{max} - p_j^{min}, i = 1,..,FS, j = 1,\dots,OP \quad (3)$$

Where p is the initial population, α is randomly choose ϵ (0,1), FS indicated the number of food sources and OP represents the number of optimization parameters. In this phase, the numbers of trials of solutions are reset to 0. After that, the food source solution is forward the bee's cycle process. The termination criteria for the mABC algorithmcan be, reaching a Maximum Cycle Number (MCN) or meeting an error tolerance (ϵ).

Cycle process for EB

After initialization of food source, the EBs is sending to the food source sites. The EBs depends on the number of food sites. These EBs are produces a modified position of the food source and stored in their memory based on the local search data and find nearby food sites, and valuates the food site quality. In this proposed mABC, the each population parameter p_{ij} is a uniformly distributed random numbe R_{ij}, $(0 \leq R_{ij} \leq 1)$ is produced and if the random number is less than theModification Rate (MR), then the parameter x_{ij} is modified as in the Equation (2).

$$v_{ij} = \begin{cases} p_{ij} + \phi_{ij}(p_{ij} - p_{kj}), & if\ R_{ij} < MR \\ p_{ij} & otherwise \end{cases} \quad (4)$$

Where j represents a random integer in the range $[1, OP]$ and $k \in \{1, 2, \ldots, FS\}$ that has to be diverse from i and MR is the Modification Rate which takes value between 0 and 1. A lower value of MR may cause solutions to improve slowly while a higher one may cause too much diversity in a solution and hence in the population.

In this process, in case, the parameter value exceeded from the boundary means, it will automatically set the predefined the boundary. If$p_i > p_i^{max}$ then it indicates as $p_i = p_i^{max}$ as well as, if$p_i < p_i^{min}$ then it indicates as $p_i = p_i^{min}$. Within the boundaries, the v_i is predicted and then a fitness value can be assigned to the solution $v_i$by Equation (3) for a minimization problem.

$$fitness_i = \begin{cases} {}^{1}/_{(1+f_i)} & if\ f_i \geq 0 \\ 1 + abs(f_i) & if\ f_i < 0 \end{cases} \quad (5)$$

Where f_i indicates the cost value of the v_i solution. This cost function can be directly used in maximization problem. Here, a greedy selection is done among p_i and v_i and the selected the best one from the represented fitness value. In case, the source at $v_i$is better profitability than p_i means, the EB stored the new position and eliminates the existing position value or else the old position is kept in memory. If $p_i$cannot be better, its counter asset the number of trials is incremented by 1, or else the counter is reset to 0.

Probabilistic selection

After the EB search, they share source site information related to the nectar amounts and their positions to OBs on the dance area. The OBs are selects a food source through a probability related to its nectar amount. The probability is calculated based on the fitness of the solutions in the population. The fitness based selection is a roulette wheel selection scheme. In this selection, each solution is proportional in size to the fitness value and defined as

$$P_i = \frac{fitness_i}{\sum_{i=1}^{FS} fitness_i} \quad (6)$$

OBs selects food source site based on EBs information

In this proposed mABC, a random real number $\in$ range of [0, 1] is generated for each source. In case, the probability P_i value of the source is exceeding than random number R_{ij} means, the OB generates a modification on the position of this food source via Eq. (4) as in EB case. After the source is evaluated, a greedy selection is done among p_i and v_i and the selected the best one from the represented fitness value. In case, the source at $v_i$is better profitability than p_i means, the OB stored the new position and eliminates the existing position value or else the old position is kept in memory. If $p_i$cannot be better, its counter asset the number of trials is incremented by 1, or else the counter is reset to 0. This process is repeated until all OBs are distributed onto food source sites.

The pseudo code of the modified ABC is given below:

1. **initialize population** $\boldsymbol{p_{ij}}$
2. evaluate population $\boldsymbol{p_{ij}}$
3. cycle=1
4. repeat

// to produce new food source population for employed bee//

5. for i=0, i++, i=FS
6. do
7. **produce new food source by the below eq**.

$$\boldsymbol{v_{ij}} = \begin{cases} \boldsymbol{p_{ij}} + \boldsymbol{\phi_{ij}}(\boldsymbol{p_{ij}} - \boldsymbol{p_{kj}}), & \boldsymbol{if\ R_{ij}} < MR \\ \boldsymbol{p_{ij}} & \boldsymbol{otherwise} \end{cases}$$

8. apply greedy($\boldsymbol{p_i}$, $\boldsymbol{v_i}$) → select better one
9. if solution $\boldsymbol{x_i}$ does not improve $\boldsymbol{trial_i = trial_i + 1}$
10. else
11. $\boldsymbol{trial_i = 0}$
12. end for
13. calculate $\boldsymbol{P_i}$ by the blow Eq.

$$\boldsymbol{P_i} = \frac{\boldsymbol{fitness_i}}{\sum_{i=1}^{FS} \boldsymbol{fitness_i}}$$

//produce a new food source population for onlookers//

14. t=0,i=1
15. repeat
16. if random<$\boldsymbol{P_i}$
17. then
18. go to step 7 to 11
19. t = t +1
20. end if
21. until t= FS

// Determine Scout//

22. **if max** ($\mathbf{trial_i} > \boldsymbol{limit}$)
23. then
24. replace $\boldsymbol{p_i}$ with a new randomly produced solution by:

$$\boldsymbol{p_{ij}} = \boldsymbol{p_j^{min}} + \boldsymbol{\alpha}(\boldsymbol{p_j^{max}} - \boldsymbol{p_j^{min}}$$

25. end if
26. the best solution achieved and stored so far
27. Cycle=cycle+1
28. Until (cycle=MCN)

Desertion criteria

After the completion of EB and OB searches, this algorithm verifies there is any exhausted source to be abandoned. If source is abandoned, then counter values updated during search. Checked the counter value is exceed than the control parameter of mABC, and is called as limit, and then the source is associated with this counter. It is assumed to be exhausted and is abandoned. The source is abandoned by it bee and is changed with a new source by the scout and is fluctuations property in the self-organization of mABC. This is implemented via generating randomly food source site position and is replaced with abandoned source. The abandoned source is mentioned as p_i and then randomly discovers a new site by scout to be replaced with p_i.

3.3.2. The Kernel Extreme Learning Machine (KELM)

Given samples$\{(x_i, t_i): i = 1,2 \ldots . N;\ x_i \in R^P, t_i \in R^q\}$,where x is the feature vector and $t$is the class label vector, the below SLFN is used to identify the sample [26].

$$\sum_{i=1}^{m} \beta_i g\left(\alpha_i^T x_j - b_i\right) = o_i,\ j = 1,2 \ldots . . N \qquad (7)$$

Where,m is the number of hidden neuron; $o_i$is the output of jth sample; $g(.)$ is the activation function$b_i$is the threshold of the ith hidden neuron; α_i and $\beta_i$are the input and output weight vectors, respectively. If the output $o$can approximate t, we derive:

$$\sum_{i=1}^{m} \beta_i g\left(\alpha_i^T x_j - b_i\right) = o_i = t_j,\ j = 1,2 \ldots . . N \qquad (8)$$

(6) can be written compactly as:

$$G\beta = T, \qquad (9)$$

Where,

$$G = \begin{bmatrix} g(\partial_1^T x_1 - b_1) & \cdots & g(\partial_m^T x_1 - b_m) \\ \vdots & \ldots & \vdots \\ g(\partial_1^T x_N - b_1) & \cdots & g(\partial_m^T x_N - b_m) \end{bmatrix}$$

$\beta = [\beta_1, \beta_2, \ldots ., \beta_m]^T$ And $T = [t_1, t_2, \ldots . t_N]^T$

To solve (7), the ELM adopts a least squares error to get solution$\hat{\beta}$:

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{G}^{\dagger}\boldsymbol{T} \qquad (10)$$

Where $G^{\dagger}$ is the Moore-Penrose generalized inverse of G. Function $g(.)$ is usually unknown, we can incorporate kernel functions in $g(.)$ .This is the so called KEML. The kernel matrix $\mathrm{K} = [\mathrm{K}(\mathrm{x}; \mathrm{x}_1) \quad \ldots \quad \mathrm{K}(\mathrm{x}; \mathrm{x}_\mathrm{N})]^\mathrm{T}$ ($K(.)$ is the kernel function) is introduced into (9) and (10) toestimate the output of the KELM:

$$\mathrm{o} = \mathrm{KT} \qquad (11)$$

Herein, the Gaussian kernel function (RBF) is adopted.

$$K(x_1; x_2) = e^{\left(\frac{-\|x_1 - x_2\|^2}{2\sigma}\right)} \quad (12)$$

In above, σ is the width of RBF. The number of hidden neuron *m* needs to be optimized for KELM. To do so, the mABC is used to optimize *m* in the training processing of the KELM.

3.3.3. The New Method Based on mABC-KELM

The proposed network intrusion detection method can be summarized as follows:

Step 1: Format the intrusion data set into standard form through pre-processing step.

Step 2: Fuse the data using principal component analysis (PCA) to obtain the feature vector. PCA is used to determine a subset of features based on a feature reduction method.

Step 3: The selected principal components called bacteria principal components are the basis of feature subsets. BF is used to search the PCA space for feature selection. In turn, the feature set obtained through this process is presented to the classifier.

Step 4: Train the KELM using the feature vectors, and optimize the hidden neuron number using modified ABC.

Step 5: Test the performance of the ABC-KELM detection model. A workflow block of the proposed mABC-KELM intrusion detection method has been mentioned.

4. EXPERIMENTAL RESULTS

In this section, the proposed mABC-KELM performance is evaluated and compared with existing with existing intrusion detection algorithms such as Fuzzy Neural Network with Expectation Maximization (FNN-EM), Genetic Algorithm based IDS (GA-IDS) [27] and EAACK [28] in presence of malicious node environment. The ADFA-LD consists of 833 normal traces for training the IDS, 4373 normal traces for evaluating FAR and 60 different attack sets, each consisting of multiple traces. Each DE method was trained using the same set of normal traces, with false alarm rates calculated by then processing a separate set of normal traces and calculating the number of alerts. The attack traces were then classified, with detection rate calculated from the number of alerts arising from this assessment.

The proposed IDS is simulated with Network Simulator tool (NS 2.34). In proposed simulation, 101 sensor nodes move in a 1000 meter x 1000 meter square region for 100 seconds simulation time. Here, each node moves independently with the same average speed. Every one of nodes has the same transmission range of 250 meters. The simulated traffic is Constant Bit Rate (CBR). Proposed simulation settings and parameters are summarized in table 1.

Performance evaluation

The given below parameters are used to evaluated the proposed scheme.

Detection Rate Comparison

The graphical representation for comparison of detection rate of proposed mABC-KELM and existing methods such as GA-IDS and EAACK are shown in figure 2. The proposed mABC-KELM method has attained better detection rate than previous schemes because of preprocess. It reduced the attacks.

Table 1
simulation parameters

No. of Nodes	101
Area Size	1000 X 1000
Mac	802.11
Radio Range	250m
Simulation Time	100 sec
Traffic Source	CBR
Packet Size	80 bytes
Mobility Model	Random Way Point
Protocol	LEACH

Communication overhead

The graphical representation for comparison of overheads of proposed mABC-KELM and existing methods such as GA-IDS and EAACK are shown in figure 3. It demonstrates that the proposed mABC-KELM scheme has attained less overhead than existing schemes because the data effective time is small enough to meet the desired event data.

Packet Delivery Ratio (PDR)

The graphical representation for comparison of PDR of proposed mABC-KELM and existing methods such as GA-IDS and EAACK are shown in figure 4. It shows the proposed mABC-KELM scheme has high PDR rate than existing schemes. Because, the packets are transformed with reliable nodes through stable link and successfully delivered to the destination node in proposed scheme.

End to end delay

The graphical representation for end to end delay comparison between proposed mABC-KELM and existing such as GA-IDS and EAACK are shown in figure 5. The proposed mABC-KELM has less delay than existing schemes. To achieve good QOS, the end to end delay should be less in the system. In proposed system, cluster based routing reduced the delay so it attained best routing.

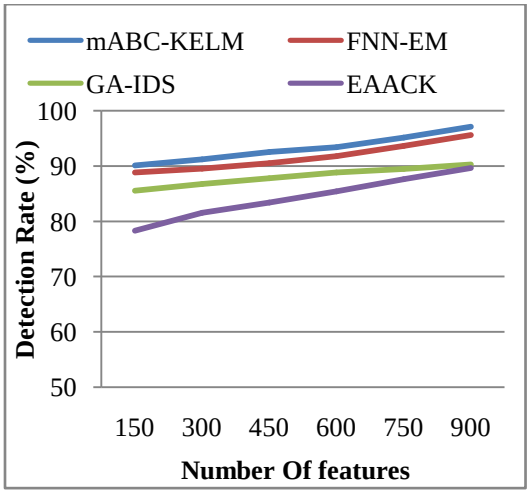

Figure 2: Comparison of Detection Rate

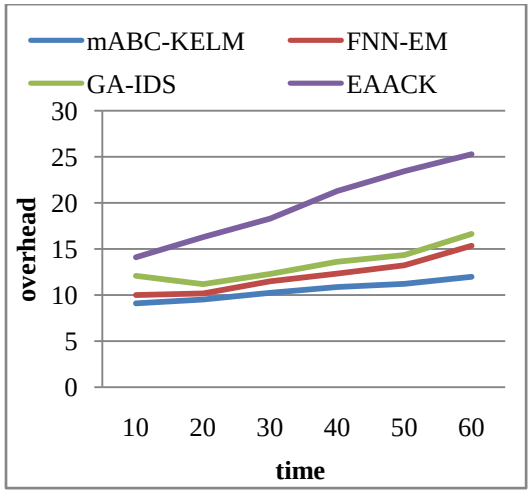

Figure 3: comparison of time vs. overhead

Packet Integrity Rate

The graphical representation for Packet Integrity Rate of proposed mABC-KELM and existing methods such as GA-IDS and EAACK are shown in figure 6. It shows the proposed mABC-KELM has attained better packet integrity than existing schemes because of encryption and decryption mechanism.

Network Lifetime

The graphical representation for Network Lifetime comparison of proposed mABC-KELM and existing methods such as GA-IDS and EAACK are shown in figure 7. The mABC-KELM has attained high lifetime than previous schemes because of adding link stability rate.

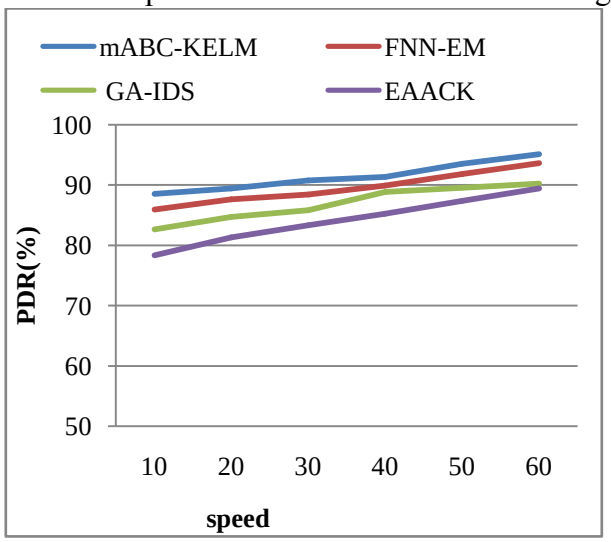

Figure 4: comparison of speed vs. PDR

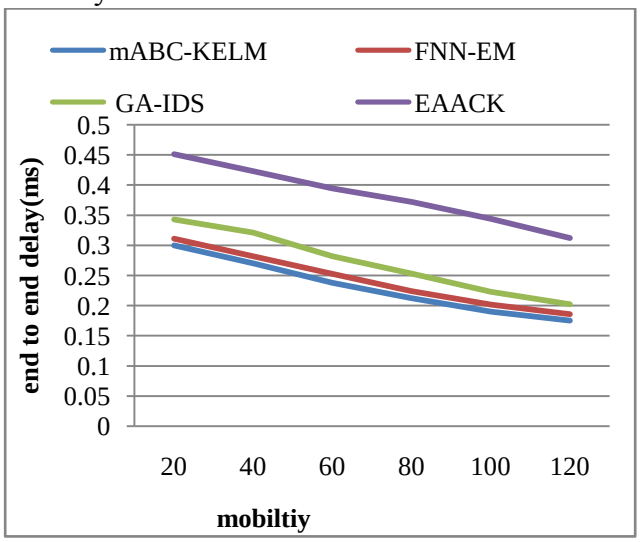

Figure 5: comparison of Mobility Vs End to end delay

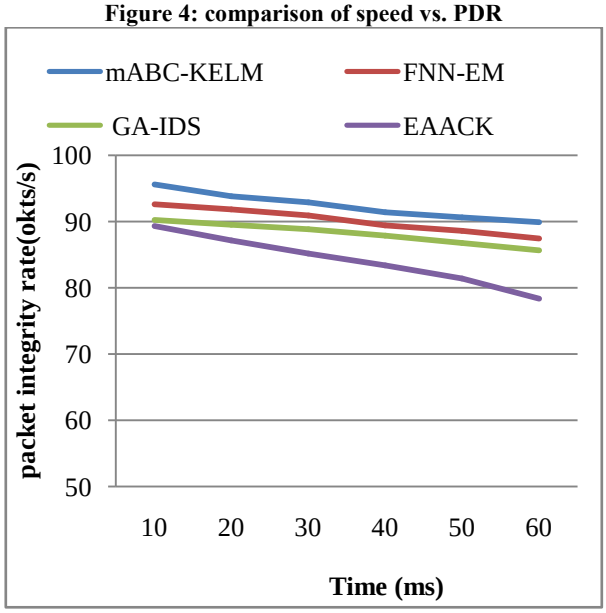

Figure 6: comparison of Time Vs Packet Integrity Rate

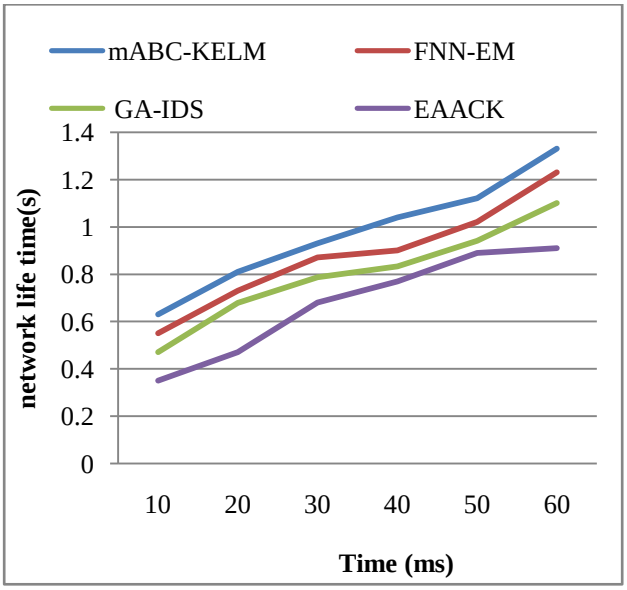

Figure 7: comparison of Pause time Vs Network Lifetime

5. CONCLUSION

Intrusion detection is very important for the computer security. In this research, a performance enhancement model is proposed for intrusion detection system based on an optimal feature subset selection using several bacteria principal components. The feature selection has been accomplished using the techniques of PCA and GA. The selected features subsets are presented to modified ABC optimized KELM. The innovation of this work lies in the development and implementation of the PCA and mABC-KELM in the intrusion detection for the first time. Experimental tests have been carried out to calculate the performance of the new method. The test result has showed satisfactory and effective intrusion detection performance of the proposed ABC-KELM method. In addition, when compared with PCA-ABC-KELM, it proves that the performance of the proposed PCA-mABC-KELM method is superior to its rivals in terms of both detection accuracy and training speed. Thus, the Proposed PCA-mABC-KELM method shows promising applications in the domain of intrusion detection.

REFERENCES

[1] Bace, R. and Mell, P. (2001) Intrusion Detection Systems. NIST Special Publications SP 800, U S Department of Defence, 31 November 2001.

[2] W. Stallings, Cryptography and Network Security Principles and Practices (Prentice Hall, Upper Saddle River, 2006).

[3] B. Rhodes, J. Mahaffey, J. Cannady, Multiple self-organizing maps for intrusion detection, Paper presented at the Proceedings of the 23rd National Information Systems Security Conference, Baltimore, 16–19, 2000.

[4] S. X. Wu and W. Banzhaf, The use of computational intelligence in intrusion detection systems: A review, Applied Soft Computing, vol. 10, pp. 1-35, 2010.

[5] C. Tsang, S. Kwong and H. Wang, Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection, Pattern Recognition, vol. 40, pp. 2373-2391, 2007.

[6] G. Wang, J. Hao, J. Ma and L. Huang, A new approach to intrusion detection using Articial Neural Networks and fuzzy clustering, Expert Systems with Applications, vol. 37, pp. 6225-6232, 2010.

[7] T. Abbes, A. Bouhoula and M. Rusinowitch, Protocol analysis in intrusion detection using decision tree, Inform.Technol. Coding Comput. vol. 1, pp. 404-408, 2004.

[8] K.Y. Chan, C.K. Kwong, Y.C. Tsim, M.E. Aydin and T.C. Fogarty, A new orthogonal array based crossover, with analysis of gene interactions, for evolutionary algorithms and its application to car door design, Expert Systems with Applications, vol. 37, pp. 3853-3862, 2010.

[9] R. B. Dubey, M.Hanmandlu and S. K. Gupta, An Advanced Technique for Volumetric Analysis International Journal of Computer Applications, vol. 1, pp. 91-98, 2010.

[10] A. Abduvaliyev, A.S. K. Pathan, J. Zhou, R. Roman, and W.C. Wong, "On the vital areas of intrusion detection systems in wireless sensor networks." IEEE Communications Surveys and Tutorials, vol. 15, no. 3, pp. 1223–1237, 2013.

[11] A. H. Farooqi and F. A. Khan, "A survey of intrusion detection systems for wireless sensor networks," International Journal of Ad Hoc and Ubiquitous Computing, vol. 9, no. 2, pp. 69–83, 2012.

[12] H. Moosavi and F. Bui, "A game-theoretic framework for robust optimal intrusion detection in wireless sensor networks," 2014.

[13] E. Karapistoli and A. A. Economides, "Anomaly detection and localization in uwb wireless sensor networks," in Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on. IEEE, 2013, pp. 2326–2330.

[14] L. Coppolino, S. DAntonio, A. Garofalo, and L. Romano, "Applying data mining techniques to intrusion detection in wireless sensor networks," in P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on. IEEE, 2013, pp. 247–254.

[15] S. Nishanthi and T. Virudhunagar, "Intrusion detection in wireless sensor networks using watchdog based clonal selection algorithm," 2013.

[16] A. Chittur, "Model Generation for an Intrusion Detection System Using Genetic Algorithms," High School Honors Thesis, Ossining High School, Ossining, NY, 2001.

[17] L. De Castro and F. Von Zuben, "Learning and Optimization Using the Clonal Selection Principle," IEEE Transactions on Evolutionary Computation, vol. 6, no. 3, pp. 239–251, 2002.

[18] P. A. Diaz-Gome and D. F. Hougen, "Improved Off-Line Intrusion Detection Using a Genetic Algorithm," in Proceedings of the Seventh International Conference on Enterprise Information Systems, Miami, USA, 2005.

[19] T. Xiao, G. Qu, S. Hariri, and M. Yousif, "An Efficient Network Detection Method Based on Information Theory and ISSN : 2028-9324 Vol. 5 No. 3, Mar. 2014 240 Genetic Algorithm," in Proceedings of the 24th IEEE International Performance Computing and Communications Conference, USA, 2005.

[20] Liu G., Zhang Y.I., Yang S. (2007). A hierarchical intrusion detection model based on the PCA neural networks. Neurocomputing, 70(7-9): 1561-1568.

[21] Horng S.J., Yang S.U., Chen Y.H., Kao T.W., Chen R.J., Lai J.L., Perkasa C.D. (2010). A novel intrusion detection system based on hierarchical clustering and support vector machines. Expert Syst. Appl., 38(1): 306-313.

[22] Tong X., Wang Z., Haining Y.U. (2009). A research using hybrid RBF/Elman neural network for intrusion detection system secure model. Comput. Phys. Commun., 180(10): 1795-1801.

[23] I. Ahmad, A. Abdullah, and A. Alghamdi, M. Hussain, "Optimized Intrusion Detection Mechanism Using Soft Computing Techniques", Telecommunication Syst., 52, 2187-2195 (2013)

[24] D. Karaboga, "An Idea Based On Honey Bee Swarm for Numerical Optimization", Technical Report TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.

[25] T.D. Seeley, "The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies", Harvard University Press, 1995.

[26] Huang G., Chen L. Enhanced Random Search Based Incremental Extreme Learning Machine. Neurocomputing. 2008; 71: 16–18.

[27] Nirmaladevi, P., & Tamilarasi, A. (2013). An Efficient Intrusion Detection System Based on GA to Recognize Attacks in User Privileges. International Review on Computers and Software (IRECOS), 8(8), 1917-1922.

[28] M.Shakshuki, Nan Kang, T.R.Sheltami, " EAACK – A Secure Intrusion Detection System for MANETs", IEEE Transactions on Industrial Electronics, Vol. 60, Issue 3, 2013, pp. 1089-1098.