

Energy Efficient Cluster based Routing Protocol with Secure IDS for IoT Assisted Heterogeneous WSN

Sultan Alkhliwi

Assistant Professor at Faculty of Science
Northern Border University, Arar, Kingdom of Saudi Arabia

Abstract—Currently, wireless sensor networks (WSNs) and the Internet of Things (IoT) have become useful in a wide range of applications. The nodes in IoT assisted WSN commonly operate on restricted battery units, meaning energy efficiency is a major design issue. Clustering and route selection processes are commonly utilized energy-efficient techniques for WSN. Although several cluster-based routing approaches are available for homogeneous WSN, only a limited number of studies have focused on energy efficient heterogeneous WSN (HWSN). Moreover, security poses a major design issue in the HWSN. This paper introduces an energy efficient cluster-based routing protocol with a secure intrusion detection system in HWSN called EECRP-SID. The proposed EECRP-SID technique involves three main phases: cluster construction, optimal path selection, and intrusion detection. Initially, the type II fuzzy logic-based clustering (T2FC) technique with three input parameters are applied for cluster head (CH) selection. These parameters are residual energy level (REL), distance to the base station (DTBS), and node density (NDEN). In addition to CH selection, the salp swarm optimization (SSO) technique is utilized to select optimal paths for inter cluster data transmission, which results in energy efficient HWSN. Finally, to achieve security in cluster based WSN, an effective intrusion detection system (IDS) using long short-term memory (LSTM) is executed on the CHs to identify the presence of intruders in the network. The EECRP-SID method was implemented in MATLAB, and experimental outcomes indicate that it outperformed the compared methods in terms of distinct performance measures.

Keywords—Wireless Sensor Networks (WSN); Clustering; Routing; Type II fuzzy logic; salp swarm algorithm; long short-term memory (LSTM)

I. INTRODUCTION

Progressive development of the Internet of Things (IoT) and wireless sensor networks (WSNs) has proven beneficial for a broader set of real-time data gathering and tracking applications. Further, WSN has gained considerable attention from developers of various applications such as automated irrigation organization, target monitoring, landslide tracking, clinical observation, forest fire prediction, and disaster management. The major benefit of WSN is that it is embedded with a massive number of effective sensor nodes (SN), which help to monitor climatic disasters emerging in remote or harsh regions. Then, the SN also investigates atmospheric factors such as temperature, pressure, humidity, moisture content, and sound that represent the catastrophic symptoms. Once the sensing operation is completed by SN, the details are then collected and sent to the base station (BS). The sensor unit and data communication unit in SN consume maximum energy

which is an energy utilization module. Fig. 1 shows the structure of WSN. When all the energy is depleted, then it is considered expired or unable to process. A node is only referred to as dead when it is not suitable for replacement or it cannot be recharged by any other power source. Therefore, it is more essential to balance the power utilized by SN.

To resolve these issues, the clustering technique has been applied by many developers [1]. This is one of the best topologies for accomplishing an extended network duration and excellent efficiency. Moreover, it helps to preserve the power of SN under the development of several reliable clusters. The number of clusters developed might be either temporary or permanent, depending on the clustering mechanism used. In addition, clustering also segregates the jointly placed nodes into clusters. This distribution is determined according to the similarity metrics such as distance from the BS (DBS), transmission radius, and cluster density. Once the clusters are developed, a node inside a cluster is chosen as a head node termed the cluster head (CH). The CH has the responsibility of organizing the data collected from cluster members (CM) and forward them to the BS node. For instance, if a single cluster has a single node, then it is mandatory to communicate with the sink node rather than the SN communicating with the BS. Finally, it is applicable to reduce power utilization.

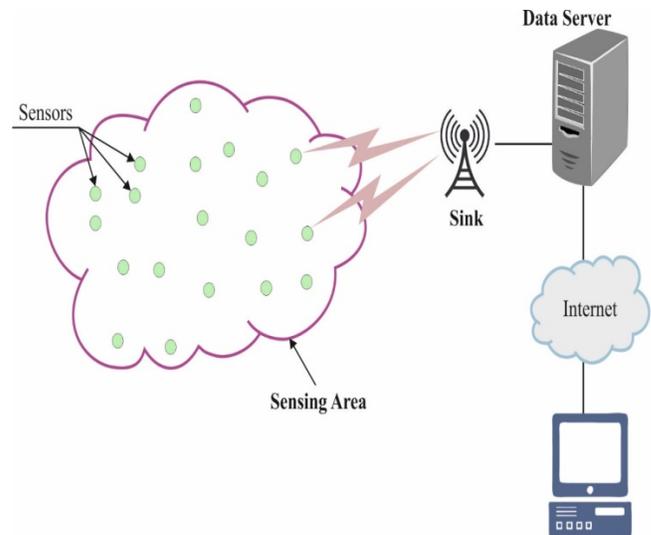


Fig. 1. Architecture of WSN.

Some of the classical (yet ineffective) routing protocols are direct transmission and minimum transmission energy types, which are suitable for preserving the energy of SN [2]. This problem is resolved with the help of clustering and cluster-related hierarchical routing protocol, such as low energy adaptive clustering hierarchy (LEACH) [2]. This is an effective model that ensures efficiency over previous protocols with the objective of energy preservation. Basically, it is computed for an identical system where the nodes have similar abilities, such as power, processing, and coverage sources. Consecutively, numerous protocols similar to LEACH have been presented to enhance the network lifetime significantly. Even though LEACH is a homogeneous protocol, the similarity of SN does not exist because of factors including manufacturing variations, diverse morphological aspects, and non-uniform physical terrain [3].

When the nodes have various configurations by means of classical factors, nodes are referred to as heterogeneous nodes and the network is framed as a heterogeneous network. Between the genres of heterogeneity, energy heterogeneity is highly applied because of its frequent dependence on non-replaceable batteries [4]. In the case of cluster-based routing, CH has been selected from all clusters by considering a few variables. When the CH collects from a cluster, then data aggregation is carried out to eliminate data redundancy. However, the CH experiences massive power exhaustion when compared to CM. Next, information gathered from CM is transmitted to BS using either single or multi-hop data transmission forwarded to users by internet sources [5, 6]. Developers have presented research principles on optimizing the CH election in homogeneous networks with the help of meta-heuristic approaches [7]. Thus, in heterogeneous WSN (HWSN), CH election is accomplished by using tactical amendments in the threshold-relied function used for the nodes.

The aim of the optimization module for CH election is to reduce power utilization. Further, optimized CH election for energy effective routing is assumed to be a non-deterministic polynomial-time hard (NP-hard) issue. Regardless of this, CH election is optimized using meta-heuristics that are embedded with major fitness functions [8]. At this point, the meta-heuristic model has been employed because of these features to convert the best solution. Thus, multitudinous optimization routing principles have been deployed to optimize CH selection and to ensure the scalability of a system that transmits data [9]. Security is another concurrent problem in HWSN, which can be tackled by using an intrusion detection system (IDS) [10]. This is mainly employed for tracking abnormal events that take place in the network. To attain effective performance in the IDS, machine learning (ML) and deep learning (DL) approaches have been found useful. DL is a modified form of ML, which is used to extract the features proficiently. Accordingly, the classical ANN is employed for handling nonlinear cases and DL methodologies are utilized for the extraction of features and to make decisions in a similar way to the human brain.

The contribution of this paper is defined as follows. An energy efficient cluster-based routing technique (EECRP-SID) with a secure intrusion detection system is introduced in

HWSN. Initially, the type-II fuzzy logic clustering (T2FC) technique with three input parameters is applied for CH selection. These parameters are residual energy level (REL), distance to base station (DTBS), and node density (NDEN). In addition to CH selection, the salp swarm optimization (SSO) technique is employed to select optimal paths for inter cluster data transmission, which results in energy efficient HWSN. Finally, in cluster based WSN, a new IDS using long short-term memory (LSTM) is executed on the CHs to classify the existence of intruders in the network. The EECRP-SID method is implemented in MATLAB. The experimental outcomes demonstrated that the EECRP-SID model outperformed the compared methods in terms of distinct performance measures.

II. RELATED WORKS

Recently developed work is relevant to the newly deployed approach, as defined in the following. In [11,21], a hybridized grey wolf and crow search method depending on optimal CH selection (HGWCSSOA-OCHS) was developed and used for enhancing network lifespan. This was achieved by reducing delays, distance between nodes, and energy utilization. The hybrid GWO and CSO approach in CH election maintains the trade-off between exploitation and exploration in a searching area. Moreover, a novel technique called PSO-based selection (PSOBS) for selecting optimal rendezvous points has been proposed [12]. Using PSO, the developed approach was used to find the best rendezvous points to accomplish remarkable network resources. Additionally, a weighted score was estimated for SN according to the data packets obtained from various sensors.

In Edla et al. [13], a clustering method was presented using shuffled complex evolution of PSO (SCE-PSO), which is an effectual FF using average cluster distance, CHs load, and number of loaded CHs in a system. The approach in [14] focused on CH selection using GA and KH methods for WSN. The key objective of this model was to enhance the lifespan of WSN routing with the BS node by designing an effectual routing approach on the basis of hybridized metaheuristic optimization models [15]. This originated as an initial framework with innovative and global searching capabilities of particle swarm optimization (PSO), diverse operator of differential approach, and pheromones of ant-colony optimization (ACO) technology to limit local searches and retain population diversity.

In Dattatraya and Rao [16], a new CH election process was developed to extend network duration and energy efficiency using fitness based glowworm swarm and the fruitfly algorithm (FGF) for WSN. In [17], an energy effective CH was selected based on the whale optimization algorithm (WOA), termed WOA-Clustering (WOA-C) was projected. Then, newly presented technology was used in the election of energy-aware CHs according to the FF, which applied the RE of a node and summation of the power of adjacent nodes.

III. PROPOSED MODEL

The proposed EECRP-SID technique involves three main phases: cluster construction, optimal path selection, and intrusion detection.

A. Clustering Methodology

The T2FC technique incorporates two stages: the CH selection and cluster formation phases [18, 22]. First, the BS node used the T2FC model to select appropriate CHs to accomplish even load distribution. Hence, the selected CHs are readily available for developing clusters and adjacent nodes. To select CH and cluster size, fuzzy logic and three input features (REL, DTBS, and NDEN) were employed. When compared with all other sources, energy is one of the most essential resources required by WSN. The process involved in T2FL is depicted in Fig. 2. Maximum power is consumed by CH (compared to CM) when computing certain operations such as aggregation, processing, and routing data. Here, REL is measured with the following function:

$$E_r = E_0 - E_c, \quad (1)$$

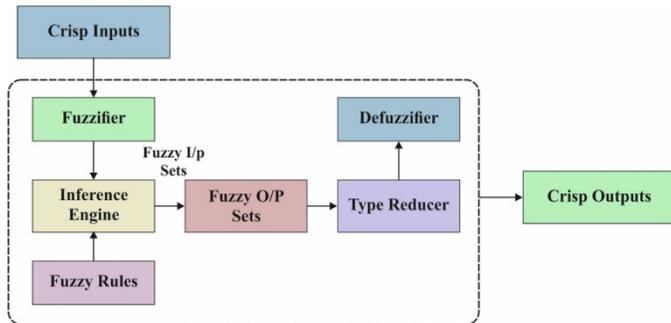


Fig. 2. Block Diagram of T2FL.

where E_0 and E_c are the energy and power consumed by a node, respectively, and E_r denotes the REL of the regular node.

The transmitted message utilizes power, which is directly proportional to the square of the distance over the candidate as well as the source node. Then, DTBS is evaluated as follows:

$$C = \frac{d_{avg}^2}{d_0^2} \quad (2)$$

Here, d_{avg} represents the higher distance between node and neighbours and d_0 implies the broadcasting radius of a node.

The NDEN implies that closer neighbours develop a better score and a node with higher probability is considered as CH. The NDEN is measured using the following expression:

$$NDEN = \frac{|D_i - D_0|}{D_0} \quad (3)$$

Here, D_i is the count of neighbours and D_0 is the optimal number of neighbours. In this approach, T2FL is composed of the following four principles:

- Fuzzifier

The main aim of the fuzzifier is to convert the exact input as fuzzified measures. The input parameters with linguistic variables for CH selection is then transformed.

- Fuzzy Inference System

The performance of T1FL and T2FL is homogeneous. In this module, a set of 27 rules are used. A sample rule for CH election is depicted in Eq. (4).

$$\begin{aligned} &Rule(i) \text{ IF } x_1 \text{ is } A_1(i) \text{ AND } x_2 \text{ is } A_2(i) \text{ AND } x_3 \text{ is } A_3(i) \\ &THEN y_1 \text{ is } B_1(i) \text{ AND } y_2 \text{ is } B_2(i) \end{aligned} \quad (4)$$

Here, i refers to the i^{th} rule in fuzzy rule, while A_1 , A_2 , and A_3 are the equivalent fuzzy set of x_1 , x_2 , and x_3 .

- Membership functions

Here, T2FL is defined by effective and poor membership functions (MFs). These functions are implied as T1FL MF. The interval between the two functions refers to the footprint of uncertainty (FOU) that describes a T2FL set. Consider the FOU is represented as f . If $f \in [0, 1]$, and $f \rightarrow 0$, then MF is implied as T1FL. If $f \rightarrow 0$ to 1, then T2FL comprises a wider range of FOU within $[0, 1]$. Therefore, the rule development in T2FL logic is similar to T1FL, as given in the following:

$$Type\ 2\ FL = Principal\ MF\ (Type\ 1\ FL) + FOU \quad (5)$$

- Type reducer/ Defuzzifier

The type-reducer produces a T1FL outcome, which is transformed as a mathematical result after the execution of defuzzifier. The node gathers PCH and cluster size, where it broadcasts a message to closer nodes, where the message is composed of Node-ID and the value of PCH. A node with a high possibility is treated as CH and sends CH_WON to adjacent nodes. A node accomplishes numerous CH_WONs from neighbouring nodes, then sends the CH_JOIN message and unifies closer CHs. Subsequently, to eliminate the premature death of CH, CH rotation is carried out. Further, when the RE of CH goes beyond the threshold measure, then CH rotation is performed. This mechanism is useful for eliminating premature death of CH and tends to enhance network lifetime.

B. Route Selection Algorithm

Developing a numerical approach that reflects the smart behaviour of a swarm is one of the fundamental steps in SI-reliant models for resolving the optimization issues. Numerical methods in addition to fish swarm, birds swarm, and ant swarm were employed in optimization issues. To model the salp chain, the individuals present in salps swarms were classified into leader and followers [19,23]. Initially, the leader is the important individual among the salps chain for computing the direction and foraging the route of a swarm and assisting the salp chain to identify food. The residual salps are considered as followers, who obey the leader to develop a chain-like structure. Therefore, while a numerical approach accelerates salps chain production, it is not suitable for resolving the optimization issues. Computing the global optimal value is one of the main aims in the optimization issues, where the global optimal score has been employed as food required by salps chain. Based on leader position, the salps chain is placed near to the food, which is expressed as follows:

$$X_j^1 = \begin{cases} F_j - c_1[(ub_j - lb_j)c_2 + lb_j] & \text{if } c_3 < 0.5 \\ F_j + c_1[(ub_j - lb_j)c_2 + lb_j] & \text{if } c_3 \geq 0.5 \end{cases} \quad (6)$$

where X_j^1 denotes the location of a leader in the j -th dimension; F_j mimics the place of food in the j -th dimension; ub_j defines the lower bound of the j -th dimension; and lb_j

refers the upper bound of the j -th dimension. These upper and lower bounds were employed to minimise excess searching space. Parameter c_2 denotes a random value from [0.1] that has been applied to control leader movement. Parameter c_3 means a random value from [0.1] that has been employed uniformly, regardless of whether the leader is moving towards or away from the food location. Parameter c_1 is depicted in Eq. (7), which is an extension factor applied for managing the global exploration and local exploitation.

$$C_1 = 2e^{-(4t/T)^2} \quad (7)$$

Here, t shows the recent count of iterations and T denotes the overall count of iterations. From Eq. (2), the adjustment factor c_1 would reduce gradually for every iteration. To make the followers obey the instructions of a leader to make a chain, Newton's law of motion was applied to upgrade the place of followers, defined as follows:

$$X_j^i = \frac{1}{2}a \cdot t^2 + v_0 \cdot t \quad (8)$$

where X_j^i is the place of the i -th follower in the j -th dimension if $i \geq 2$; t implies time; v_0 mimics the first speed, and simulation of the follower's direction $a = v_{final}/v_0$. The speed of a follower is defined by $v = (x - x_0)/t$. Subsequently, the time variable of the optimization issues is depicted as an iteration count; thus, the iteration interval is from $t = 1$. The follower's basic speed is $v_0 = 0$. Eq (8) is improvised as follows:

$$X_j^i = \frac{1}{2}(X_j^i + X_j^{i-1}) \quad (9)$$

In the case of followers, a better position might be reached instead of using a recent solution (food). Then, the food is interchanged by an optimal location and upgraded by the leader to move in the direction of the food. The benefits of SSA are given in the following:

- It can be acquired from the numerical approach of SSA with simple structure and can be trained using parameters and operators, which makes the implementation easier.
- In optimization, SSA applies the best solution in the present iteration as food. Even the fitness values are applicable in a complete population that affects the quality of the food.
- Leaders are capable of exploring and moving closer, according to the position of the food. The followers have to move in a chain format, which demonstrates the model's simplicity. Under the consideration of benefits, the demerits are also defined in SSA.

The principle behind the best SSA is defined in the following:

Step 1. Allocate the algorithm variables: No. of iterations T , No. of ascidian populations N , test function dimension D .

Step 2. Allocate the salp population based on the maximum and minimum bounds, $t = 1$.

Step 3. Estimate the fitness value of a searching individual, and consider them with optimal fitness value in recent population as food F_j .

Step 4. Upgrade c_1 on the basis of Eq. (2) and produce random values c_2 and c_3 .

Step 5. When $i = 1$, upgrade the leader's place on the basis of Eq.(1). When $i \geq 1$, upgrade the follower's place based on the Eq. (4). $t = t + 1$.

Step 6. Evaluate whether a method has attained a higher count of iterations. When the procedure is terminated, then the optimal value is provided back; else, repeat Step 3.

A flowchart of SSA is illustrated in Fig. 3. The task of finding an effectual path using SSA is explained in the following:

- Every CH sends the route information (such as node-ID, RE, and distance) from closer CHs. Thus, CH saves the details about a routing table.
- A route from CHs to BS is computed using a salp suited in the CHs occasionally. The fitness function is derived as

$$FF_{ij} = \frac{1}{d_{ij}} \quad (10)$$

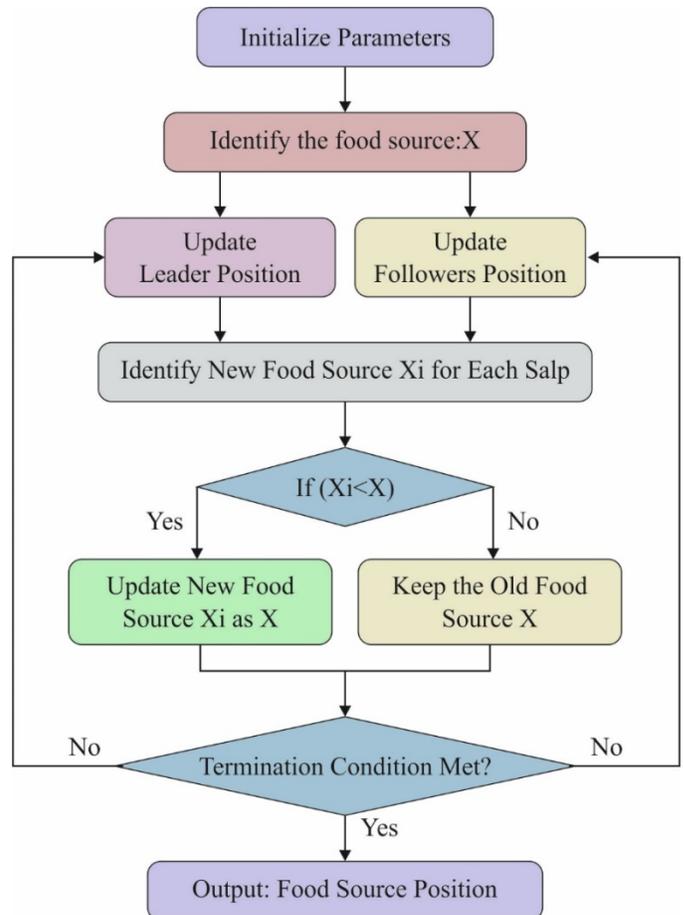


Fig. 3. Flowchart of SSA.

where d_{ij} denotes a distance between source node i to equivalent CH. The distance d_{ij} can be evaluated as a Euclidian distance, as in Eq. (11).

$$d_{ij} = \sqrt{(S(i).xd - s(j)xd)^2 - (S(i).yd - s(j)yd)^2} \quad (11)$$

The energy level of nodes is given by

$$E = \frac{E_0 - E_{residual}}{\sum_{k \in N} E_k} \quad (12)$$

where E_0 signifies an initial power and $E_{residual}$ refers to the remaining energy of a node. Terms α and β are parameters employed for regulation.

- During data transmission, a node with a higher probability is selected as a relay node from the CHs to BS.

C. LSTM based IDS

To determine the existence of intrusions in the network, the LSTM model is employed [20,24]. LSTM is a special extension of the recurrent neural network (RNN) structure, which is a DL model. It is composed of feedback connections used for eliminating long term dependencies, and is capable of computing the data points and a complete series of knowledge. For example, LSTM can be operated on un-segmented data, correlated recognition patterns, audio recognition, and abnormal prediction in network traffic or IDSs (intrusion detection systems). A typical LSTM unit is composed of four major sections: cell, input gate, output gate, and forget gate. Initially, cell memories over rare time intervals, and hence, 3 gates are standardized the data flow into and out of a cell.

LSTM systems are particularly suitable for classifying, computing, and developing predictions according to time-sequential data, because they suffer from scarce duration among significant events in a time series. Initially, the LSTM is deployed for affecting the diminishing gradient issues that identify classical RNN. The relative insensitivity to gap length is a combination of LSTM with RNNs, hidden Markov models (HMMs), and alternate series learning models in massive domains. An LSTM with four stages in a cell is illustrated in Fig. 4. In a cell state, the horizontal topmost line is implemented in the cell, which refers to the cell state. Then, the LSTM is not capable of eliminating the data from a cell state, which is carefully regularized by gates. A cell is composed of massive structures.

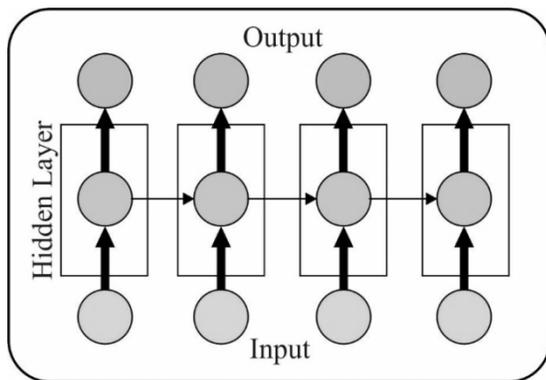


Fig. 4. Structure of LSTM Cell.

Step 1: The Initial step in LSTM is to find the unwanted decision made by a distant cell state. This is developed by a sigmoid layer (forget gate layer) as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (13)$$

where h_{t-1} implies the result from the existing time stamp, x_t = new input, and b_f = bias.

Step 2: Next, a decision is saved in a cell state. This is composed of two parts. Initially, a sigmoid layer (input gate layer) that selects upgrading values and a tanh layer is applied to make a vector of novel contender measures, which is included in the state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (14)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c) \quad (15)$$

Step 3: The previous cell state is updated, $C(t-1)$, as a novel cell state. Then, steps are selected for performing some events. Initially, the previous state is increased by the removal of images, then it is included in the cell state. This is referred to as novel candidate measures, which is scaled by a proportion for deciding a state update in all measures.

$$C_t = f_t * C_t + i_t * \tilde{C}_t \quad (16)$$

Step 4: This is the final stage, and the output is supported by a cell state that is an extended version. First, a sigmoid layer is implemented for selecting the portions of a cell state as a result. Then, the cell state is fixed by tanh and increases the final result of a sigmoid gate; thus, the consequent portions.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (17)$$

$$h_t = o_t * \tan h(C_t) \quad (18)$$

IV. PERFORMANCE VALIDATION

An analysis of the results of the EECRP-SID model was conducted in terms of energy efficiency with varying rounds, as shown in Fig. 5. The energy should deplete at a slower rate with an increased number of rounds. The figure shows that the MS-GAOC model exhibited worse outcomes and depleted all its energy at a faster rate compared to existing techniques. Next to that, the IABC model tried to provide better results over the MS-GAOC model by exhausting its energy at a slightly reduced rate. Moreover, the KHA algorithm accomplished somewhat satisfactory performance by attaining energy exhaustion at a moderate rate. Further, the MO-PSO algorithm reached near optimal results by attaining energy depletion at a minimum rate. However, the EECRP-SID model demonstrated superior energy efficiency by depleting the energy at a slower rate.

An average delay analysis of the EECRP-SID model was conducted and compared with the other methods under a varying node count, as illustrated in Fig. 6. As shown, the MO-PSO algorithm was found to have poor performance and resulted in a higher average delay, tending to increase significantly with an increase in the number of nodes. Next, the KHA algorithm required a somewhat less average delay compared to the MO-PSO model, but not compared to the other methods. The IABC model achieved reasonable results

with a moderate average delay, while the MS-GAOC model demonstrated competitive results by requiring a low average delay. However, the proposed EECRP-SID model only required a minimum amount of average delay.

Fig. 7 shows the PDR results analysis of the EECRP-SID model and existing techniques under a distinct node count. The figure demonstrates that the MO-PSO algorithm achieved an inferior outcome with the lowest amount of PDR. Subsequently, the MS-GAOC model surpassed the MO-PSO model by attaining slightly better PDR (but not compared to the other methods). The IABC model achieved an even higher PDR, whereas a near-optimal PDR value was achieved by the MS-GAOC model. However, the proposed EECRP-SID model proved to be superior in the clustering and routing processes by obtaining a maximum PDR.

In Fig. 8, the network lifetime analysis of the EECRP-SID model is presented and compared with existing techniques in terms of alive node count. From the figure, it can be noted that the MS-GAOC model attained a maximum number of dead nodes with an increasing number of nodes. This implies that it holds a lower number of alive nodes. Further, the IABC model achieved a somewhat better lifetime by offering a slightly higher number of alive nodes. The KHA model achieved a somewhat acceptable network lifetime with moderate alive node count. Although the MO-PSO algorithm exhibited near-optimal alive node count, superior network lifetime was demonstrated by the EECRP-SID model.

Next, the intrusion detection performance of the EECRP-SID model was evaluated using the KDD Cup 1999 dataset, in terms of different evaluation parameters (as shown in Fig. 9 and 10). The results indicate that the RF model resulted in a minimum sensitivity of 92.39%, specificity of 93.83%, F-score of 93.58%, and kappa of 85.99%. This was followed by the RBF Network, which achieved a somewhat higher sensitivity of 93.4%, specificity of 92.38%, F-score of 93.38%, and kappa of 85.79%. The RT model displayed a moderate outcome with a sensitivity of 95.68%, specificity of 95.39%, F-score of 95.84%, and kappa of 91.06%. The DT model exhibited an even better classifier outcome, with a sensitivity of 95.68%, specificity of 95.37%, F-score of 95.83%, and kappa of 91.03%. In addition, the LR achieved near optimal intrusion detection, with the highest sensitivity of 97.26%, specificity of 96.92%, F-score of 97.29%, and kappa of 94.19%. However, the proposed LSTM-IDS model achieved a higher sensitivity of 97.68%, specificity of 98.54%, F-score of 98.13%, and kappa of 96.21%.

Fig. 10 illustrates an accuracy analysis of the different IDSs on the applied test dataset. The resultant values meant that the CSA-PSO model demonstrated ineffective detection results, with the lowest accuracy of 75.51%. The GBT-IDS model achieved an accuracy of 84.25%, while the RBFNetwork and RF models achieved near identical accuracies of 92.93% and 93.04%, respectively. The FCM-IDS, DT, and RT models exhibited closer accuracy of 95.3%, 95.53%, and 95.55%, respectively. Although the existing CSA-IDS and LR models demonstrated competitive intrusion detection with accuracies of 96.88% and 97%, respectively, the presented LSTM-IDS model achieved the highest accuracy of 98.43%. These results

verify the effectiveness of the proposed methods with respect to energy efficiency, lifetime, and intrusion detection.

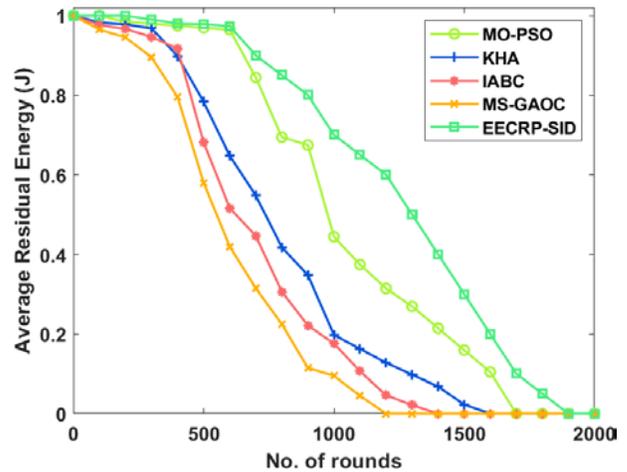


Fig. 5. Average Residual Energy Analysis of the EECRP-SID Model Compared with Existing Techniques.

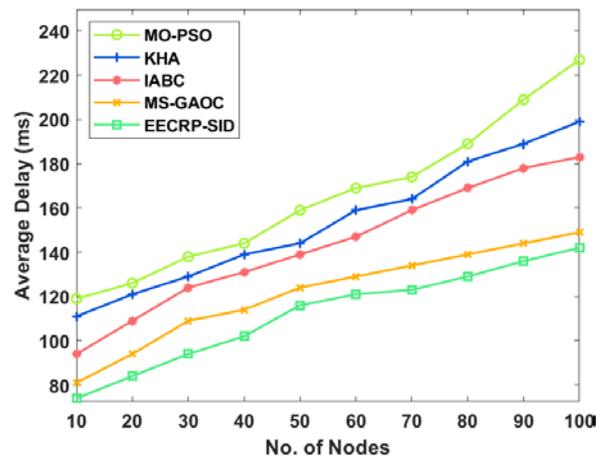


Fig. 6. Average Delay Analysis of the EECRP-SID Model Compared with Existing Techniques.

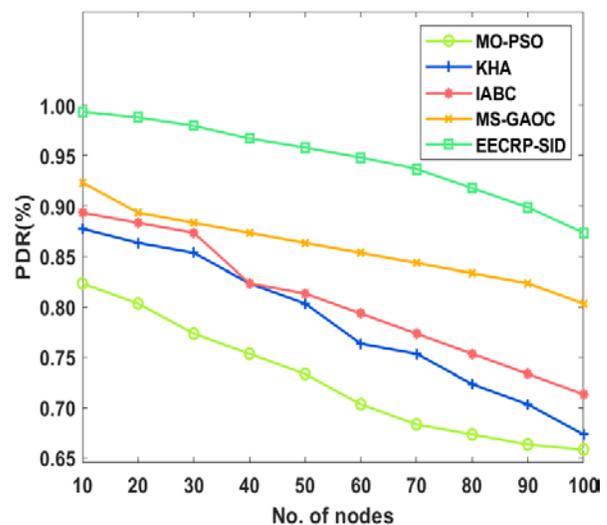


Fig. 7. PDR Analysis of the EECRP-SID Model Compared with Existing Techniques.

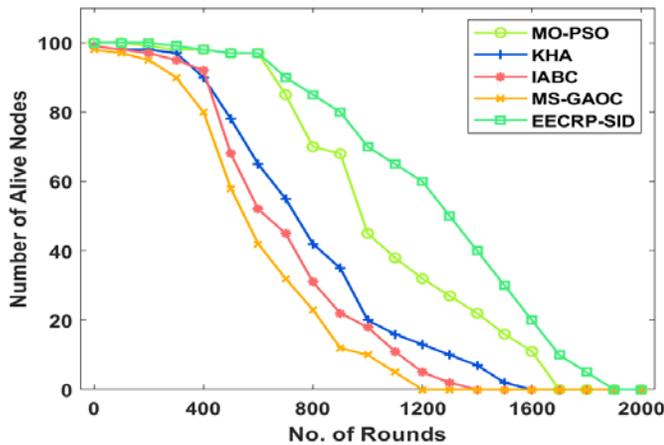


Fig. 8. Lifetime Analysis of the EECRP-SID Model Compared with Existing Techniques.

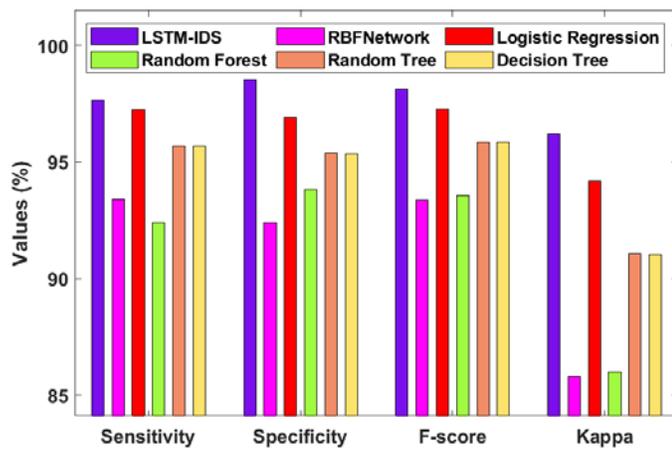


Fig. 9. Intrusion Detection Analysis of the LSTM-IDS Compared with Existing Techniques.

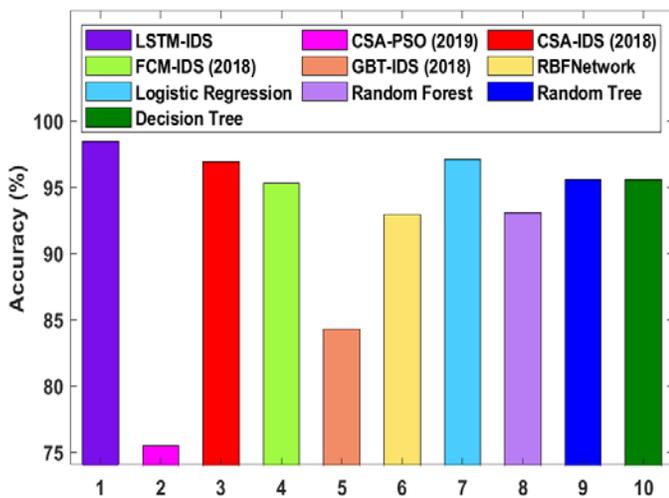


Fig. 10. Accuracy Analysis of the LSTM-IDS Compared with Existing Techniques.

V. CONCLUSION

This paper presents the development of a novel EECRP-SI model (EECRP-SID) for energy-efficient cluster-based routing protocol with secure IDS in HWSN. The proposed EECRP-SID technique involves three main phases: cluster construction, optimal path selection, and intrusion detection. First, the T2FC technique is applied to form clusters and elect proficient CHs. Next, an SSA based inter clustering routing task is employed. Finally, to determine the existence of intrusions in the network, the LSTM model is used. During intrusion detection in the HWSN process, the presented LSTM-IDS model achieved the highest accuracy of 98.43%. The experimental results analysis of the EECRP-SID model showcased superior results in terms of energy efficiency, lifetime, and intrusion detection. In the future, the performance of intrusion detection can be improved by the hybridization of metaheuristic algorithms for tuning the hyperparameters.

REFERENCES

- Sharma, R., Vashisht, V., Singh, U., "EEFCM-DE: energy efficient clustering based on fuzzy C means and differential evolution algorithm in wireless sensor networks," IET Commun. 13 (8), 996–1007, 2019b
- W.R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in: Proc. of 33rd Annu. Int. Conf. on Syst. Sci. Hawaii, IEEE, 2000.
- V. Mhatre, C. Rosenberg, Homogeneous vs heterogeneous clustered sensor networks: a comparative study, in: Int. Conf. on Commun., IEEE, pp. 3646–3651, 2004.
- M. Yarvis, N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu, S. Singh, Exploiting heterogeneity in sensor networks, in: Proc. of 24th Annu. Jt. Conf. on Comput. Commun. Soc., INFOCOM 2005, IEEE, pp. 878–890, 2005
- J. Huang, D. Ruan, Y. Hong, Z. Zhao, H. Zheng, IMHRP: Improved multi-hop routing protocol for wireless sensor networks, in: J. Phys. Conf. Ser., IOP Publishing, p. 012054, 2017
- S. Kumar, P. Ranjan, R. Ramaswami, M.R. Tripathy, Resource efficient clustering and next hop knowledge based routing in multiple heterogeneous wireless sensor networks, Int. J. Grid High Perform. Comput. 9, 1–20, 2017
- A.A. Bara'a, E.A. Khalil, A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks, Appl. Soft Comput. 12, 1950–1957, 2012
- C.-W. Tsai, T.-P. Hong, G.-N. Shiu, Metaheuristics for the lifetime of WSN: A review, IEEE Sens. J. 16, 2812–2831, 2016
- B.P. Deosarkar, N.S. Yadav, R.P. Yadav, Clusterhead selection in clustering algorithms for wireless sensor networks: A survey, in: Int. Conf. on Comput. Commun. Netw., ICCN, IEEE, pp. 1–8, 2008.
- Butun, I., Morgera, S.D. and Sankar, R., A survey of intrusion detection systems in wireless sensor networks. IEEE communications surveys & tutorials, 16(1), pp.266-282, 2013
- Subramanian, P., Sahayaraj, J.M., Senthilkumar, S. and Alex, D.S., A Hybrid Grey Wolf and Crow Search Optimization Algorithm-Based Optimal Cluster Head Selection Scheme for Wireless Sensor Networks. Wireless Personal Communications, pp.1-21, 2020
- Tabibi, S. and Ghaffari, A., Energy-efficient routing mechanism for mobile sink in wireless sensor networks using particle swarm optimization algorithm. Wireless Personal Communications, 104(1), pp.199-216, 2019
- Edla, D.R., Kongara, M.C. and Cheruku, R., SCE-PSO based clustering approach for load balancing of gateways in wireless sensor networks. Wireless Networks, 25(3), pp.1067-1081, 2019.

- [14] Karthick, P.T. and Palanisamy, C., Optimized cluster head selection using krill herd algorithm for wireless sensor network. *Automatika*, 60(3), pp.340-348, 2019
- [15] Wang, H., Li, K. and Pedrycz, W., An Elite Hybrid Metaheuristic Optimization Algorithm for Maximizing Wireless Sensor Networks Lifetime With a Sink Node. *IEEE Sensors Journal*, 20(10), pp.5634-5649, 2020
- [16] Dattatraya, K.N. and Rao, K.R., Hybrid based cluster head selection for maximizing network lifetime and energy efficiency in WSN. *Journal of King Saud University-Computer and Information Sciences*, 2019
- [17] Jadhav, A.R. and Shankar, T., Whale optimization based energy-efficient cluster head selection algorithm for wireless sensor networks. *arXiv preprint arXiv:1711.09389*, 2017
- [18] Nayak, P. and Vathasavai, B., Energy efficient clustering algorithm for multi-hop wireless sensor network using type-2 fuzzy logic. *IEEE Sensors Journal*, 17(14), pp.4492-4499, 2017
- [19] Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H. and Mirjalili, S.M., 2017. Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software*, 114, pp.163-191, 2017
- [20] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H. and Xu, B., August. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 207-212, 2016
- [21] Alghamdi, T.A. Energy efficient protocol in wireless sensor network: optimized cluster head selection model. *Telecommun Syst* **74**, 331–345 (2020). <https://doi.org/10.1007/s11235-020-00659-9>
- [22] Devaraj, A.F.S. Energy aware reliable route selection scheme with clustered RP model for wireless sensor networks to promote interaction between human and sensors. *J Ambient Intell Human Comput* (2020). <https://doi.org/10.1007/s12652-020-02147-z>
- [23] B.K. Patle, Ganesh Babu L, Anish Pandey, D.R.K. Parhi, A. Jagadeesh, A review: On path planning strategies for navigation of mobile robot, *Defence Technology*, Volume 15, Issue 4, 2019, Pages 582-606, ISSN 2214-9147
- [24] Jesús Gonzalez, Wen Yu, Non-linear system modeling using LSTM neural networks, *IFAC-PapersOnLine*, Volume 51, Issue 13, 2018, Pages 485-489, ISSN 2405-8963,