Intrusion Detection System Attack Classification with Optimization Model for WSN Security

Abidullha Adel, Sohel Rana, Jayastree

Abstract: ***Wireless Sensor Network (WSN) subjected various challenges during data transmission between nodes deployed in a network. To withstand those security challenges Intrusion Detection System (IDS) is designed. IDS is involved in attack detection and classification but is subjected to a lack of effective classification techniques for attack prevention. To overcome those challenges associated with security this research presented an effective clustering technique known as Centred-Order Node Clustering (CONC). Also, Cluster Head (CH) is elected based on the Improved Flower Pollination Algorithm (IFPA) with multi-objective characteristics. By this proposed method lifetime of the network is improved. Additionally, a supervised classification technique called AdaBoost Regression Classifier (ABRC) is developed with the Intrusion Detection System (IDS). The developed ABRC is constructed for malicious node detection with the prediction of several attacks using IDS. Through improved security mechanisms sensor nodes are involved in effective data transmission between sensor nodes. The simulation analysis stated that the proposed mechanism provides better results rather than the existing technique.***

Keywords: Wireless Sensor Network (WSN), Intrusion Detection System (IDS), Clustering, AdaBoost Regression Classifier (ABRC), Centred-Order Node Clustering (CONC), Improved Flower Pollination Algorithm (IFPA).

I. INTRODUCTION

In a computer system, security is considered a major issue due to the dynamic and decentralized ad-hoc infrastructure. Ad-hoc networks are self-organized networks without any fixed security scheme which increases the security risk of the network [1]. Even though massive security schemes were developed for Ad-hoc networks those are all not effective for guaranteed security [2]. However, those security schemes are not effective for a dynamic network. Among various ad-hoc Wireless Sensor Network (WSN) provides an infrastructure-less, dynamic, and distributed network [3]. Due to WSN's vulnerable characteristics network is always subjected to several security threats that impact whole network functionality. WSN network provides secure data transmission with the implementation of the authentication protocol and secure routing protocol for data security for insider or passive attacks. Even those protocols are subjected to several challenges in their functionality due to the limited energy and mobility of nodes. The data received from intruders cause passive attack in-network but inside node attack cannot be eliminated. WSN networks are subjected to several types of security attacks such as routing attacks, Sybil attacks, and Denial of Service (DoS) so on. In the WSN network system, four major types are existing those are Denial of Service (DOS), User to Root (U2R), and probe attack which relies on OSI layers of the WSN network. The WSN network comprises 4 layers such as a physical layer, data link layer, network layer, and transport layer. DoS attack occurs in both the physical and network layer. The attack is stated as DoS if it satisfies three conditions such as [4]:

i. Selective data forwarding - based on the pre-defined criteria packets are dropped selectively.

ii. Tampering - without any encryption scheme tampering will occur.

iii. Jamming - interference of frequencies due to network nodes.

Another attack U2R occurs mainly in the network layer of WSN. The U2R attack in WSN is identified whenever an illegal node transmits Hello flood requests to any legitimate node. In the network layer, another attack arises which is defined as R2L. Usually, the attack is stated as R2L when the network is subjected to Sybil attack, wormhole attack, spoofing, and R2L attack [5]. Further, within the network layer, another attack arises such as a probing attack that spoofed routing information or altering path for information routing or replay routing information or sinkhole attack [6]. To detect suspicious activity in WSN network Intrusion Detection System (IDS) is utilized within the WSN [7]. The cluster-based WSN network minimizes the load of the WSN node with a reduction of aggregate computation and energy consumption [8]. Due to technological advancement, WSN has been widely used in a vast range of applications this leads to security challenges in a network focused on reliable node performance of the network. IDS focused on the detection of malicious activity within the node and protect the whole network from preventing malicious activity in nodes. The incorporated IDS agents collect abnormal characteristics of nodes and analyze with the period for performing appropriate action [9].

Manuscript received on October 03, 2021.
Revised Manuscript received on October 11, 2021.
Manuscript published on October 30, 2021.
* Correspondence Author

Abidullha Adel*, Assistant Professor and Lecturer in Kunduz University. Afghanistan

Md. Sohel Rana, Department of Computer Science and Engineering from Daffodil International University of Bangladesh

Jayastree. J, B.Tech, Electronics and Communication Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology. Chennai, Tamil Nadu

The IDS agents are deployed in WSN in three possible ways such as centralized, distributed, and hybrid. Those agents are effective within the deployed base station, the centralized approach does not impact small node performance [10]. Situation awareness (SA) is considered a basic element due to space and time for acting as the network. In a cluster-based network implementing security factors in Cluster Head (CH) is considered beneficial due to a centralized approach for addressing security mechanisms against several threats. The CH collects information about the behavior of the node to determine the operation. It incorporates collected information of nodes in Knowledge-based systems (KBSs) for gathering and data storage in the symbolized form at different scenarios [11]. A created KDSs provides the scheme to overcome security threats that occurred from internal and external intruders [12]. Intrusion within the network is detected due to the presence of unknown attacks and anomalous activity of the rooted unknown threats in WSN. Those unknown attacks can be denial-of-service, wormhole attack, botnet, malware attack, and so on [13]. Anomaly in the WSN network is detected by IDS based on malicious traffic patterns, identification of anomalies from unknown attacks through learning obtained from historical data [14]. IDS must identify known attacks and provides realized robustness for the unknown anomaly in the network. For threat, detection IDS is considered an essential part of a dynamic network with limited energy levels. The limited energy level of WSN affects the lifetime of the network. In case, if threats or attacks are incorporated within WSN it consumes a huge amount of energy which limits the WSN network lifetime [15]. IDS are designed in such a way to detect malicious activity or policy within the network for anomalous detection of attack in the network. In existing, IDS is implemented with conventional statistical learning methods such as Naive Bayes, Decision Tree, Support Vector Machine (SVM), and Random Forest [16]. Due to the remarkable performance of deep learning algorithms in IDS, several studies adopt a neural network for anomaly or threat detection. The deep learning mechanism utilized is Multilayer Perceptron, Convolutional Neural Network, and Recurrent Neural Network [17]. This paper proposed an efficient classification technique for improving data transfer security in WSN. The proposed classifier is stated as the AdaBoost regression classifier (ABRC) for detection in the network. The proposed ABRC deployed in Intrusion Detection System (IDS) to provide high-level security. Initially, this research employs Cantered-Order Node Clustering (CONC) for the clustering of nodes. To select a cluster head (CH) Improved Flower Pollination Algorithm (IFPA) is employed. Through the developed approach it is expected that the proposed scheme provides a high-level security scheme. The comparative analysis expressed that the proposed algorithm provides significant performance rather than an existing technique. This paper is organized as follows: In section 2 presented exiting literature related to CH selection and classification. In section 3 illustrated about overall research methodology followed by the description of individual methods. In section 4 presented performance metrics and simulation results are presented. The performance of the proposed approach is comparatively examined with an existing technique. Section 5 presented the overall conclusion of the research followed by references.

II. RELATED WORKS

Researcher, [18] developed an integrated algorithm for energy-efficient clustering and routing in WSN networks. The proposed algorithm uses particle swarm optimization (PSO) for Linear/Nonlinear Programming (LP/NLP) formulations. The proposed algorithm performance is based on the multi-objective fitness function for the efficient encoding of WSN. The developed clustering is based on the conversion of node energy with load balancing schemes. Another researcher [19] developed a swarm-based intelligence for effective cluster head election in the WSN network. The developed algorithm is based on LEACH-based clustering with a modified Ant Colony Optimization algorithm. Through the deployed optimization algorithm residual energy of nodes is estimated with consideration of different nodes in the network. Energy-Harvesting Stable Election Protocol (EH-SEP) heterogeneous EH-WSN environment based on SEP which is termed as EH-SEP. This belongs to the class of clustering protocol for energy harvesting. Moreover, the nodes with higher residual energy probability are selected as CH. Data transmission in a multi-hop scenario is utilized for energy balancing in the network. In [20], novel Energy Efficient Dynamic Algorithm with dynamic clustering in a distributed manner. The variation of energy level in each node is comparatively examined with nearby nodes. The objective is to reduce energy consumed with increased network efficiency. In [21], Energy Neutral Clustering (ENC) with CH group (CHG) mechanism. The cluster is involved in construction CH for sharing various traffic loads. However, control overhead is involved in minimal information sharing with other protocols. Some researchers [22] examined the performance of the existing intrusion detection scheme. In research conducted by [23] Naive Bayes based intrusion detection system for WSN security. Another researcher [24] adopted Random Forest (RF) based automatic intrusion detection system for comparative performance analysis. Additionally, in IDS other classification techniques such as SVM are exhibit improved accuracy of the system with improved lifetime. To improve the detection accuracy of the system [26] applied Least Square SVM (LSSVM) for Mutual Information-based Feature Selection (MIFS) for improving detection efficiency. However, conventional techniques are limited for limited data flow within the network for imbalanced data that technique performance is degraded. To improve the performance of the IDS system deep learning-based schemes are employed based on the neural network [27]. The neural network with Multilayer Perceptron (MLP) provides significant performance with the construction of various network layers that exhibit the applicability of deep learning intrusion detection. In an intrusion detection system, Convolutional Neural Network (CNN) provides superior performance rather than a conventional technique for performance improvement [28]. Another researcher, [29] uses CNN with Recurrent Neural Network (RNN) based on consideration of local features and sequential correlation estimation.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Through a random search neural network, another researcher [30] random neural network-based intrusion detection system is adopted with superior performance characteristics rather than traditional techniques. Another researcher [31] adopted SVM, RF, and extreme learning machines (ELM) in the intrusion detection system. The simulation results exhibited that performance of ELM is significantly higher.

III. PROPOSED ATTACK CLASSIFICATION METHOD

Security is considered a major factor in any wireless communication system. In WSN, the energy of the network is limited due to the battery power supply which impacts the overall lifetime of the network. Attacks within the network affect the overall performance of the WSN network and consume more energy which reduces the network lifetime. To improve the security of the WSN network with minimal energy utilization of data transfer IDS is employed. IDS is involved in the identification and prevention of attacks in WSN. The information from different sources is examined for security breaches for both insider and outsider attacks, misuse, and vulnerability assessment of the network.

At present, WSN is subjected to security issues during data transfer from sensor networks to others. To overcome those issues, an effective classification technique ABRC is proposed. Initially, a group of sensor nodes is deployed for clustering and CH election form improving the lifetime of the network with improved network scalability. Here, clustering in WSN is achieved through Center-ordered node clustering (CONC) concerning the weights of nodes. Through the formation of clustering, this research adopts the CH election using a modified Improved Flower Pollination Algorithm (IFPA). After, the election of cluster head (CH), a developed AdaBoost Regression Classifier (ABRC) is employed for the reduction of databases and classification. The proposed classifier technique uses AdaBoost and Regression classifier to detect whether the node has an attack or not. In the second stage, sensor nodes that all have the possibility of attack are detected based on given conditions. This research considers four attacks such as DoS, U2R, Probe, and R2L for attack detection to improve security in WSN. The overall flow of the proposed technique is illustrated in figure 1.

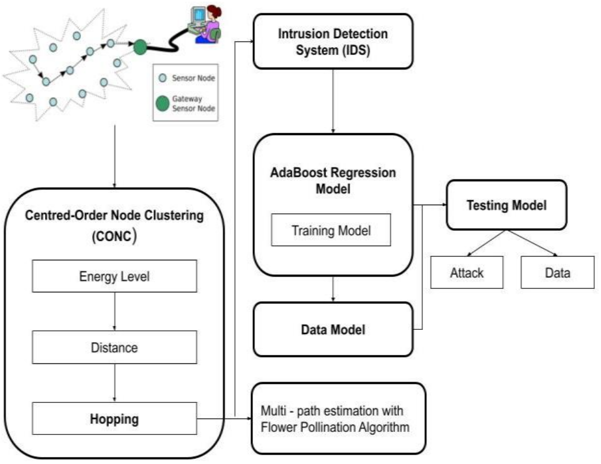
Sensor Node
Gateway Sensor Node
Intrusion Detection System (IDS)
AdaBoost Regression Model
Training Model
Testing Model
Attack
Data
Data Model
Centred-Order Node Clustering (CONC)
Energy Level
Distance
Hopping
Multi - path estimation with Flower Pollination Algorithm

Figure 1. Overall Proposed Methodology

3.1 Datasets

This work utilizes the CICIDS 2019 dataset which comprises 14 features that are labeled for differentiation of attack or normal behavior of node. The selected dataset features are categorized into four subcategories such as intrinsic features, content features, similar service features, and host features. The CICIDS 2019 dataset with attack types for testing and training is presented in Table 1.

Table 1: Data Distribution of Attacks

	Distribution of Data	Value
Training Set	Normal	66,479
	Anomaly	59,479
	DoS	44,662
	Probe	13,573
	User to Root (U2R)	54
	Root to Local (R2L)	983
Testing Set	Normal	9,231
	Anomaly	11,683
	Denial of Service (DoS)	7,268
	Probe Attack	2,268
	User to Root (U2R)	63
	Root to Local (R2L)	2,657

In the U2R attack, attackers utilize normal user account and gain information related to systems. The U2R attacks targeted password sniffing, dictionary attack, and social engineering. Hence, for testing and training, a minimal range of U2R was selected. Table 2 presented about attributes of selected features of CICIDS 2019 datasets were presented.

Table 2. Attributes of CICIDS 2019

Name of Attribute for CICIDS2019 dataset	Description of attributes in CICIDS2019 dataset
Src IP	Network IP address
Src Port	Address of port
Dest IP	The IP address of the destination
Dest Port	Destination
Proto	Transport Layer
Date first seen	Initial data flow in the network
Duration	The complete duration of data transmission
Bytes	Data count those are transmitted
Packets	Total packed transferred
Flags	TCP flags c
Class	Classification of label whether it is attack, normal and suspicious
Attack Type	Attack type
AttackID	Estimation of attacker id
Attack Description	Evaluation of network attack

Based on the assigned attributes dataset is classified and evaluated for the attack in the WSN network.

3.2 WSN Clustering based on Centre-ordered node clustering (CONC)

This paper involved the construction of WSN for clustering of nodes for effective achievement of data transmission through Centred-Ordered Node Clustering (CONC). With the estimation of weights for estimated values of each node, weights are evaluated with the network function Upon the estimation of WSN weights and hopping data loss within the network is computed. The sender and destination node weightage functions are computed based on the computation of distance probability between nodes and hopping between sender and receiver. Within the clustered network to perform effective communication between nodes middle-order cluster head. The transmission of data between sender and receiver is computed based on the weighting function for estimation of network efficient performance based on probabilistic function. Additionally, network improvement is enhanced through reduction of data loss and increase in the lifetime of the network. Also, the performance of the network is increased and minimizes the data loss for improving the lifetime of the cluster head. This paper developed a CONC-based technique for clustering a node with the estimation of node position and location. To compute node position and location within the node dimensionality reduction computation is performed using the covariance matrix. Consider the covariance matrix $S_{m*n} \in R^{m*n}$ in which observation and variable operation are represented as 'm' and 'n'. The standardized matrix for the network is the normalized value of a network is given as zero with the elimination of biased values and principal weights concerning observation. Also, the estimated sample vector $C = E\{S(m)S^T(m)\}$ covariance matrix is computed based on n*n. The matrix of unknown values are measured with n*n with covariance value of μin which C is computed based on given equation (1)

$$\lambda\mu = C\mu \tag{1}$$

In equation (1), the computed eigenvalues are denoted as λ. The node location and position are evaluated based on eigenvalues estimated using equation (2)

$$f(m) = \frac{1}{m}\sum_{i=1}^{m} S(i)S^T(i)\,\frac{f(i-1)}{\|f(i-1)\|} \tag{2}$$

Similarly, in equation (2) node estimation steps are represented as f and node position, and time is denoted as S. The data transmission within the network for the first data transmission is denoted as in equation (3)

$$f(m) = \frac{m-1}{m}S(m-1) + \frac{1}{m}S(m)S^T(m)\,\frac{f(m-1)}{\|f(m-1)\|} \tag{3}$$

The $\frac{m-1}{m}$ is stated as the final estimation of values and $\frac{1}{m}$ provides the new location.

With the establishment of MONC location and position of the node is evaluated based on the weighted probability with weighted factor as represented in equation (4) stated as follows:

$$w(i) = S(i)S^T(i)\,\frac{f(i-1)}{\|f(i-1)\|} \tag{4}$$

The node weights positive values are represented in equation (5) and equation (6) for transmission of data is represented as follows:

$$f(m) = \frac{m-1-l}{m}S(m-1) + \frac{1+l}{m}S(m)S^T(m)\,\frac{f(m-1)}{\|f(m-1)\|} \tag{5}$$

$$S_2(m) = S_1(m) - S_1^T(m)\,\frac{f_1(m)}{\|f_1(m)\|}\,\frac{f_2(m)}{\|f_2(m)\|} \tag{6}$$

In the above equation (7)$S_1(m) = S(m)$ the term $S_2(m)$ over iterative steps provides the residual energy. The overall node data is computed as in equation (7)

$$f(m) = \frac{m-1}{m}f(m-1) + \frac{1}{m}S(m) \tag{7}$$

With consideration of equation (7), the data transmission is computed concerning the energy and position of the node.

3.2 Cluster Head Election based on Improved Flower Pollination Algorithm (IFPA)

In this paper, the CH election is subjected to a multi-label optimization problem. The fitness function is shown in equ (8).

$$\chi^* = min\,P(\chi) \tag{8}$$

$$P(\chi) = I(\chi) + a_1 S(\chi) \tag{9}$$

Where χrepresented as label function for the assigned variable at each point in the cluster$\Psi$to a label$L$. From that, optimized values of nodes are represented in equ (10).

$$\chi : q \in \Psi \rightarrow \chi(q) \in \Re \tag{10}$$

The flower pollination algorithm (FPA) is considered a natural process of a multi-objective optimization algorithm. This has been widely adopted in a vast range of multi-objective optimization processes. This performance of FPA is categorized into phases such as biotic or cross-pollination and biotic or self-pollination for the reproduction process. In the case of the cross-pollination process pollens transfer is considered for bees, bats, butterflies, and beetles for longer distances. In the case of self-pollination, pollens are transferred through water or wind for a shorter distance. The flower or solution for FPA is stated in equation (11).

$$P = \begin{bmatrix} \overrightarrow{Y^1} = [y_1^1 y_2^1 \dots y_{DV}^1] \\ \overrightarrow{Y^2} = [y_1^2 y_2^2 \dots y_{DV}^2] \\ \vdots\ \vdots\ \vdots\ \ddots\ \vdots \\ \overrightarrow{Y^m} = [y_1^m y_2^m \dots y_{DV}^m] \end{bmatrix} = \begin{bmatrix} F\left(\overrightarrow{Y^1}\right) \\ F\left(\overrightarrow{Y^2}\right) \\ \vdots \\ F\left(\overrightarrow{Y^m}\right) \end{bmatrix} \tag{11}$$

The above equation m is denoted as a solution number or flower population stated as $(\overrightarrow{y^1}, \overrightarrow{y^2}, \dots \overrightarrow{y^m})$ and DVrepresents the number of the decision variable· The fitness function $k = \{1, 2, \dots, m\}$ is denoted as$F(\overrightarrow{Y^k})$. The characteristics of Levy flights with global search characteristics are represented as biotic or cross-pollination, with the updated Levy flights function described in equation (12).

$$y_{i+1}^t = y_i^t + \gamma L(y^* - y_i^t) \tag{12}$$

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The iteration value is represented as$i$, and solution optimal solution is denoted as y_i^t for solutiony^*. The iteration pollination values with Levy distribution are denoted with scaling factor L andγ.

$$L \sim \frac{\chi\Gamma(\chi)\sin(\pi\chi/2)}{\pi}\frac{1}{R^{1+\chi}} \qquad (13)$$

In equation (13), Γ presented about standard gamma function with step distribution of$R > 0$. The gamma function estimation is based on consideration of two random solutions with consideration of local procedures with abiotic and self-pollination. The equ (11) presents FPA local pollination value.

$$y_{i+1}^t = y_i^t + s_1 \times (y_i^j - y_i^k) \qquad (14)$$

Here, two solutions were selected randomly with consideration of various flowers of the same species y_i^j andy_i^k. The value is distributed evenly between the interval 0 and 1 i.e. [0, 1]. Through an efficient search strategy of exploitation and exploration efficiency of search within the algorithm is improved. Moreover, FPA exploration with the current solution is effectively utilized for achieving new solutions. The FPA algorithm convergence speed is significantly improved with the exploration of the search process. The selection of search space bounded through heuristic approach and general global pollination generates search space problem in Improved Flower Pollination Algorithm (IFPA). The information selected from similar parents is narrowed with heuristics-based bounded search space with consideration of space area, which is stated in equ (15).

$$y_{i+1}^{t_a} = \left(max\left(y_i^{j_a}, y_i^{k_b}\right) - min\left(y_i^{j_a}, y_i^{k_b}\right) \cdot s_2 + min\left(y_i^{j_a}, y_i^{k_b}\right)\right) \qquad (15)$$

Through equation (15), the $i$iteration with t^{th} and a^{th} variable are denoted as$y_i^{t_a}$. The randomly selected solutions y_i^a and y_i^b with s_1 and s_2 in the uniform interval are denoted as 0 and 1. Based on population heuristics-based search space is evolved for search space evaluation. With an exploration of search space, global pollination needs to maintain a path with local minima values. The IFPA steps are explained in Algorithm 1.

Algorithm 1: Improved Flower Pollination Algorithm (IFPA)

```
Start
Initialization of FPA parameters such as number of flowers$(m)$,
    switching probability$(SP)$
Initialization of $m$flower population $(P)$
    $[P = \overrightarrow{y^1}, \overrightarrow{y^2}, \dots \overrightarrow{y^m}]$
Examine the optimal solution $y^* \in P$
$i \leftarrow 1$
do
    for every $t \in (1, m)$
        if $V(0,1) < SP$
            for every $a \in (1, DV)$ do        \*Global pollination
                if $V(0,1) < 0.5$
                    $y_{i+1}^{t_a} = y_i^{t_a} + \gamma L(y^* - y_i^{t_a})$
                else
                    $y_{i+1}^{t_a} = \left(max\left(y_i^{j_a}, y_i^{k_b}\right) - min\left(y_i^{j_a}, y_i^{k_b}\right) \cdot s_2 + min\left(y_i^{j_a}, y_i^{k_b}\right)\right)$
                    Where $j, k \in (1, m), j \neq k$
                end if
            end for
        else
            $y_{i+1}^t = y_i^t + V(0,1) \times (y_i^l - y_i^n)$        \* Local pollination
            Where$l, n \in (1, m), l \neq n$
        end if
        if $(Fy_{i+1}^t) < F(y_i^t)$ then
            update $y_i^t$with $y_{i+1}^t$ then
        end if
        if $(Fy_{i+1}^t) < F(y^*)$ then
            update $y^*$with $y_{i+1}^t$ then
        end if
    end for
    $i \leftarrow i + 1$
met $i > i_{max}$
End
```

Initially, WSN node parameters are fed into IFPA for evaluation of network parameters and to elect cluster head (CH) of the network. The Improved Flower Pollination Algorithm (IFPA) is utilized as an optimization model for the determination of Cluster Head with minimal energy utilization. Through iterative optimization, CH is elected in the WSN network with an objective function of improved security and reduced energy utilization of improved network lifetime. The proposed model for the CH election using IFPA is presented in figure 2**.**

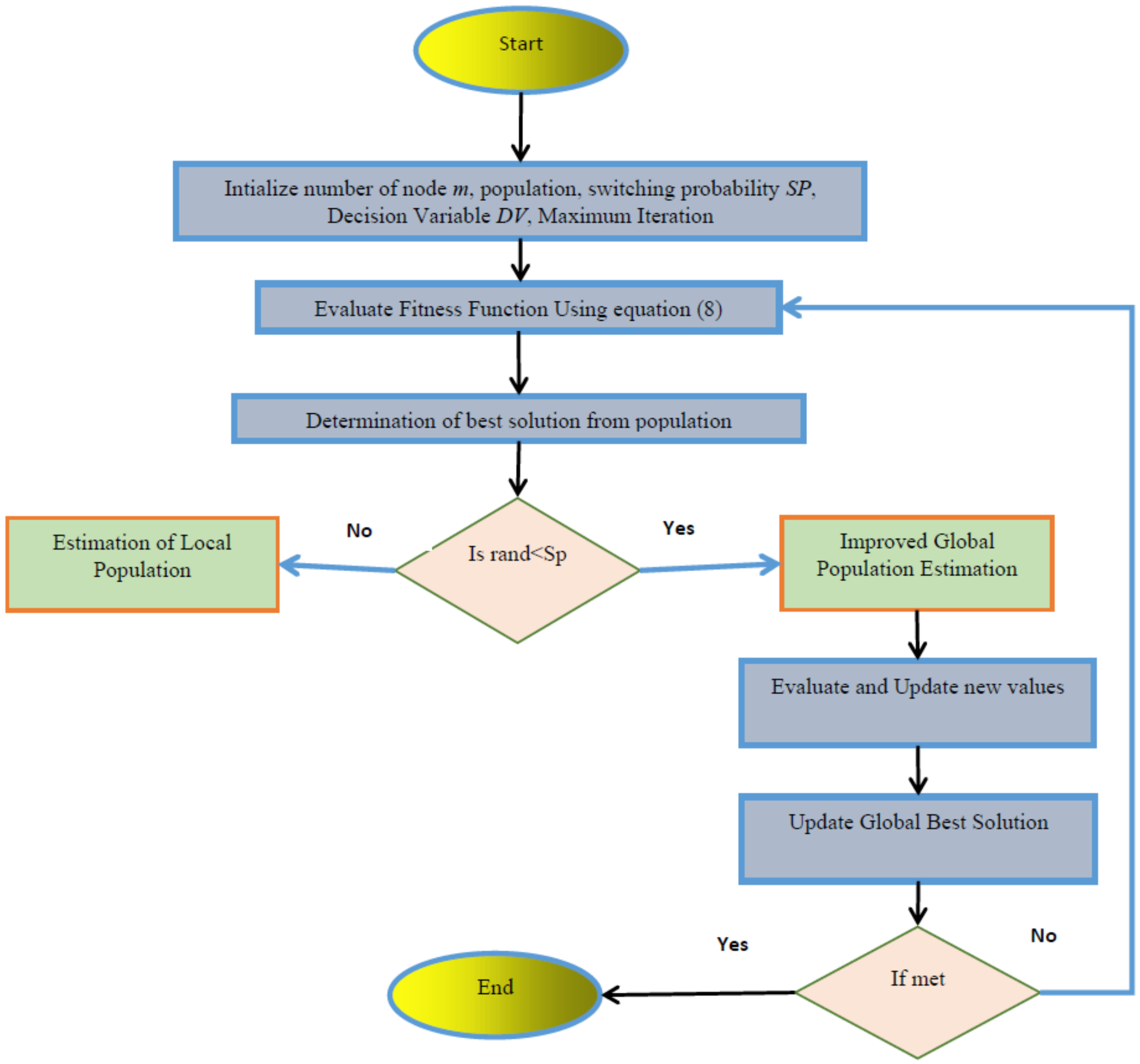
Start
Intialize number of node *m*, population, switching probability *SP*, Decision Variable *DV*, Maximum Iteration
Evaluate Fitness Function Using equation (8)
Determination of best solution from population
Is rand<Sp
No
Yes
Estimation of Local Population
Improved Global Population Estimation
Evaluate and Update new values
Update Global Best Solution
If met
Yes
No
End

Figure 2: Overall Flow of IFPA

3.3 Attack classification using Sigmoidal AdaBoost Regression Classifier (ABRC)

This research developed a sigmoidal-based classification technique for improving security in the IDS system specifically for the WSN network. The proposed approach integrates both Adaboost and regression classifier for improving security in IDS of WSN. Logistics regression is involved in the conversion of mismatched data into either 0's and 1's to improve the accuracy of the classifier. In the case of the AdaBoost scheme, it eliminates all unwanted data from the network. To withstand continuity in the network this research uses a sigmoidal approach for the elimination of limitations associated with the classifier. The general steps involved in the proposed sigmoidal AdaBoost regression classifier are presented as follows:

The attacks in the network are classified based on the AdaBoost classifier with the estimation of the weak learner. Using binary classification training is a performance for a strong classifier through estimation of error. With consideration of weak learner accuracy of classification in AdaBoost algorithm with consideration of decision tree at different levels. The estimation of attacks for classification is presented as in equation (16) as follows:

$$H = sign\left(\sum_t \alpha_t h_t(x_t)\right) \tag{16}$$

The developed AdaBoost classifier is applied with an integrated logistic regression model for multiclass performance. The prediction values are estimated with logistics function between the values of 0's and 1's. The logistic regression values computed are stated on equation (17)

$$h_\theta(x) = g\left(\frac{1}{1+e^{-\theta T_x}}\right) \tag{17}$$

The developed integrated classified is evaluated for classification of attack based on the consideration of regression classifier(18):

$$v = \omega^T x \tag{18}$$

The classifier model aimed to derive the sigmoidal function between the value range of$(-\infty, \infty)$, by logistics characteristics, the sigmoidal function is derived using equation (19),

$$1 + e = \frac{1}{e} \qquad (19)$$

The estimation of classifier model derivatives is presented in equation (20) - equation (24). Based on consideration of regression function characteristics probability ranges are defined in equations (20) and (21):

$$P = a_0 + a_1x_1 + a_2x_2 + \ldots\ldots + a_kx_k \qquad (20)$$

$$\left[\frac{P}{(1-P)}\right] = b_0 + b_1x_1 + b_2x_2 + \ldots\ldots + b_kx_k \qquad (21)$$

Applying log on both sides of equation (21)

$$log\left(\frac{P}{1-P}\right) = log(\omega^T x) \qquad (22)$$

In the above equation, applying exponential property

After applying natural exponential property it is stated as in equation (23),

$$log\left(\frac{P}{1-P}\right) = \sum b_jx_j \qquad (23)$$

Where $P = \sum b_ix_i$ is defined as a logistics regression function presented in equation (24)

$$P = \frac{exp(b_jx_j)}{[1+exp(b_jx_j)]} \qquad (24)$$

Based on chain rule and maximum likelihood relation classification model is estimated as in equation (25) – equation (29):

$$F'(x) = F'g(x)\,g'(x) \qquad (25)$$

$$P = P(k)(1 - P(k)) \qquad (26)$$

For P maximum likelihood estimation is,

$$\hat{l}(\theta; x) = \frac{1}{n}\sum_{i=1}^{n} ln\, f(x_i|\theta) \qquad (27)$$

$$P = \sum log\, P(k_i) + \sum log(1 - P_i(k_i)) \qquad (28)$$

Elimination of negative term which provides the equation (29),

$$\sum P = \sum P_i \qquad (29)$$

Now, $P = (a_0 + a_1x_1 + a_2x_2 + \ldots\ldots + a_kx_k)\sum P_i$

The developed model computed classification model for the attack classification model is presented in equation (30):

$$H = Sigmoid\left(\sum_{j=1}^{N} P_i a_t h_t(x)\right) \qquad (30)$$

With consideration of equation attack classification is performed in the WSN network.

IV. RESULTS

This paper focused on improving the security scheme in WSN with the integration of clustering and CH election with attack classification. Initially, CONC is adopted for WSN node clustering with the CH election using the Improved Flower Pollination Algorithm (IFPA). Through the integration of both algorithm performances of WSN for data, the transmission is significantly improved. The performance of the *proposed* technique significantly minimizes energy consumption with reduced time consumption with an effective CH selection mechanism. The CH is selected based on IFPA with consideration of weighted factors of nodes. After estimation of clustering and CH election ABRC a supervised classifier is adopted for high-level security. The proposed methodology is implemented in MATLAB 2019 and the system configuration is listed below in Table 3.

Table 3: Simulation Parameters

Parameters	Values
Operating System	Windows 7
Processor	Intel Core i7
RAM	4GB
Node Deployment	Random
Number of Nodes	60
Area	100x100m^2

4.1 Performance Metrics

Initially, sensor nodes are deployed and estimation of parameters is performed with varying different node sizes such as 20, 40, 60, 80, and 100. The parameters measured for analysis are energy consumption, end-to-end delay, packet delivery rate (PDR), and Throughput

4.1.1 Intrusion detection parameters

The intrusion detection parameters within the WSN network were estimated based on the different parameters such as True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Predicted yes value (N), and actual yes value (P). *Those* values are utilized for the prediction of other classification parameters such as accuracy, specificity, sensitivity, False Positive Rate, Negative Prediction value, and Positive Prediction Value. The other parameters measured for estimation of prediction parameters of proposed ABRC are listed as follows:

Specificity: The specificity of the network is considered a statistical method to estimate the performance of the *classification* test. IT provides the measure of accuracy for a specific class of datasets. Generally, the term specificity is defined as a class of achieving negative results when an attack is negative or minimal. It is defined as in (31)

$$Specificity = \frac{TN}{TN+FP} \qquad (31)$$

Sensitivity:

Sensitivity is the statistical method utilized for classification performance estimation. Sensitivity is defined as recall involved in model prediction for instances selected based on consideration of class in the dataset. This offers a prediction of attack possibility within the network. This is stated as a rate of correct event classification among all events in the network and it is given in equation (32)

$$Sensitivity = \frac{TP}{TP+FN} \qquad (32)$$

Retrieval Number: 100.1/ijeat.A31801011121
DOI: 10.35940/ijeat.A3180.1011121
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Accuracy:

Accuracy provides the constructed network WSN performance. It provides a ratio between the number of correctly classified samples to several total sample counts and it is presented in equation (33)

$$Accuracy = \frac{TP+TN}{P+N} \qquad (33)$$

False-positive rate (FPR): To test data FPR is utilized and estimate the ratio between FP values to the sum of false-positive and true negative values which is given in equ (34)

$$FPR = \frac{FP}{\quad} \qquad (34)$$

Fault detection rate (FDR): FDR is for prediction of test data classification based on detection rate prediction to estimate FP ratio based on sum average of false-positive and true positive as stated in equation (35)

$$FDR = \frac{FP}{FP+TP} \qquad (35)$$

Error rate: This provides misclassification for test data classification for error prediction and it is estimated based on the ratio of false-positive value and false negative values which is defined in equation (36)

$$ErrorRate = \frac{FP+FN}{P+N} \qquad (36)$$

Precision: Precision provides the estimated ratio value of True Negative to Negative values and it is given in equation (37)

$$Pr\,e\,cision = \frac{TN}{N} \qquad (37)$$

4.1.2 Attack Detection Parameters

To detect malicious nodes in WSN network attacks are differentiated into known and unknown attacks for the prevention of attacks within the node. The parameters considered for analysis of attack in the network are throughput, End-to-End delay, transmit energy, Packet Delivery Ratio (PDR), and Bit Error Rate (BER).

Throughput

Throughput provides the successful reception of data transmitted through a communication medium. The term throughput is measured using unit bits per second (bits/sec or bps). The throughput is calculated using the below equation (38)

$$\text{Throughput} = \frac{\text{Sum of successful packet} * \text{Avg packet}}{\text{Total Packet}} \qquad (38)$$

End to End Delay

The end-to-End delay provides time taken by the network for transmission of a packet across sender to receiver. It is utilized effectively in the IP network for the transmission of data from sender to receiver. The end to end delay is calculated using equation (39):

$$d_{end-end} = N\left[d_{trans} + d_{prop} + d_{proc} + d_{queue}\right] \qquad (39)$$

In the above equation (36), $d_{end-end}$ describes end-to-end delay, d_{trans} illustrate transmission delay, d_{prop} describes propagation delay, d_{proc} represent processing delay and d_{queue} provides queuing delay.

Transmit energy

In sensor communication, various sink nodes are deployed for data transmission from sender to receiver. The total energy utilized for inter-node communication is calculated using equation (40):

$$TransmitEnergy = NodeWeight \times ExecutionTime \qquad (40)$$

4.2 Simulation Results

This research focused on improving the performance of attack classification in the WSN network. To improve attack detection in WSN with improved security this research uses CONC integrated with IFPA for clustering and CH election. To improve security ABRC classifier is developed. The parameter selected for analysis of the proposed approach is presented. The performance is stated in terms of parameter measurement under consideration of various attacks in the WSN network. The presence of attack in WSN and its performance of network are presented. In table 4 presented about observed throughput value observed for the attack detection environment along with the proposed ABRC mechanism.

Table 4: Parameter Measurement for Throughput

	No.of Nodes	DoS	U2R	R2L	Probe	Unknown Attack	Proposed ABRC
Throughput	20	55.87	57.56	56.45	54.45	53.67	57.89
	40	54.86	54.89	55.24	55.27	54.76	56.96
	60	53.67	57.23	56.87	57.67	57.34	63.37
	80	64.23	63.98	63.25	62.78	61.78	63.78
	100	63.56	65.87	64.78	63.56	62.56	64.45

In figure 3 throughput comparison of varying number of nodes are presented. The performance of proposed ABRC stated that for 60 node the throughput is maximized.

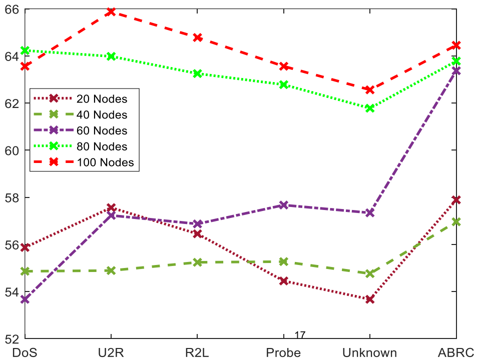

Figure 3: Comparison of Throughput

In table 5 end-to-end delay measurements for the varying number of nodes are presented. The end-to-end delay is measured from the varying number of nodes. In figure 4 comparison of end-to-end delay for varying nodes is presented.

Retrieval Number: 100.1/ijeat.A31801011121
DOI: 10.35940/ijeat.A3180.1011121
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 5: Parameter Measurement for End-to-End Delay

	No.of Nodes	DoS	U2R	R2L	Probe	Unknown Attack	Proposed ABRC
End-to-End Delay	20	1.7456	1.7357	1.7398	1.7432	1.8253	1.8456
	40	1.8563	1.8354	1.7689	1.7753	1.7634	1.8267
	60	1.7856	1.7648	1.8146	1.7984	1.6975	1.8576
	80	1.6972	1.7386	1.7158	1.7895	1.7456	1.8692
	100	1.7592	1.7694	1.7486	1.7695	1.7284	1.8375

In the proposed ABRC approach end-to-end delay measurement is minimal for 40 nodes. The average end-to-end delay measured for DoS, U2R, R2L, probe, and Unknown attacks are significantly minimal for 40 nodes,

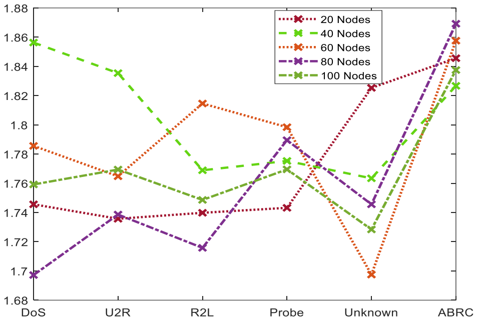

Figure 4: Comparison of End-to-End Delay

In figure 4 for node 60 end-to-end delay is significantly minimal for unknown attack environments. This implies that for 60 nodes the proposed ABRC offers balanced network conditions. In table 6 energy utilized is presented, the analysis of results stated that for the varying number of nodes energy utilization level is significantly higher. The proposed ABRC approach exhibits higher energy utilization level rather than other attack exists in the network.

Table 6: Parameter Measurement for Energy Utilized

	No.of Nodes	DoS	U2R	R2L	Probe	Unknown Attack	Proposed ABRC
Energy Utilized	20	0.1678	0.4268	0.2784	0.3786	0.1894	0.5786
	40	0.1478	0.3675	0.4789	0.1766	0.2789	0.4975
	60	0.2789	0.1789	0.3782	0.2789	0.1784	0.5378
	80	0.1978	0.2486	0.1756	0.4864	0.3756	0.6723
	100	0.3756	0.1894	0.1865	0.3756	0.2894	0.5786

As stated BER is measured for varying numbers of nodes with different attack scenarios with the proposed ABRC scheme. In table 7 comparative analysis of different attack environments with proposed ABRC is presented.

Table 7: Parameter Measurement for BER

	No.of Nodes	DoS	U2R	R2L	Probe	Unknown Attack	Proposed ABRC
Bit Error	20	1.8264	1.8975	1.6756	1.9726	2.1756	1.3782
	40	2.6852	1.9647	1.7895	1.6762	1.9575	1.3156
	60	2.1687	1.8615	2.1675	2.3475	1.4246	1.2974
	80	1.9782	1.7895	1.9354	1.8462	1.8864	1.3984
	100	1.8756	2.0186	2.2734	2.2756	1.8364	1.2592

In figure 5 BER measurement of the proposed ABRC for varying nodes is presented. From the analysis, it is observed that for deployed 60 nodes BER is minimal for an unknown attack scenario.

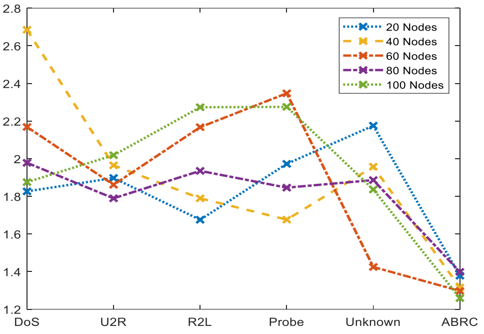

Figure 5: Comparison of BER

The BER analysis stated that the proposed ABRC provides minimal BER value rather than other attack scenarios. In table 8. PDR measurement is presented for the different attacks in the network. The PDR measurement provides the successful reception of the data packet. In figure 6 comparative analysis of varying nodes is presented.

Table 8: Parameter Measurement for PDR

	No.of Nodes	DoS	U2R	R2L	Probe	Unknown Attack	Proposed ABRC
Packet Delivery Ratio (PDR)	20	86.4512	85.3489	79.7512	88.7923	78.6148	92.6745
	40	78.6543	78.6214	82.1134	87.6942	83.6745	94.6722
	60	82.9456	81.9756	88.6755	91.4562	88.9452	93.7561
	80	79.8456	83.9745	91.7594	89.4756	91.5674	95.6812
	100	84.6978	85.7892	78.9463	93.7564	87.6426	96.7161

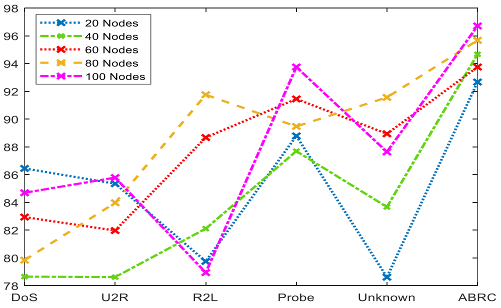

Figure 6: Comparison of PDR

In figure 6 PDR is higher for 60 nodes for probe attack environment. This implies that the proposed ABRC offers significant performance for 6 nodes especially for probe and Unknown attacks. The PDR analysis stated that the proposed ABRC provides a significant improvement in PDR. The increase in PDR implies that the security scheme of the proposed ABRC is significantly improved. In table 9 comparative analysis of proposed ABRC parameters is presented with existing techniques.

Table 9: Comparative Analysis of ABRC

Methods	Sensitivity	Specificity	PDR	Accuracy
GA	0.2	0.95	0.03	0.82
PSO	0.8	0.984	0.35	0.89
ABC	0.75	0.983	0.2	0.93
Proposed ABRC	0.85	0.997	0.39	0.96

In table 9 overall performance of the proposed ABRC approach with the existing technique is presented. The cluster head selection is comparatively examined with GA, PSO, and ABC is performed. The comparative analysis illustrated that the proposed approach exhibits improved performance. In figure 7 and 8 comparative illustrations of sensitivity and specificity of proposed ABRC with existing techniques are presented.

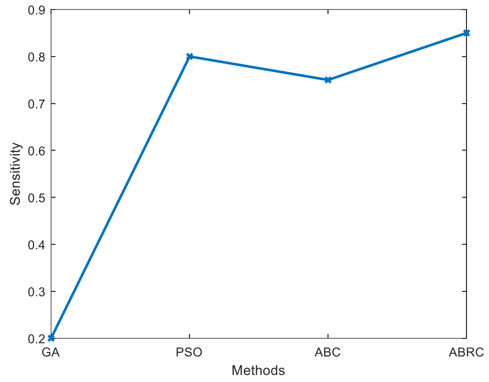

Figure 7: Comparison of Sensitivity

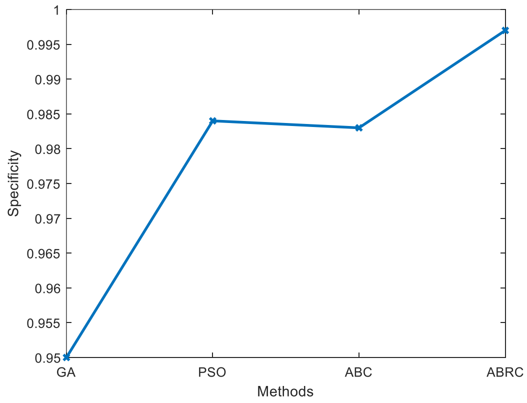

Figure 8: Comparison of Specificity

In figure 7 it is observed that the proposed ABRC provides a higher sensitivity rate rather than GA, PSO, and ABC. Also, in figure 8 specificity of the proposed ABRC is significantly higher than GA, PSO, and ABC. The sensitivity performance of the proposed ABRC is approximately 30% higher than existing and for specificity performance of ABRC is approximately 20% higher. In figure 9 comparative analysis of the proposed ABRC with the existing technique is presented. In figure 10 PDR analysis of the proposed ABRC with the existing technique is provided.

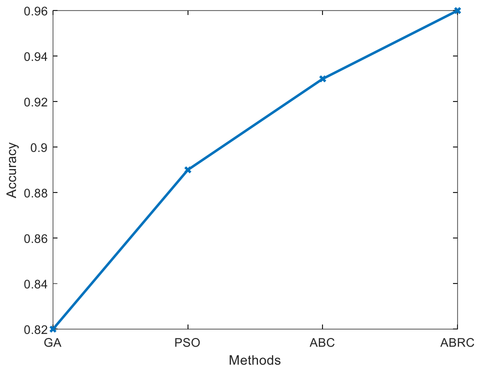

Figure 9: Comparison of Accuracy

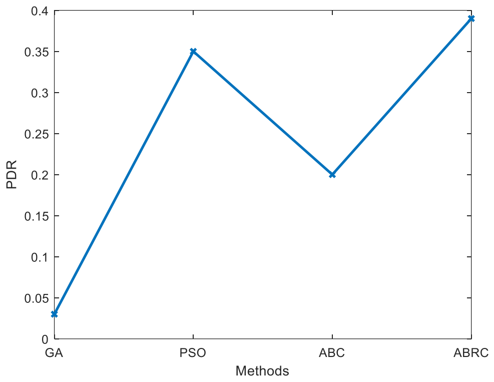

Figure 10: Comparison of PDR

In the analysis of figure 9, it is observed that the proposed ABRC provides a higher accuracy rate of 0.96 which is significantly higher than the existing technique. The accuracy is approximately 25% higher than existing techniques. From figure 10 it is observed that ABRC offers a higher PDR rate than the existing technique approximately around 20%. From the analysis, it is concluded that the proposed ABRC offers significant performance for varying attack scenarios. Further, the performance is effective for probe and unknown attack environments.

V. CONCLUSION

WSN clustering is performed with the aid of centered-order node clustering (CONC) concerning assigned weights of nodes. To elect effective CH from constructed cluster Improve Flower Pollination Algorithm (IFPA) is constructed which minimizes time consumption of node for data transfer. To improve WSN security CICIDS 2019 datasets are utilized for the classification of attacks in the network. Through CICIDS 2019 dataset supervised AdaBoost Regression Classifier (ABRC) is adopted.

The proposed classifier performance is applied in anomaly-based IDS which detects the presence of attack in the network. The performance of the proposed algorithm is comparatively examined with the existing technique stated that the proposed mechanism provides a higher PDR rate of 0.33 whereas the existing technique offers minimal PDR values. Furthermore, the performance of the proposed approach for accuracy, sensitivity, and specificity provides higher performance rather than existing techniques. The improved efficiency of is proposed approach is achieved due to the incorporation of ABRC which increases the prediction accuracy. On the whole, it is stated that the performance of the proposed algorithm is significantly higher than the existing techniques.

REFERENCES

1. Huang, S., & Lei, K. (2020). IGAN-IDS: An Imbalanced Generative Adversarial Network towards Intrusion Detection System in Ad-hoc Networks. *Ad Hoc Networks*, 102177.
2. Hosseini, S., & Zade, B. M. H. (2020). New hybrid method for attack detection using combination of evolutionary algorithms, SVM, and ANN. *Computer Networks*, 107168.
3. Mehmood, A., Lloret, J., & Sendra, S. (2016). A secure and low-energy zone-based wireless sensor networks routing protocol for pollution monitoring. *Wireless Communications and Mobile Computing*, *16*(17), 2869-2883.
4. Gao, Y., Ao, H., Feng, Z., Zhou, W., Hu, S., & Tang, W. (2018). Mobile network security and privacy in WSN. *Procedia Computer Science*, *129*, 324-330.
5. Bhatt, R., Maheshwary, P., Shukla, P., Shukla, P., Shrivastava, M., & Changlani, S. (2020). Implementation of Fruit Fly Optimization Algorithm (FFOA) to escalate the attacking efficiency of node capture attack in Wireless Sensor Networks (WSN). *Computer Communications*, *149*, 134-145.
6. Singh, D. P., Goudar, R. H., & Wazid, M. (2013). Hiding the sink location from the passive attack in WSN. *Procedia Engineering*, *64*, 16-25.
7. Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Computer Networks*, 107247.
8. Jeske, M., Rosset, V., & Nascimento, M. C. (2020). Determining the Trade-offs Between Data Delivery and Energy Consumption in Large-scale WSNs by Multi-Objective Evolutionary Optimization. *Computer Networks*, 107347.
9. Singh, S. (2019). A sustainable data gathering technique based on nature inspired optimization in WSNs. *Sustainable Computing: Informatics and Systems*, *24*, 100354.
10. Gill, K. S., Saxena, S., & Sharma, A. (2020). GTM-CSec: Game theoretic model for cloud security based on IDS and honeypot. *Computers & Security*, *92*, 101732.
11. Dua, M. (2020). Attribute Selection and Ensemble Classifier based Novel Approach to Intrusion Detection System. *Procedia Computer Science*, *167*, 2191-2199.
12. Kasongo, S. M., & Sun, Y. (2019). A Deep Long Short-Term Memory based classifier for Wireless Intrusion Detection System. *ICT Express*.
13. Aburomman, A. A., & Reaz, M. B. I. (2017). A novel weighted support vector machines multiclass classifier based on differential evolution for intrusion detection systems. *Information Sciences*, *414*, 225-246.
14. Aslahi-Shahri, B. M., Rahmani, R., Chizari, M., Maralani, A., Eslami, M., Golkar, M. J., & Ebrahimi, A. (2016). A hybrid method consisting of GA and SVM for intrusion detection system. *Neural computing and applications*, *27*(6), 1669-1676.
15. Gao, Y., Ao, H., Feng, Z., Zhou, W., Hu, S., & Tang, W. (2018). Mobile network security and privacy in WSN. *Procedia Computer Science*, *129*, 324-330.
16. Borkar, G. M., Patil, L. H., Dalgade, D., & Hutke, A. (2019). A novel clustering approach and adaptive SVM classifier for intrusion detection in WSN: A data mining concept. *Sustainable Computing: Informatics and Systems*, *23*, 120-135.
17. Donkal, G., & Verma, G. K. (2018). A multimodal fusion based framework to reinforce IDS for securing Big Data environment using Spark. *Journal of information security and applications*, *43*, 1-11.
18. Al-Sodairi, S., & Ouni, R. (2018). Reliable and energy-efficient multi-hop LEACH-based clustering protocol for wireless sensor networks. *Sustainable Computing: Informatics and Systems*, *20*, 1-13.
19. Sharma, S., & Kaul, A. (2018). Hybrid fuzzy multi-criteria decision making based multi cluster head dolphin swarm optimized IDS for VANET. *Vehicular Communications*, *12*, 23-38.
20. Aburomman, A. A., & Reaz, M. B. I. (2016). A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Applied Soft Computing*, *38*, 360-372.
21. Thaseen, I. S., & Kumar, C. A. (2017). Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *Journal of King Saud University-Computer and Information Sciences*, *29*(4), 462-472.
22. RM, S. P., Maddikunta, P. K. R., Parimala, M., Koppu, S., Reddy, T., Chowdhary, C. L., & Alazab, M. (2020). An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. *Computer Communications*.
23. Mehmood, A., Lloret, J., & Sendra, S. (2016). A secure and low-energy zone-based wireless sensor networks routing protocol for pollution monitoring. *Wireless Communications and Mobile Computing*, *16*(17), 2869-2883.
24. Kang, S. H., & Kim, K. J. (2016). A feature selection approach to find optimal feature subsets for the network intrusion detection system. *Cluster Computing*, *19*(1), 325-333.
25. Kuila, P., & Jana, P. K. (2014). Energy efficient clustering and routing algorithms for wireless sensor networks: Particle swarm optimization approach. *Engineering Applications of Artificial Intelligence*, *33*, 127-140.
26. Janakiraman, S. (2018). A hybrid ant colony and artificial bee colony optimization algorithm-based cluster head selection for iot. *Procedia computer science*, *143*, 360-366.
27. Ahmad, I., Hussain, M., Alghamdi, A., & Alelaiwi, A. (2014). Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components. *Neural computing and applications*, *24*(7-8), 1671-1682.
28. Ni, Q., Pan, Q., Du, H., Cao, C., & Zhai, Y. (2015). A novel cluster head selection algorithm based on fuzzy clustering and particle swarm optimization. *IEEE/ACM transactions on computational biology and bioinformatics*, *14*(1), 76-84.
29. Wang, G., Hao, J., Ma, J., & Huang, L. (2010). A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering. *Expert systems with applications*, *37*(9), 6225-6232.
30. Kalaiselvi, K., Suresh, G. R., & Ravi, V. (2019). Genetic algorithm based sensor node classifications in wireless body area networks (WBAN). *Cluster Computing*, *22*(5), 12849-12855.

AUTHORS PROFILE

Abidullha Adel, is assistant professor and Lecturer in Kunduz University. He has hold 2 years of university-level teaching experience in computer science and network technology and has a strong CV about research activities in computer science and information technology projects. He has had several research in the field of IoT and network technology. He has a patent and has received several medals and awards due to his innovative work and research activities. He has good skills in software engineering including experience with: .Net, SQL development, database management and designing many MIS for Hospital, Bank, School and library by using strong programming language (C#, Java, Python, MVC). His brilliant personal Strengths are in highly self-motivated team player who can work independently with minimum supervision, strong leadership skills, and outgoing personality with 2 years' experience in national and international organization in Afghanistan. He got his B.Sc. in information technology field from Badakhshan University in Afghanistan. Then he got his M. Sc. in Computer Science field from Nanjing University of Information and Technology (NUIST) Nanjing, China.

Md.Sohel Rana is a Senior Software Engineer . He has 5 years job Experience. He has completed a B.Sc degree in Computer science and engineering from Daffodil international university of Bangladesh in July 2016 and also he has got a M.S degree in Computer Science and Technology in July 2020 from Nanjing Univesity of Information & Science,Nanjing, China. He as received Chinese Government scholerhip and sevarel awards due to he's study and research .His research interests on those fields like Machine Learning, Artificial Intelligence, Data Mining, Object Detections and Digital Image Processing as well as Network Security.

Jayastree. J is a research associate with 5 year of experience. She completed her U.G in B.Tech - Electronics and Communication Engineering in B.S. Abdur Rahman Crescent Institute of Science and technology. Also, she pursued her P.G in M.Tech - Electronics with Communication Systems as major subjects in B.S. Abdur Rahman Crescent Institute of Science and technology. She completed a project on Monopole antenna design for multi-band applications with both software and hardware design. She scored 71% in GATE exam. At present, she registered in Post Graduate programme in Wireless Communication. Her major area of interest are antenna design, wireless communication, digital processing and wireless sensor network.