Alexandria Engineering Journal 82 (2023) 82–100

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Original Article

Secure localization techniques in wireless sensor networks against routing attacks based on hybrid machine learning models

Gebrekiros Gebreyesus Gebremariam [a,b,*], J. Panda [b], S. Indu [b]

[a] Department of Electronics and Communication Engineering Raya University, Maychew 7020, Tigray, Ethiopia
[b] Department of Electronics and Communication Engineering Delhi Technological University, Shahbad Daulatpur, Main Bwana Road, Delhi 110042, India

ARTICLE INFO

Keywords:
Secure localization
Network model
Average detection accuracy
Routing attacks
Hybrid machine learning models

ABSTRACT

The identification and localization of malicious nodes in wireless sensor networks (WSNs) is a hot area of research that can considerably extend the network's lifetime and make it more valuable. We use sensors whose positions are known, or anchor nodes, to make educated guesses about the positions of the unknown nodes. Several localization methods have been developed for precise estimation of the unknowable nodes. So, during the network setup process, finding suitable network parameters for node localization with the requisite accuracy in a short amount of time remains a tough task. Due to the fact that they manipulate network resources and routing protocols, routing assaults like wormhole attacks, Sybil attacks, blackhole attacks, and replay attacks are just a few examples of the types of attacks that have the potential to hinder the accuracy of localization and the quality of service provided by WSNs. This work proposes safe localization and detection of routing threats in wireless sensor networks by utilizing hybrid optimized machine learning approaches for optimal distance, position, and data communication. These approaches aim to find the optimal distance between sensors and the optimal position of sensors. Calculating the average localization accuracy and finding malicious nodes both need the use of the benchmark datasets CICIDS2017 and UNSW NB15. The machine learning algorithms that have been provided can be utilized with these datasets. The cluster labelling K-means clustering technique is applied to binary classification in the system that has been proposed. As a consequence, the system achieves an average detection accuracy of 100%. The findings of the simulation indicate that the proposed hybrid strategy is capable of achieving a higher level of localization accuracy of the unknown nodes, with an average localization error of 0.191.

1. Introduction

Wireless sensor networks (WSNs) consist of many sensor nodes providing a self-organizing network having multi-hop transmission and standard functioning for discovering the location of the data [1]. The network consists of randomly distributed and positioned sensor nodes for collecting data from the environment with limited battery constrained and computational data processing for scientific and military applications. WSNs are essential for scientific research and industrial areas that have recently attracted considerable attention [2]. The concept of localization in wireless sensor networks is necessary for computing and identifying malicious attacks that degrade the network's performance. The characteristics of random distribution and limited resource constraints of computational data processing and power consumption make the localization techniques in WSNs essential for detecting and classifying malicious nodes. The localization techniques are critical for computing the exact position of the nodes upon the transmission of the radio signals that are vulnerable to various security attacks and environmental losses in WSNs. WSNs that rely heavily on the location of sensor nodes include security applications that protect sensitive data from unauthorized users and various assaults in vehicle Adhoc networks (VANETs) and mobile Adhoc networks (MANETs). Localization, node deployment, routing, energy consumption by sensor nodes, and limited lifetime of sensor nodes are just few of the issues that WSNs confront. Energy consumption by sensor nodes is another challenge in WSNs, and adopting clustering-based routing protocols is one way to extend the network's lifetime [3].

The randomly deployed sensor nodes cannot find the exact position within the target area and transmission radius. The localization schemes used to locate these sensor nodes are categorized as range-based and

* Corresponding author.
E-mail addresses: kiros2004comp@gmail.com, kiros@rayu.edu.et (G.G. Gebremariam), jpanda@dce.ac.in (J. Panda), s.indu@dtu.ac.in (S. Indu).

https://doi.org/10.1016/j.aej.2023.09.064
Received 18 October 2022; Received in revised form 3 May 2023; Accepted 24 September 2023
Available online 5 October 2023

range-free localization approaches. The Metrics like receiver signal strength indicator (RSSI) and real-time data like the angle of arrival and time of arrival are commonly used in range-based localization methods. Regarding pinpointing the location of malicious nodes, range-based localization methods perform admirably for detecting and localizing malicious nodes. The localization technique is utilized to monitor sensor nodes' movement with minimum energy, distance, cost, and distance through connectivity. The anchor node is location-aware and helps estimate the localization and position of randomly deployed unknown nodes. The primary goal of the proposed technique is to identify and pinpoint the origin of routing attacks in WSNs using hybrid approach based on distance vector routing since the current localization techniques are vulnerable to various routing threats and attacks, including wormhole attacks, blackhole attacks, and Sybil attacks.

2. Routing attacks

Different types of routing attacks, such as wormhole, Sybil, black hole, and sinkhole, are discussed in the context of wireless sensor networks. By inserting rogue nodes that generate incorrect routing information and data transfer between sensor nodes and the base station, routing attacks can significantly affect the performance of wireless sensor networks.

2.1. Sybil attack

Sybil attack is the most harmful routing attack in wireless sensor networks. Sybil attacks can be easily deployed in the network affecting the localization accuracy by compromising the anchor nodes [3]. The Sybil attack generates multiple identities from a single geographical position by compromising the beacon node having the same distance to the unknown node. The compromised Sybil attacks have the same received signals strength indicator (RSSI) measured concerning the unknown node. Sybil attacks can also remain concealed through the network. The beacon node, which acts as the cluster head and has more processing power than the sensor nodes, is the target of the Sybil attack, which generates many bogus routing paths and identities, as depicted in Fig. 1. Sybil nodes (S) serve as beacon nodes, injecting a malicious node (M), and allowing the normal nodes (N) to pick up the malicious node's broadcast message. This makes the busy the normal operating of the network and denies the legitimate nodes from accessing the network.

2.2. Wormhole attack

A wormhole attack is a harmful routing attack that disrupts the normal function of the network by creating a wormhole tunnel between two malicious nodes and modifying the packets [4]. The wormhole tunnel is located between two nodes with a significant distance to capture and replay the packets of each other along the tunnel. Fig. 2 shows the two ends of malicious nodes A and B. Wormhole attack facilitates a high-speed, long link for transmitting packets between the nodes making a good network service for the attacker's advantage. The attacker attracts the network traffic and drops data packets disrupting the flow of packets breaking and analyzing the information in the network. The wormhole attack also turns off the wormhole tunnel to create false routing events and a fake list of neighbors.

2.3. Sinkhole attacks

The sinkhole node pretends to be an ordinary node and sends out routing advertisements to the base stations, leading the other nodes in the network into a false sense of security. A routing gap caused by malicious nodes might disrupt the normal functioning of the network. A sinkhole attack uses a compromised node near the target to boost the legitimacy of a route to the victim. The real node is duped into a false sense of safety by this alteration of routing information. As seen in Fig. 3, a sinkhole attack can be used to lure and steal packets from nearby nodes. Sinkhole attacks use a hidden tunnel to lure in nodes and steal data packets. When the base station received these packets, it was fooled by the rogue node.

2.4. Blackhole attack

Blackhole attacks capture and reprogram sensor nodes to block packets instead of receiving and forward to the base station [5]. In a black hole attack, the malicious node compromises the data after it enters the black hole zone. By dividing the network in this way, the black hole attack prevents critical updates from reaching the base station, hence degrading the network's performance. It has a major impact on network throughput and performance indicators. In Fig. 4, we observe that a packet is transmitted from node S, the transmitter, to nodes C and D, the receivers, on route to the base station. The Blackhole node consumes all information rather than passing it along.

The black hole attack performs suspicious activity using loopholes for discovering the routing [6]. To prevent packets from reaching their intended destinations, a black hole attack replaces a trusted node with a hostile one. The suspicious node selectively disables packet forwarding for a chosen set of nodes. The data is dropped from the black hole and transmitted to the control center. Black hole attacks cause the order numbers of routing requests and responses to be higher than those of a typical node. The normal node will not respond to the routing request with a higher order number, which causes routine deletion from the networks [7].

The contribution of the proposed framework for secure localization in wireless sensor networks is listed below:

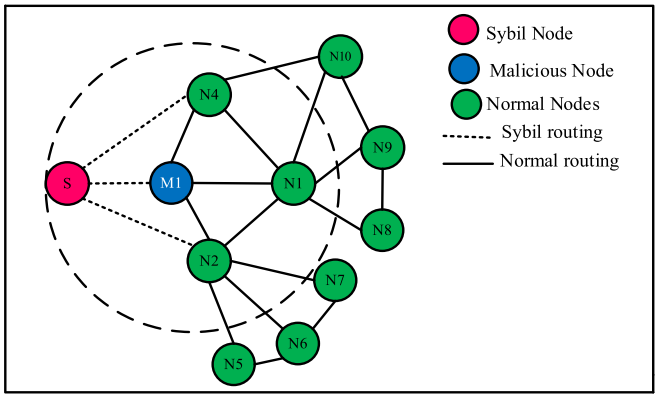


Fig. 1. Illustration of Sybil attack in wireless sensor networks generating multiple identities and routing paths.

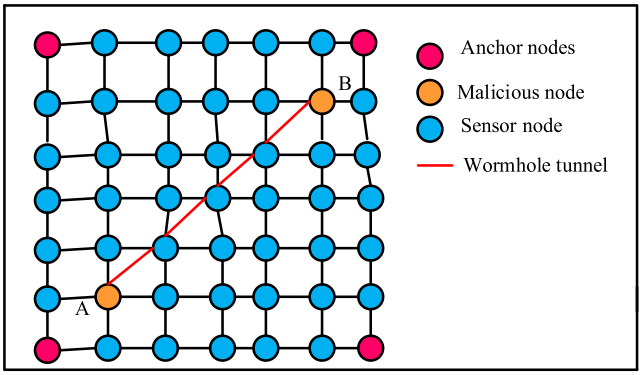


Fig. 2. Illustration of wormhole attacks in wireless sensor networks.

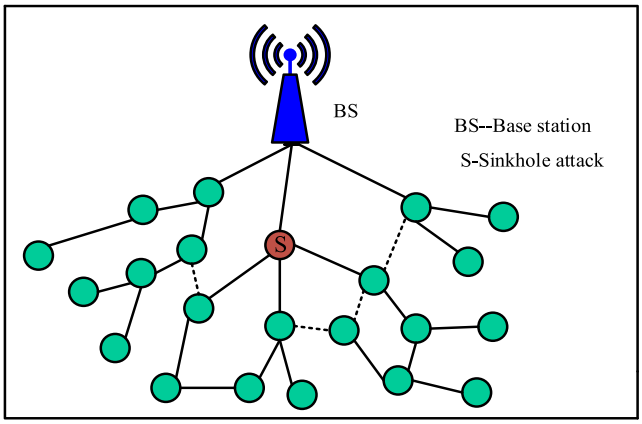


Fig. 3. Sinkhole attack and its rout advertising in WSNs.

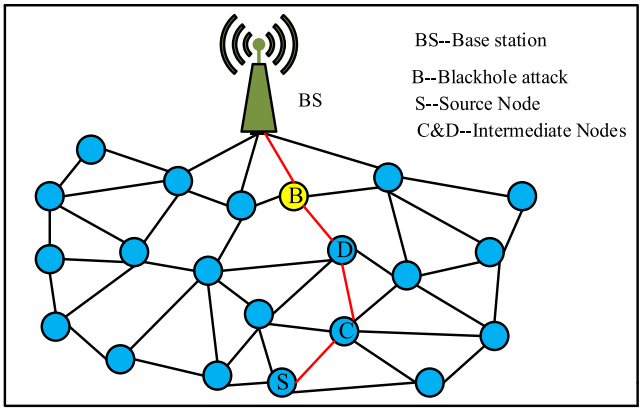


Fig. 4. Blackhole attack scenario in WSNs.

- Explore a range-free secure localization scheme using benchmark datasets based on federated machine learning techniques.
- Network planning and sensor deployment are suitable for computing the position and location of unknown nodes utilizing the minimum distance reference point.
- Enhance detection and localization accuracy of routing attacks in wireless sensor networks.
- To identify potentially harmful nodes, it is necessary to calculate the trustworthiness of the positions and locations of both known and unknown sensor nodes.
- Explore clustering and election of the beacon nodes for the localization process in the target field in wireless sensor networks to enhance trust value and security using a threshold reference point.
- Effective data processing techniques using feature engineering for classification by training the machine learning techniques
- Machine learning techniques are utilized to compute the average detection accuracy of malicious nodes in wireless sensor networks.

This paper aimed to use hybrid machine learning techniques to cut down on the estimation error when trying to locate nodes. This study models the proposed method in a WSN environment, where hidden sensor nodes can be located. The ramifications of using state-of-the-art algorithms to analyses data from unknown sensor nodes are discussed. This article contributes two major improvements to the localization of WSNs. To begin, we presented a simple and efficient localization algorithm that reduces the effect of localization mistake on WSN precision. Second, the proposed method minimizes the potential for error in WSN node localization. The outline of this paper consists of the following five parts: Section two is about the literature of the related works, followed by the third section details the network and attack models giving the routing information. Section four discusses the materials and methods, followed by section five with results and discussion. The last part of this work is the conclusion and remarks.

3. Literature review

The precision of localization outcomes is closely related to the quality of the anchor data used in the localization process. If some network anchors are compromised as malicious nodes sending false information to others, it is improbable that the unknown nodes will learn their exact location. As a result, the schemes need to consider security measures, including detecting and isolating harmful network anchors. Several distinct localized security techniques were presented, each catering to a certain use case.Y. H. Robinson et al. [1] Proposed a machine learning and 3-dimensional manifold localization technique for detecting and localizing unknown nodes in wireless sensor networks. They also discussed the detection of the fault nodes, improving localization accuracy, and reducing energy consumption. J. Chen et al. [2] Presented distance weighted hop distance using chicken swarm optimization to maximize unknown nodes' localization accuracy in wireless sensor networks. They also discussed the computation and estimation of unknown nodes using two-dimensional hyperbolic and nature inspire optimization techniques. A. Giri et al. [3] Proposed information-theoretic technique for detecting and localization Sybil attack in wireless sensor networks. They also discussed the entropy correlation coefficient for detecting Sybil attacks in the network and computing the localization accuracy of the malicious nodes. H. Chen et al. [8] addressed the problem of secure localization of the position of wireless sensor networks in an unattended environment against wormhole attacks using a distance-vector scheme with little computational processing and hardware cost. They also discuss the simulation and implementation of wormhole attacks to improve the malicious node's detection accuracy by computing the position of each node in the network using the distance vector protocol. B. Hasan et al. [9] examined intrusion detection techniques using artificial neural networks based on optimized energy in wireless sensor networks. They also discussed malicious node detection based on energy consumption and packet delivery using artificial neural networks and regression analysis for improving network security and performance. G. Farjamnia et al. [10] presented detection and localization of wormhole attacks using distance vector-hop using gap measurement and hop-size correlation with minimum distance and varying topology configuration to improve localization accuracy. They also discuss finding the shortest routing path to the anchor nodes and computing the localization error based on the distance vector information. R. Goyat et al. [11] Presented a secured range-free localization scheme based on the blockchain technique in WSNs to compute the unknown nodes' localization accuracy. They also discussed the trust values of the beacon nodes against malicious energy and mobility parameters in the network to select the miner node acting as the base station with topology variation. X. Li et al. [12] proposed a secured and range-free localization technique using an outlier elimination and vectorization approach to filter the beacon's position and malicious nodes in the network. They also discussed improving the localization accuracy of the unknown sensor nodes with and without malicious attack nodes. M. Beko and S. Tomic [13] addressed the secure target localization problem for randomly deployed wireless sensor networks in the presence of malicious attacks that manipulates the estimation process and disable the accurate localization of the unknown sensor nodes. They also presented the clustering, bisection techniques, and weighted central mass for finding the precise localization of the malicious node in the network. V. P. Kavitha and J. Katiravan [14] examined the hybrid adaptive network fuzzy logic control inference system (ANFIS) for secure and optimized localization techniques in wireless sensor networks. The scheme is also used to find the best-optimized location, cost, and distance with minimum distance energy

and better communication to the base station. They also discussed the selection and optimization techniques for enhancing the network lifetime. S. T. Patel and N. H. Mistry [15] presented Sybil node detection [16] using various schemes. F. Y. Yavu et al. [17] proposed detecting IoT routing attacks using a deep learning machine learning technique. Coojia's simulator creates realistic assault scenarios on an Internet of Things network equipped with a thousand sensors. V. Sujatha and E. A. M. Anita [18] examined the detection of Sybil attack detection using hybrid fuzzy and powerful extreme learning machines. L. Wang et al. proposed a real number hop-count technique for computing the location of unknown nodes by training hop count quantization using kernel extreme learning machines [19]. They tested certain algorithms and examined trade-offs. They define numerous feature vectors for transferring and dimension reduction the localizing problem onto distinct machine learning models in a unique way for accurate prediction and localization [20]–[22]. They examined how network size, anchor population, transmitted signal power, and wireless channel quality affect these models' localizing accuracy. The rest of the literature works are summarized as shown in Table 1.

4. Network model

The design and planning of wireless sensor networks that meet the security and performance for practical, secure location and position is a challenge in recent research [11]. The proposed network model consists of N number of sensor nodes randomly deployed without central management into two-dimensional environments with random mobility. The sensor nodes are organized into m beacon nodes and n unknown nodes in the network. The sink node broadcasts information to all the nodes and selects the beacon nodes based on the residual energy, minimum distance, trust value, and the number of neighbor nodes. The beacon nodes continuously monitor the sensor nodes to detect the malicious activity of the sensor node using the trust model. Sink nodes and beacon nodes are location and position-aware of their sites. They have higher energy and a large communication radius, enabling the establishment and estimate of randomly deployed sensor nodes, as shown in Fig. 5. Threats to the safety of wireless sensor networks arise when calculating the estimated location and position of the network's unidentified nodes. In WSNs using beacon nodes, the localization accuracy can be improved and the location can be calculated more precisely if malicious sensor nodes are detected.

The three criteria used to choose the beacon node from the pool of sensor nodes are as follows:

- The minimal distance a node must travel to reach the base station, as determined by the distance-vector.
- The potential energy left behind after activating a node.
- The intensity of the sensor node's incoming signal.
- Number of nearby nodes that are neighbors to sink node

Using the distance vector method, we can determine the separating distance, D, between any pair of sensor nodes (1):

$$D = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (1)$$

Where u and v represent the positions of the nodes i and j. It's likely that the node closest to the base stations is the cluster head. The multipath model is used to determine threshold values that characterize the amount of effort required for network model communication and activation. For a data transmission of k bits across a distance D and a threshold distance D_o we get (2) where E_{TX} is the energy required for transmission:

$$E_{TX} = \begin{cases} k \times E_e + k \times E_f \times D^2, \text{if} D \leqslant D_0 \\ k \times E_e + k \times E_m \times D^4, \text{if} D > D_0 \end{cases} \quad (2)$$

Table 1
Summary of recent works and research findings for secure localization nodes in WSNs.

Author,Year and Method	Performance	Research Findings	Limitation
T. K. Mohanta and D. K. Das, [23] 2023, selective opposition class topper optimization	The average localization error was 0.05 % for random beacon node deployment, 0.04 % for circular, and 0.08 % for spiral.	Spiral anchor deployment increases localization accuracy. The shortcut creates a localization network with model-specific inputs.	The idea is successful when applied to a specific kind of input traffic and when deployed using a certain number of nodes.
H. Sun et al. [24], 2023, Improved Adaptive Genetic Algorithm	The approach has the best convergence speed and is 11.3 %, 26.5 %, and 16.6 % more stable than the GA, PSO, and IAGA DV-Hops.	Improved adaptive evolutionary algorithms optimize unknown node coordinate values to increase localization accuracy.	uses a lot of computing power, which is noticeable when applied to larger tasks which requires assessing the fitness of population.
E. T. Fute et al. [25], 2023, Tabu based optimization method using RSSI	The unique optimization method reduces localization error by 39 % compared to network anchoring. Signal fluctuation affects localization precision.	range-based sensor localization technique. The technique improves the PSO algorithm by employing tabu search to find a particle's best local neighbour and particle evolution.	The RSSI signal strength typically has a limited range, which may not cover the entire area of interest, and this can result in incomplete localization
Y. Jin et al. [26], 2022, successive convex approximation method is designed to iteratively solve the optimization problem	By averaging RMSE across all unknown nodes, the suggested technique beats the PSO-DV-hop, CC-DV-hop, and DV-hop algorithms by 23.92 percent, 36.67 percent, and 45.90 percent.	Successive convex approximation solves the optimization issue iteratively. Localization accuracy is better than the particle swarm optimization approach with less computational effort.	SCA solves convex optimization issues, which can be computationally expensive, especially for big tasks. Each iteration's convex approximations determine the solution.
A. Singh et al. [27], 2020, machine learning approach based on Support Vector Regression (SVR) model.	Modified Cuckoo Search simulations yield these feature values. R-SVR had the highest correlation coefficient and Root Mean Square Error among the three approaches.	Three feature-standardization-based ALE prediction methods were provided. Anchor ratio, transmission range, node density, and iterations are ALE training and prediction features.	Training an SVR model can be computationally expensive, especially when dealing with large datasets. The algorithm requires significant amounts of time and computational resources to search for the optimal hyperparameters.
S. Messous and H. Liouane [28], 2020, online successive distance-vector hop scheme	The method's average localization error is 30 % lower than the DV-Hop, 5 % lower than the literature methodology, and 10 % lower.	progressively calculate node positions to improve multihop wireless sensor network node localization. Algorithms handle network anchor availability fluctuation.	In vast networks, the method converges slowly. This can delay network traffic, making it unsuitable for big, complicated networks.

(continued on next page)

Table 1 (*continued*)

Author,Year and Method	Performance	Research Findings	Limitation
S. Dong et al. [29], 2020, distance vector-hop by computing the distance and position of nodes	WSN node localization security is improved by the algorithm. The proposed approach lowers localization error by 3 % compared to DV-Hop with 50 beacon nodes.	The suggested approach outperforms the DV-Hop localization algorithm against Sybil assaults. The approach minimizes localization error by 3 % compared to DV-Hop with 50 beacon nodes.	Distance vector algorithms are not efficient in using network resources as they may send packets through longer paths due to incomplete or outdated routing information causing loops and congestion.
O. Cheikhrouhou and A. Koubaa [16], 2019, Blockchain-based secure localization using a public ledger that contains nodes position and the list of their neighbor nodes.	Anchor rate increases the safe algorithm's accuracy and decreases the unsecure algorithm's. Malicious anchor rates rise with overall anchor rates, and anchor node errors are worse.	These security techniques eliminate misleading data and prevent localization errors. The simulation results suggest that implementing the proposed security mechanism increases the localization accuracy in hostile nodes.	As the number of nodes in the network grows, the size of the public ledger will also increase, which could make it challenging to scale the system. The network resources are concentrated in the hands of a few entities.
X. Qi et al. [30], 2018, Multilateral positioning algorithm and improve the nodes' positioning accuracy	Distance connectivity and grid and random deployment simulations of the MA-MDS algorithm determine the threshold k and Error 95 % confidence interval is calculated.	Even with fewer anchor nodes, noise-sparse networks operate better. Prussian Analysis is used in the MDS positioning step to improve coordinate transformation and eliminate errors.	signals can be affected with by other signals or obstructions, causing position estimation inaccuracies. The number of anchor nodes may be limited, making node placements difficult to estimate.
P. Li, X. Yu et al. [31], 2017, Secure localization based on trust valuation process	Secure localization model inaccuracy decreases with attack power. Attack power increases secure localization model localization error if hostile nodes perform normally.	Normal localization algorithm error increases with attack power. Secure localization model localization error increases with attack power if malicious nodes.	Scalability becomes an issue as network nodes grow. Sybil attacks, when an attacker creates numerous identities to manipulate the trust valuation process, are vulnerable.
L. Song et al. [32], 2019, Improved localization algorithm based on hybrid chaotic strategy (MGDV-Hop)	Changing beacon nodes can calculate location inaccuracy. From 15 to 50 beacons, the hybrid chaotic strategy.	Chaos mutation and inertial weight optimize firefly travel distance in the hybrid chaotic strategy algorithm. Hybrid chaotic technique improves convergence, precision.	The hybrid chaotic localization algorithm may be difficult to understand and implement. Sensitive initial circumstances.

Where E_{TX} is the energy expended in sending a single bit of data E_f is the energy expended in receiving it, and E_e is the energy wasted by either the transmitter or the receiver. Various factors contribute to the energy dissipated, including the distance travelled by the signal, its filtering, its modulation, and its channel coding. With a data length of k, we have the following expression for the minimum acceptable transmission distance, d_o:

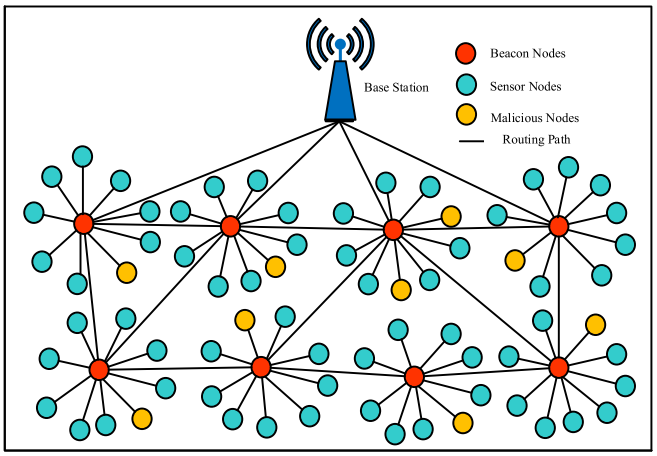


Fig. 5. Secure network model for localization of sensor nodes in wireless sensor nodes.

$$D_0 = \sqrt{\frac{E_f}{E_m}} \tag{3}$$

The amount of power needed by the receiving node to process k bits of a message is:

$$E_{RX(k)} = k \times E_e \tag{4}$$

Certain algorithms are required for precise target location, as well as for finding and evaluating the nodes' locations and positions. Both range-based and range-free localization methods are included in the system. The second choice, while more economically viable, requires specialized hardware. Received signal strength indicator (RSSI) and distance vector hop localization methods are examined for their ability to pinpoint and analyze the precise location of wireless sensor nodes. The distance vector localization procedure is essential to compute the coordinates of the sensor nodes and cluster heads using the beacon nodes [28]. The system determines and adjusts the distance between the unnamed nodes based on their calculated positions. In WSN, the distance vector hop process is used to locate empty spots among the beacon nodes. The average hop size is utilized in the distance vector approach to compute the smallest possible distance. This algorithm was first identified by [29]. The distance vector localization scheme is a range-free strategy [33] with a series of steps.

Routing phase: The beacon node initiates the routing process by broadcasting a message to all of the sensor nodes. As soon as it enters the network, the hop count associated with its position data is set to 0. [33]. A node that gets such a message will determine where the beacon node is located, append its own identifier to the message, and then broadcast it to its neighbors while increasing the hop count [8].

Computing distance: Using the beacon node, we can determine the average hop size and the distance to the unknown node in the network as in (5).

$$D(N(i), N(Sink)) = D(N(i), CH(i)) + \sum_{j=1}^{n} D(CH(j), CH(j+1)) + D(CH(n-1), Sink) \tag{5}$$

Where D is distance, to the Sink – depicts the total distance travelled by all sensor nodes (N) in order to reach the sink node. It is represented by the equation (5), and it indicates how far each sensor node has travelled. Whereas n is the number of hops from the cluster head CH (i, j) to the destination node i.e. the sink node.

Position computation and estimation: Geometric calculations like triangulation, the polygon technique, and trilateration are frequently employed as the underlying basis for determining the locations of unknown nodes [34]. There are a number of methods that can be used to calculate the separation of two nodes; they include synchronization, radio signal intensity, and the physical features of the carrying wave [35]. The average distance hop for the anchor node is computed and obtained relative to another beacon with the minimum hop count given by equation (6):

$$HS_i = \frac{\sum_{i^1 j} \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}}{\sum_{i^1 j} hij} \tag{6}$$

Where i and j true anchor nodes, (u_i, v_i),(u_j, v_j), the known and true coordinates for i and j, hij hop counts of the anchor nodes and HS_i average hop distance

Anchor node transmits its information followed by hop-size calculation [29]. The distance between sensor node and anchor computed with hop-size details is given by equation (7)

$$D_{pk} = H_p Si_i \times hp_{pk} \tag{7}$$

Each node's estimated position (P) can be calculated using the polygon method. An unnamed node, marked by the coordinates (u, v), and its surrounding region, denoted by the symbol Di, are located at point P. Given a network with n beacon nodes, the predicted location of the out-of-place node p is given by the equation (8) [29].

$$\begin{aligned} (u - u_1)^2 + (v - v_1)^2 &= D_1^2 \\ (u - u_2)^2 + (v - v_2)^2 &= D_2^2 \\ &\cdot \\ (u - u_n)^2 + (v - v_n)^2 &= D_n^2 \end{aligned} \tag{8}$$

We can get a set (n-1) of expressions subtracting from the first equations to make the system linear, given as depicted in equation (9):

$$\begin{aligned} u_1^2 + v_1^2 - u_n^2 - v_n^2 - 2(u - v_n)u - 2(v_1 - v_n)v &= D_1^2 - D_n^2 \\ u_2^2 + v_2^2 - u_n^2 - v_n^2 - 2(u_2 - v_n)u - 2(v_2 - v_n)v &= D_2^2 - D_n^2 \\ &\cdot \\ u_{n-1}^2 + v_{n-1}^2 - u_n^2 - v_n^2 - 2(u_{n-1} - v_n)u - 2(v_{n-1} - v_n)v &= D_{n-1}^2 - D_n^2 \end{aligned} \tag{9}$$

Rearranging the previous equations into the formula ofui = B A^{-1}, where A, ui,and B are expressed as in equations (10)and (11), separately.

$$A = \begin{Bmatrix} 2(u - u_n)2(v_1 - v_n) \\ 2(u_2 - u_n)2(v_2 - v_n) \\ \cdot \\ 2(u_{n-1} - u_n)2(v_{n-1} - v_n) \end{Bmatrix} \tag{10}$$

$$B = \begin{Bmatrix} u_1^2 + v^2 - u_n^2 - v^2 + D_n^2 - D_1^2 \\ u_1^2 + v^2 - u_n^2 - v^2 + D_n^2 - D_1^2 \\ \cdot \\ u_{n-1}^2 + v_{n-1}^2 - u_n^2 - v_n^2 + D_n^2 - D_{n-1}^2 \end{Bmatrix} \tag{11}$$

$$U_i = \begin{pmatrix} u \\ v \end{pmatrix} \tag{12}$$

By calculating the least squares equation, we may determine the node's location by equation (13).

$$U = (A^{'}A)^{-1}A^{'}B \tag{13}$$

In contrast to its range-based equivalents, RSSI-based localization algorithms have gained a lot of traction in the academic community for a variety of compelling reasons [36]. Measurements of received signal strength indicator (RSSI) and data transmission to higher stack layers are commonplace in modern wireless sensor nodes. There is no requirement for precise time-of-arrival calculations, Ultrawide band (UWB) radios, or antenna arrays with RSSI-based localization. Both the software and hardware requirements for achieving node localization are minimal. The DV-Hop algorithm, on the other hand, does not take advantage of the actual distances between nodes in the immediate vicinity to improve localization accuracy in large-scale wireless sensor networks.

To improve localization accuracy and malicious node detection, the hybrid strategy adds two stages beyond just employing the DV-Hop technique to pinpoint WLAN nodes. We start by using the RSSI data to estimate the distances between the anchor nodes and their one-hop surrounding sensor nodes, as opposed to relying on the average hop distance as the original DV-Hop algorithm did. In most modern wireless sensor nodes, the MAC sub-layer automatically calculates the RSSI value for each received packet, so there's no need for any additional hardware or money to use this information. Second, after identifying a sensor node N, it is promoted to the anchor status and used to help identify other sensor nodes. The remaining sensor nodes can be localized more precisely with the use of additional (repurposed) anchor nodes. When there are fewer anchor nodes in a wireless network, this is very helpful. In order to arrive at the best possible answer to a problem, evolutionary computers employ a method called differential evolution (DE), which involves repeatedly trying to improve a candidate solution with regard to some quality parameter. Metaheuristics are methods that make little or no assumptions about the problem at hand in order to search through vast spaces of potential solutions. Unfortunately, metaheuristics such as DE cannot guarantee that you will always obtain the optimal solution.

Unlike conventional optimization methods such as gradient descent and quasi-newton methods, DE does not require knowledge of the gradient of the optimization problem, making it applicable to problems involving optimization of multidimensional real-valued functions. In this way, DE can be used to solve optimization issues that are fundamentally discrete, noisy, dynamic, etc. DE optimizes a problem by retaining a population of candidate solutions, creating new candidate solutions by the merger of existing ones, and finally retaining the candidate solution with the best score or fitness. Since the optimization problem is treated as a black box that provides only a quality measure once a candidate solution is provided, the gradient is superfluous.

4.1. Attack model

Several attack localization schemes are conducted in wireless sensor networks for various classes of attacks. The proposed approach aims to enhance the detection accuracy of security localization [29] to routing attacks using the distance vector hop procedure and clustering protocols in WSN. Fig. 6 depicts a model of a Sybil attack using three groups of wireless nodes. The attack model illustrates how malicious nodes employ deceptive behaviors by generating numerous identities to target the real node's position and location via different routing methods. The malicious node produces multiple fake identities similar to the honest

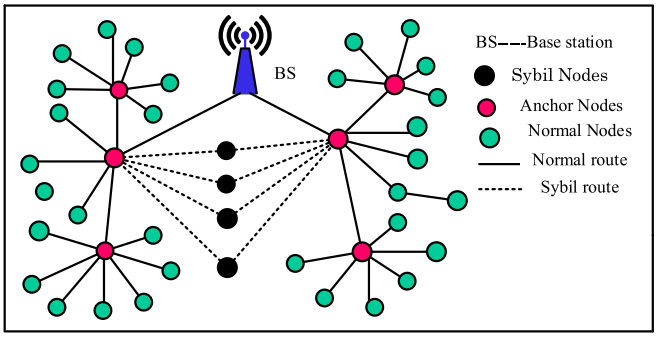


Fig. 6. Illustration of routing and information manipulation Sybil attacks in WSNs [21].

nodes with different locations and routing information. This malicious node is Sybil attacks acting as the beacon node and sends a message to the sensor nodes gaining the network resources and information that makes the network busy.

The Sybil attack steals the location and identity of legitimate nodes by creating new identities as new beacon nodes, making itself a normal node.

The malicious attack interrupts the normal functioning of the scheme by manipulating the routing information, computing the location and distance information, and the verification process [11]. The malicious attack compromises the beacon node to acquire the identity of the legitimate node for accomplishing the activities of the genuine node, including:

- Malicious nodes captured and manipulated false packets, forged locations, and fabricated identities.
- The attacker acquires the transmission range of the beacon nodes for targeting its location
- Malicious nodes disrupt the evaluation process by altering the integrity and information of the sensor nodes
- Malicious nodes broadcast and forward false energy information in the network, tampering with the integrity of the information

5. Materials and methods

The proposed system, as shown below in Fig. 8, depicts the overall architecture and various phases and procedures, including sensor deployment and data collecting, computing position and location, data transmission and aggregation, data processing using feature engineering and sampling techniques, attack detection analysis, and classification for implementing secure localization and detection of in wireless sensor network using machine learning models [1]. Hyperparameter and Bayesian optimization techniques are used to enhance the performance of hybrid machine learning classification models. The k-means clustering is also used for binary classification and further improves the proposed scheme's attack detection performance.

5.1. Sensor deployment and clustering

Wireless sensor nodes are deployed randomly, containing the base station, cluster head, and sensor nodes in two dimensional with transmission radius in the target field. We Assumed that the cluster head and base station are location-aware and can compute the location and position of the other unknown nodes. The cluster head processes, verify, and forwards the data to the base station.

The proposed system's efficiency is demonstrated by deploying sensor nodes, authenticating nodes, collecting information, and selecting beacon nodes for estimating the location of unknown sensor nodes. We also implement the processes of secure clustering and routing, estimation of sink node location, and updating the location.[37]. Sensor nodes are deployed and clustered as in Fig. 7. Here, we look at WSNs with a hexagonal grid. In a two-dimensional network, "N" nodes are dispersed randomly, with "M" nodes serving as anchor nodes because their locations are known. The total number of unidentified nodes that need to be pinpointed is (N-M). All nodes first have their data enrolled at the Sink Node. After the cluster head is selected based on calculated parameters, including distance, residual energy, and node degree, the authentication procedure is built upon the first stage of the sensor deployment to reduce network connection overhead and maximize effective data transmission. Registering a node at a sink node allows it to obtain Pseudo-Random Numbers, which can be used to distinguish between different nodes. Nodes have been selected as anchor nodes with high computational processing and power. For reliable data transmission, nodes constantly moving throughout the network will adopt a cluster-based routing strategy.

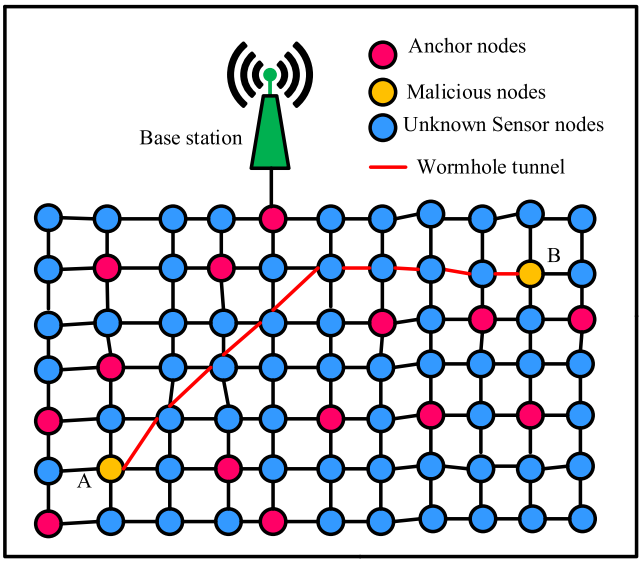


Fig. 7. Configuration of dimensional sensor deployment for hierarchical topology in wireless sensor networks.

5.2. Routing and communication strategies

Clustering-based hierarchical routing protocol uses overall transmission distance minimization and balanced node energy consumption across the network's lifespan to maximize energy utilization and extend its lifetime [38]. This demonstrates that the protocol works quite well. Clustering procedure is used to divide the region being sensed into smaller, more manageable sections. Each cluster has a designated leader, known as the cluster head (CH), who is in charge of passing on information to the other members. After collecting data from the nodes in the cluster, the head node (CH) sends the resulting fusion to the base station (BS). It is important to effectively disperse the cluster structure and optimize the selection of CHs when using a clustering-based protocol. The optimal choice of CH is seen as the most important component in clustering-based routing protocols, among the many other current difficulties.

The chosen nodes can then keep an eye on the closest sensor nodes within their range of communication [39]. Ad hoc sensor networks can benefit from this method of node election. Due to the open nature of the wireless medium, the security strategy is put to the test in both methods in order to protect the confidentiality of the communication channels. In order to protect the monitor node selection process, which is still vulnerable to sophisticated attacks, modern methods employ the key management methodology. The original data and the secret keys can be easily revoked by an opponent who has detected the unprotected channel and has access to the necessary cryptanalysis hardware and software components. An intelligent attacker can disrupt data transmission and sensor node functionality by learning where these nodes are located.

5.3. Characterization of routing attacks

Wireless Sensor Networks (WSNs) are vulnerable to different types of attacks due to the open nature of wireless communication. Routing attacks are one of the most common types of attacks in WSNs, where an attacker tries to disrupt the communication between nodes by compromising the routing protocol. These attacks can be characterized based on their pre-attack and post-attack behaviors.

Pre-Attack Characterization:

Vulnerability Analysis: In this stage, the network is analyzed to identify potential vulnerabilities that can be exploited by the

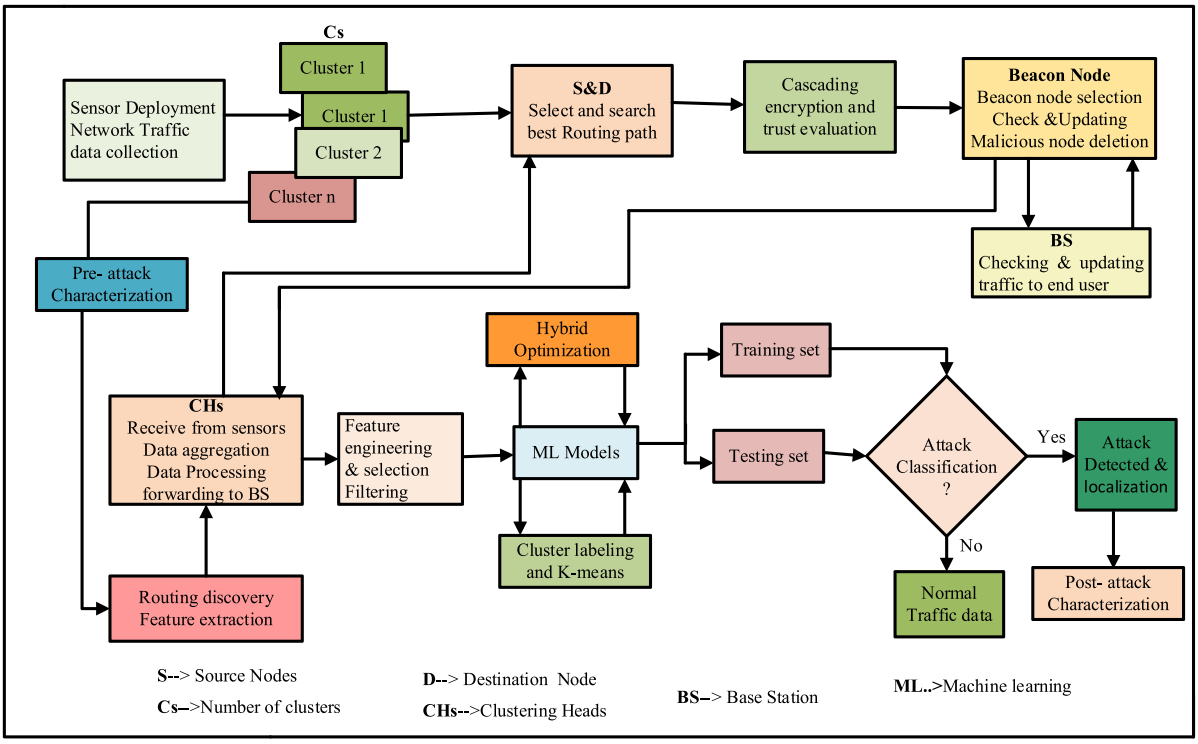


Fig. 8. Illustration of blockchain-based proposed localization technique in wsns using federated and hybrid machine learning (ml) models.

attacker. This involves examining the network topology, routing protocols, and security mechanisms deployed in the network.

Attack Preparation: In this stage, the attacker gathers information about the network topology, the location of the nodes, and the routing protocols used. The attacker also selects the attack method and determines the required resources.

Attack Initiation: In this stage, the attacker launches the attack on the network by exploiting the identified vulnerabilities. The attacker can use various techniques to disrupt the network, such as selective forwarding, sinkhole attacks, and Sybil attacks.

Post-Attack Characterization:

Detection: In this stage, the network administrator detects the attack by monitoring the network traffic and observing abnormal behavior. Various detection techniques can be used, such as anomaly-based and signature-based detection.

Analysis: In this stage, the network administrator analyzes the attack to determine its scope and impact. The analysis can help in identifying the compromised nodes, the type of attack, and the attacker's location.

Recovery: In this stage, the network administrator takes corrective measures to recover the network from the attack. This can involve replacing compromised nodes, reconfiguring the routing protocol, or deploying additional security mechanisms.

Overall, pre-attack and post-attack characterization of routing attacks can help in improving the security of WSNs by identifying vulnerabilities, detecting attacks, and taking corrective measures.

5.4. Cascading encryption and trust evaluation

Cascading encryption and trust evaluation are two important security mechanisms that can be used in wireless sensor networks (WSNs) to ensure the confidentiality and integrity of data transmitted over the network. Cascading encryption involves the use of multiple layers of encryption to protect data transmitted over the network. In this approach, data is first encrypted using a symmetric key encryption algorithm, such as Advanced Encryption Standard (AES), and then encrypted again using a public-key encryption algorithm, such as RSA. This double encryption provides an added layer of security that makes it more difficult for an attacker to intercept and decipher the data. Trust evaluation, on the other hand, involves evaluating the trustworthiness of nodes in the network to prevent malicious nodes from compromising the security of the network. Trust evaluation can be based on a number of factors, such as the node's behavior, history, and reputation, and can be used to detect and isolate malicious nodes from the network.

When used together, cascading encryption and trust evaluation can provide a strong level of security for wireless sensor networks. Cascading encryption can ensure that data transmitted over the network is protected from interception and deciphering, while trust evaluation can help to identify and isolate malicious nodes that may attempt to compromise the security of the network. However, it is important to note that these security mechanisms can also have a negative impact on the performance of the network, as they require additional computational resources and may introduce latency into the system. Therefore, it is important to carefully balance the level of security with the performance requirements of the network.

5.5. Transforming techniques

To transform secure localization techniques into a machine learning approach for routing attack detection and localization, we need to first collect training data as shown in Fig. 9. This training data should consist of network traffic data that includes information about the location of nodes, as well as any attacks that have been detected. Machine learning can be used to help detect and localize malicious nodes in a network. One approach is to use a supervised learning algorithm that is trained on a dataset of known malicious nodes and normal nodes. The algorithm can then be used to classify new nodes as either malicious or normal based on their features. To make this approach more secure, the training dataset should be representative of the network being protected. The dataset should be diverse and include different types of malicious nodes and normal nodes. The features used for classification should also be carefully selected and should not leak any sensitive information about

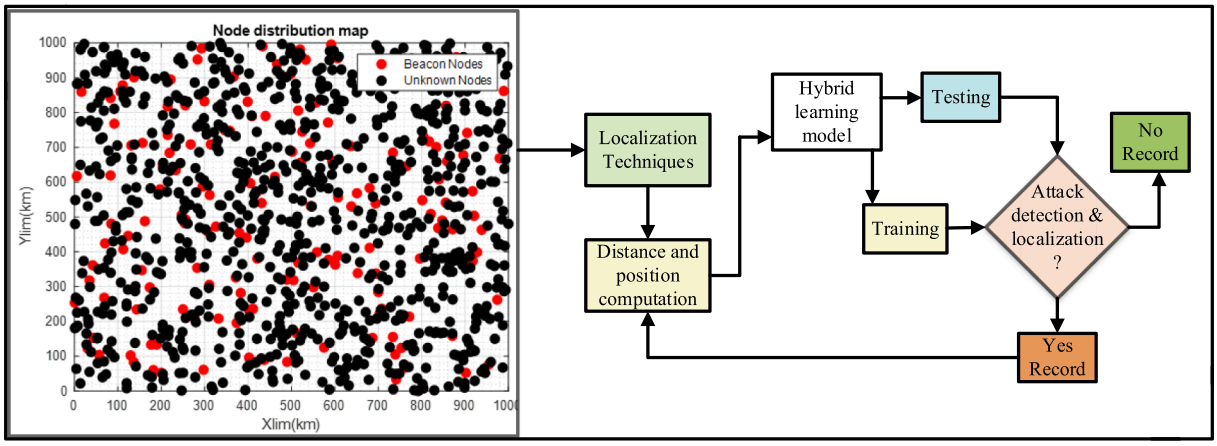


Fig. 9. secure localization techniques for detection and localization of malicious attacks using in WSNs.

the nodes. Machine learning can be a powerful tool in detecting and localizing malicious nodes in a network. Here's one approach for doing so in a secure manner:

Data collection: Collect network traffic data and identify the features that may be indicative of malicious behavior, such as excessive data transfer, abnormal network activity, or unauthorized access attempts.
Data labeling: Label the data based on whether or not each node is malicious.
Feature engineering: Extract relevant features from the labeled data to be used in the machine learning model. Feature engineering is an important step as it can greatly impact the accuracy and effectiveness of the model.
Model selection: Select a machine learning algorithm that can identify and locate the source of network intrusions.
Model training: Train the machine learning model on the labeled data using the chosen algorithm. Ensure that the training process is secure by using appropriate measures such as encryption, secure communication, and access control.
Model evaluation: Evaluate the trained model's performance using a separate set of data that was not used in the training process. Adjust the model parameters and features as necessary to improve its performance.
Deployment: Deploy the trained model to detect and localize malicious nodes in the network in real-time. Ensure that the deployment process is secure by using appropriate measures such as secure communication protocols and access control.

It is important to note that machine learning is not a silver bullet for detecting and localizing malicious nodes. It should be used in conjunction with other security measures such as firewalls, intrusion detection systems, and access control mechanisms. Additionally, the effectiveness of the machine learning model may diminish over time as attackers adapt to its detection methods, so it is important to continually update and improve the model.

This section will discuss the detection of blackhole and sinkhole attacks procedures, and techniques using the clustering and routing protocols. This section also focused on detecting and classifying routing attacks in WSNs using machine learning techniques using benchmark datasets. The proposed system in implemented using different phases for detection and analysis of routing attacks in wireless sensor networks, as shown in Fig. 9.

Sensor nodes are deployed randomly in the target environment to collect data from the surroundings and send it to the cluster head. They are grouped into clusters with the help of the routing protocol and select cluster head. The cluster head then forwards it to the sink node. The intermediate nodes are used to select the best path and reducing energy consumption in the network. The cluster head also aggregates and processes using feature selection and data transformation for verification. The autoencoder performs binary classification by training the model with an activation function. It is also used for optimization and regularization with seven layers. This enables the selection best machine learning model for the decision-making process and detection and classification of attacks using the benchmark dataset.

5.6. *Benchmark datasets*

Both the CICIDS2017 and UNSW-NB15 network intrusion detection datasets are freely accessible for use in academic studies of IDS effectiveness. The Canadian Institute for Cybersecurity (CIC) has released the CICIDS2017 dataset, which includes both benign network activity and malicious assaults like denial-of-service (DoS), probing, and malware distribution. The dataset is designed to reflect real-world network traffic and contains over 80 million network flows. The UNSW-NB15 dataset was created by researchers at the University of New South Wales (UNSW) and contains network traffic data from a simulated environment. The dataset includes both normal traffic and a range of attacks, including DoS, port scanning, and infiltration. The dataset is designed to be comprehensive and includes a variety of network protocols and attack scenarios. Both datasets are widely used in research to evaluate the performance of IDSs, and many studies have been conducted using these datasets to compare and contrast different detection techniques. These datasets are also freely available to the research community, which makes them a valuable resource for developing and testing new IDSs.

The CICIDS2017 and UNSW_NB15 sample of datasets are used as a benchmark for evaluating the effectiveness of the proposed system using machine learning models for the detection and classification of the class of attacks. The Tools for intrusion detection and network attack scenarios were used to acquire and run the dataset, which included seven types of attacks. Table 2 shows the frequency distribution of each class of

Table 2
Frequency distribution of attacks CICIDS2017 dataset.

Class of attacks	Frequency	Percent	Valid Percent	Cumulative Percent
BENIGN	7315	26.0	26.0	26.0
Bot	1806	6.4	6.4	32.5
BruteForce	2767	9.8	9.8	42.3
DoS	8958	31.9	31.9	74.2
Infiltration	36	0.1	0.1	74.3
PortScan	5041	17.9	17.9	92.2
WebAttack	2180	7.8	7.8	100.0
Total	28,103	100.0	100.0	

attack samples for training and testing. The dataset includes typical and widespread types of assaults based on real-world data, such as network traffic monitoring. The dataset is particularly diverse, with different kinds of attacks represented throughout 80 feature sets produced with CICFlowMeter and made publicly available online.

For intrusion detection, the UNSW-NB15 is a cutting-edge benchmark dataset that has been utilized extensively in recent years. The raw data traffic packets were generated using the cyber range laboratory IXIA PerfectStorm tool from the Australia Center for cyber security (ACCS) [40], creating hybrid normal and abnormal network traffic packets. The software continuously updates the vulnerabilities and exposures of collected packets in the context of information security, simulating nine different types of assaults. There are a total of 42 properties in the dataset, including 39 numerical and 3 category variables. Table 3 shows the distribution of the 10 types of attacks found in the dataset. When developing predictive models for classification and regression tasks with machine learning models, the dataset is split into a training portion (80 %) and a testing portion (20 %).

5.7. Data Pre-Processing

Improving the data quality for training and testing is the first step in the pre-processing phase of constructing predictive machine learning models. Duplicate values can be removed, the mission data can be replaced, and unneeded sample structures can be eliminated. Normalization using minimum and maximum scaling values as in after cleaning the dataset is mandatory as in equation (14).

$$Z_{norm} = \frac{Z - \min(i)}{\max(Z) - \min(Z)} \tag{14}$$

Where min(z) and max(z) represent the least and maximum values of attribute Z, respectively. Z_{norm}, is a normalized feature value, and Z is an original feature value [41].

K-means cluster sampling is utilized to improve the classification accuracy of the machine learning model by generating a small K-number of clusters of the original dataset to reduce the training complexity [42]. In order to maximize productivity, processing power, and resources, the K-means sampling method creates highly representative small groups by eliminating duplicate data. To address the issue of class bias, researchers developed a method called synthetic minority oversampling (SMOTE). Data pre-processing is followed by feature engineering to compute the correlation features between features in order to generate sensitive, high-quality features while also reducing dimensionality and removing redundant ones.

5.8. Machine learning models

To enhance performance and test the efficacy of the proposed system, a number of machine learning techniques are implemented, such as XGBoost, Decision tree, Random forest, Extra tree, and Ensemble stacking, and benchmark datasets with a number of evaluation metrics are used. These methods for extracting information from and discovering connections among wireless sensors facilitate the identification of malicious nodes in a network and are both effective and useful.

Table 3
Frequency distribution of attacks in the UNSW_NB15 dataset.

Class of Attacks	Frequency	Percent	Valid Percent	Cumulative Percent
Analysis	677	0.8	0.8	0.8
Backdoor	583	0.7	0.7	1.5
DoS	4089	5.0	5.0	6.5
Exploits	11,132	13.5	13.5	20.0
Fuzzers	6062	7.4	7.4	27.4
Generic	18,871	22.9	22.9	50.3
Normal	37,000	44.9	44.9	95.2
Reconnaissance	3496	4.2	4.2	99.5
Shellcode	378	0.5	0.5	99.9
Worms	44	0.1	0.1	100.0
Total	82,332	100.0	100.0	

Machine learning models are computer programs that use statistical algorithms and mathematical models to identify patterns and relationships in data. These models are trained on large datasets and use the patterns and relationships they discover to make predictions or classifications about new, unseen data. There are many different types of machine learning models, each suited to different types of problems and data. Some of the most common types include:

1. Linear Regression: a model that tries to find a linear relationship between a set of input variables and a target variable.
2. Logistic Regression: a model that predicts the probability of a binary outcome (e.g. whether a customer will buy a product or not).
3. Decision Trees: a model that uses a tree-like structure to make decisions based on a set of input features.
4. Random Forest: an ensemble model that combines multiple decision trees to improve accuracy and reduce overfitting.
5. Support Vector Machines: a model that tries to find the best boundary (i.e. hyperplane) to separate two classes of data.
6. Neural Networks: a model that uses layers of interconnected nodes (i.e. neurons) to learn complex relationships in data.
7. Clustering: a model that groups similar data points together based on their features.
8. Dimensionality Reduction: a model that reduces the number of features in a dataset while retaining as much information as possible.

Each of these models has its own strengths and weaknesses, and choosing the right one for a particular problem depends on the nature of the data and the goals of the analysis.

5.8.1. Decision tree

Decision Tree (DT) is a supervised machine learning technique for building classification models using tree structures having significant test attributes of the instance from each branch to connect the root node to the leaf node. A DT is applied for classification and regression tasks using a tree-like structure for computing the predictions from the input vector of the root node [40]. The leaf nodes predict the detection and classification accuracy of the various classes of attacks. The scheme is used to build other techniques, including Gradient Boosted, random forest, decision tree, etc. localize and detect the malicious nodes in the network. This technique is easy to implement and automatically manages missing values network traffic collecting. The decision tree is based on the divide and conquer strategy of decision and leaf nodes representing the test attributes and class value [43]. The classification accuracy can be improved by increasing the generated decision tree.

5.8.2. Random forest

Random Forest (RF) is an ensemble learning model that operates by constructing many Decision Trees for predicting a class [44]. In order to handle high-dimensional data, Random Forest evaluates the significance of features in addressing overfitting and stability issues, hence lowering the variance. This makes the random forest method useful for identifying and categorizing harmful assaults in wireless sensor networks based on a diverse benchmark dataset. For noisy data, the random forest is both missing-value-aware and outlier-resistant. The scheme's need for increased computational power and resources is a direct result of the many trees that must be built. Random Forest constructs a forest of several decision trees to predict the model's detection accuracy [45]. It is successful in detecting and classifying attacks using a benchmark dataset.[46].

5.8.3. Gradient boost

Extreme Gradient boosting (XGBoost) is a classification technique for large datasets with a minimum amount of time, making it popular

nowadays [47]. The Gradient Boosting technique uses the extraction of important features to improve the computational speed and provides a precise output for intrusion detection by reducing memory consumption during the training and testing of the dataset for classification [48]. As demonstrated, this machine learning method is useful for optimizing the loss function and calculated features in equation (15).

$$\Phi(X) = \sum_{k=1}^{K} f_k(X)\ \mathrm{f}_k \varepsilon F \tag{15}$$

Where $\phi(X)$ is the final result of the K sequential classifier, whereas the f_k is the threshold function for the Gradient descent algorithm with K iterations. The technique uses parallel computing to achieve the desired classification outcomes. In order to optimize the process and control the overfitting component, XGBoosting enhances the Gradient descent and regularization technique. The equation depicts the relationship between the classifiers' parameters. (16) shown below:

$$\ell(\phi)_t = \sum L(f_{t-1} - f_t) + \Omega(f_t) \tag{16}$$

Where $\mathscr{L}(\phi)_t$ is the loss function and $\Omega(f_t)$ optimally adjusting the step size t, where is the regularization term. The feature metrics are used by the Gradient Boosting method to derive the scores for each attribute based on the features.

5.8.4. Extra trees

The extra tree (ET) technique is a tree-based algorithm similar to the random forest machine learning model and utilizes an ensemble decision tree for the classification and regression process [40]. The ET technique also creates a randomization layer for maintaining and optimizing the process by combining a collection of randomized decision trees built on various subsets of the dataset [42]. It is an ensemble model with multiple decision trees for building effectively linear and complex data traffic classification. It is an estimator that fits a randomized decision tree on a different section of samples of the dataset using normalization to increase the accuracy and control the over fit of the data [49].

5.8.5. Ensemble learning

Ensemble machine learning techniques are classifiers with averaged accuracy to reduce the risk of overfitting and bias from a single classifier [50]. Accuracy is improved by a tree-like structure used in the machine learning models used for classification. Ensemble techniques are *meta*-algorithms combining various machine learning techniques into a unified machine learning prediction model for evaluating variation and stacking [51]. Using the Ensemble method, multiple machine learning outputs can be combined into a single, more accurate model. The three most fundamental ideas in ensemble techniques are stacking, bagging, and boosting. In this setup, machine learning models are stacked to increase their predictive ability.

5.8.6. Optimization techniques

Hybrid machine learning models combine two or more different types of machine learning algorithms to improve the overall performance of the model. However, training and optimizing hybrid models can be challenging due to the complexity of the model and the need to optimize multiple parameters. Hybrid optimization techniques can be used to address this challenge and improve the performance of hybrid machine learning models. One approach to hybrid optimization is to use a combination of gradient-based optimization techniques and metaheuristic algorithms. Gradient-based techniques, such as stochastic gradient descent (SGD), can be used to optimize the parameters of the individual models in the hybrid model. Metaheuristic algorithms, such as genetic algorithms, particle swarm optimization, or simulated annealing, can be used to search for the optimal combination of models and their parameters. This approach can be used to find a set of optimal parameters for each individual model and then combine these models using an optimization algorithm to find the optimal combination.

The suggested system uses a pair of optimization methods. By combining a tree based on the Parzen estimation (BO-PTE) with hyperparameter and Bayesian optimization techniques, the suggested scheme is able to improve its classification of machine learning models on the benchmark dataset. Every single machine learning task makes use of hyperparameters to fine-tune the aforementioned parameters and get optimal results. This hyperparameter optimization (HPO) reduces human efforts and improves machine learning performance [52]. Hyperparameter optimization is used in both black box and global optimization to improve performance metrics. This allows us to explain how black box Bayesian Optimization works. Bayesian optimization (BO) is a framework for global optimization that uses pricey black-box functions, and it is gaining popularity in HPO for deep neural networks. Bayesian optimization is a recursive method that employs a probabilistic surrogate model and an acquisition function to assess the junctures at which choices must be made by way of the Gaussian process. Tree-based approaches, such as random forest and tree Parzen estimators (PTE), handle the hyperparameters. Bayesian-based optimization with tree Parzen estimators (BO-PTE) is introduced in this study for determining the optimal point of assessment in automatic machine learning.

6. Experimental setup

In this section, we'll go over the configuration and evaluation metrics using the simulation environment. The sensor nodes and sink nodes are assumed to be static after deployment. After deployment, the anchor nodes are dynamic and movable for localizing the unknown nodes. The sensor nodes have their unique IDs and location and activation energy information. The sink node is known to every sensor node and is located outside the clustering boundary. Wireless sensors are distributed randomly with 60 beacon nodes, 240 unknown nodes, and 35 malicious nodes in the target field with an area of 1000x1000 m^2. The transmission radius of each beacon node and sensor node is 250 m. The Sybil node's precise location is determined by employing the Distance Vector hop algorithm. The configuration of the simulation's parameters is displayed in Table 4. Planning and simulating the network is done in MATLAB R2021a on a Windows machine running a 64-bit OS and an Intel(R) Xeon(R) Silver 4214 CPU at 2.20 GHz 2.19 GHz (2 processors) with 128 GB (128 GB useable). The python anaconda toolboxes are used for data processing and analysis to evaluate the proposed method's performance using the dataset as a benchmark [8].

6.1. Performance metrics

Using the CICIDS2017 benchmark dataset for attack categorization, we evaluate the efficacy of the proposed method by calculating the localization error (LE), average localization error (ALE), localization accuracy of the unknown sensor nodes, and the confusion matrix. Average localization error (ALE), average localization accuracy (ALA),

Table 4
Simulation parameters.

Parameters	Values
Software	MATLAB
Deployment	Random
Number of nodes	300
Simulation area	1000x1000 m^2
Protocol	Routing and clustering
Number of beacons	60
Transmission Radius	250 *m*
Unknown nodes	240
Model	Regular
Malicious nodes	35 %

accuracy, detection rate precision, and recall are the measures used for evaluation. Equation (17) is used to calculate the average error localization, abbreviated as ALE [29]. The ALE is calculated by adding the LE of all the unknown nodes and dividing by the total number of unknown nodes. The LE is defined as the discrepancy between the predicted and observed locations of any undiscovered nodes.

$$LE = \sqrt{(u_i^{'}-u_i)^2 + (v'_i-v_i)^2}$$
$$ALE = \sum_{i=1}^{n} \frac{\sqrt{(u_i^{'}-u_i)^2 + (v'_i-v_i)^2}}{nR} \tag{17}$$
$$ALA = (1 - (\sum_{i=1}^{n} \frac{\sqrt{(u_i^{'}-u_i)^2 + (v'_i-v_i)^2}}{nR})) \times 100\%$$

Where (u'_i, v'_i) are the actual coordinates of the anonymous node i and (u_i, v_i) are the calculated coordinates, n represents unknown nodes, and R is the network's communication range. Using the localization approach, we determine the position and error of each node. The performance metrics, including detection, false alarm, precision, and recall, measure the effectiveness of the proposed system. Mathematically they can be expressed as follow in equations (18–19):

$$\begin{aligned} \text{Detection Rate(DR)} &= \frac{TP}{TP + FN} \\ \text{Recall (RC)} &= \frac{TP}{TP + FN} \\ \text{Precision} &= \frac{TP}{FP + FN} \\ \text{Specificity} &= \frac{TN}{FN + TP} \\ \text{False positive rate} &= \frac{FP}{FP + TN} \end{aligned} \tag{18}$$

$$\begin{aligned} \text{Accuracy(Acc.)} &= \frac{TN + TP}{TN + TP + FN + FP} \\ F - \text{Measure} &= \frac{2 \times R \times P}{(R + P)} \end{aligned} \tag{19}$$

The detection rate is a positive proportion of correctly labelled normal traffic relative to the total number of samples in the collection. In equation (20), the true positive rate equals the detection rate. Accuracy is measured by the percentage of attacks that were correctly labelled compared to the total number of attack occurrences in the network.

Where true positive (TP) is the number of attacks correctly classified as attacks, and false positive (FP) is the number of attacks incorrectly classified in the network [53]. The legitimate nodes in the network traffic are called true negatives (TN) when they are accurately identified as legitimate, and false negatives when they are misidentified as malicious. To evaluate the performance of the proposed system, the localization error variance (LEV) and root mean square error (RMSE) which is computed by equation (20) metrics in our simulation work.

$$\text{RMSE} = \sqrt{\frac{1}{M_c} \sum_{t=1}^{M_c} \| u_i^{est} - u_i^{true} \|_2^2} \tag{20}$$

Where u_i^{true} and u_i^{est} represent the actual and estimated location coordinates of the unknown sensor nodes. Whereas M_c is the number of simulation iterations in the experiment for computing localization accuracy of unknown nodes.

The disruption of the network architecture brought on by mobile nodes is a major problem for localization in WSNs [11]. In order to avoid inaccurate localization caused by unknown nodes leaving the range of beacon nodes, it is crucial to pick the right beacon nodes, which have to be relatively slow. When nodes are highly mobile, it might disrupt the network's equilibrium. How far apart or how close together two nodes are moving is indicated by their relative mobility.

6.2. Result and discussion

The simulation results show that anchor nodes have more neighbors and connectivity than the sensor nodes, as shown in Fig. 10(a). The average connectivity of the network is 61, and the average number of the neighbor nodes to each anchor node is 3 using the regular model, as shown in Fig. 10(b).

Fig. 11(a) depicts the error map that was generated for the sensor nodes that could not be identified. As can be seen in Fig. 11(b), the average amount of localization error when using 60 beacon nodes and 240 unknown sensor nodes is 0.1908, which is considered to be immune. The proposed scheme achieved minimum error compared to Y. Jin et al. [26] with an average localization error of 0.37 using a novel iterative localization technique, transforming the matrix into an optimization approach. This demonstrates that network planning and simulation of WSNs using a secure localization technique effectively detect malicious nodes by computing the position and location of all nodes in the network. This is accomplished by utilizing a secure localization scheme. According to the findings of the simulation, range-free localization approaches are able to effectively compute the position and location of unknown sensor nodes. As a result, the beacon is better equipped to identify the rogue node by computing the position and location of each node, resulting in wireless sensor networks that are both secure and scalable.

Malicious nodes affect nodes' distribution and localization accuracy by creating the wrong position and location of the unknown sensor nodes in wireless sensor networks, as shown in Fig. 12(a). and Fig. 12(b). Malicious nodes mislead the sensor nodes' routing path and information, degrading the network service and the network performance by compromising the beacon node and sending wrong information to the base station as a legitimate node.

The suggested solution utilizes a network of twenty movable anchor nodes, which helps to reduce costs while also greatly improving the accuracy with which malicious nodes in WSNs may be located as in Fig. 13 (a) and (b) to show how the suggested scheme's average localization accuracy is enhanced by a hybrid method that combines the DV-hop technique with additional approaches like RSSI and DE for the beacon nodes and the unknown nodes, respectively.

The Hybrid DE-RSSI-DV-hop localization algorithm is a method for determining the location of a node in a wireless sensor network (WSN). It combines three techniques: Differential Evolution (DE), Received Signal Strength Indication (RSSI), and Distance Vector (DV) hop. The Hybrid DE-RSSI-DV-hop localization algorithm starts with an initial location estimate and iteratively updates the estimate until it converges to a stable solution. The algorithm uses the RSSI and DV hop information to calculate the distance between the nodes and then applies the DE algorithm to optimize the location of the node being localized. One advantage of this algorithm is that it is able to work in environments with obstacles or other sources of signal interference that may affect the accuracy of RSSI-based localization.

By using both RSSI and DV hop information, the algorithm is able to improve the accuracy of the location estimate. Overall, the Hybrid DE-RSSI-DV-hop localization algorithm is a powerful technique for localizing nodes in a wireless sensor network and has many potential applications in fields such as environmental monitoring, healthcare, and security as shown in Fig. 13.

To evaluate how well the proposed localization method works, we compared the localization error for several different hybrid approaches, such as the differential evaluation distance vector hop (DE-DV-Hop) [54], the butterfly optimization algorithm (BOA-DV-Hop), the modified Archimedes optimization algorithm distance vector hop (MAOA-DV-Hop), and the combination of received signal strength indicator with DE-DV-Hop (RSSI-DV-Hop) as shown in Fig. 13 (c) and (d). After only a few rounds, the RSSI-DE-DV-hop localization error curve is nearly horizontal

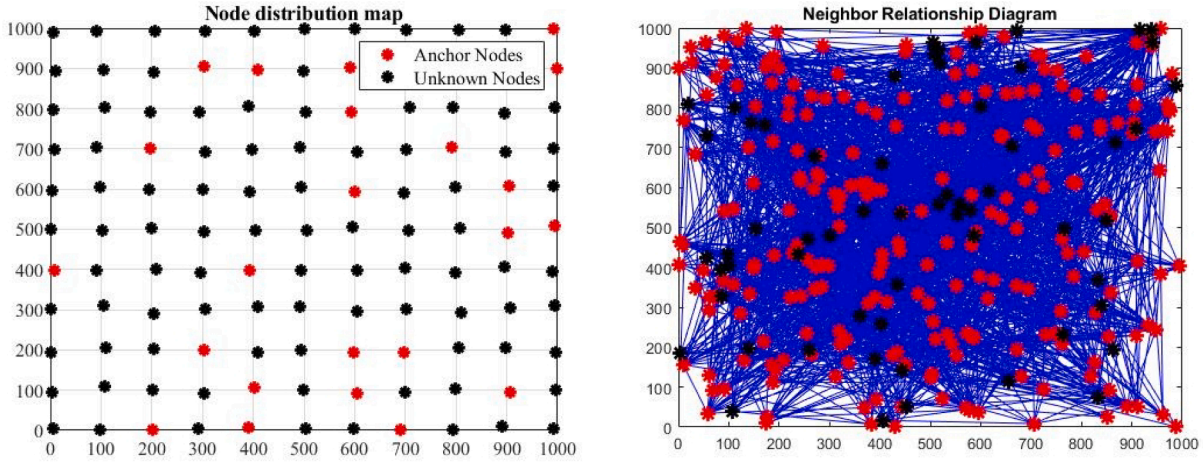


Fig. 10. The localization and computation of unknown nodes' positions in WSNs are depicted in (a) a simulated senor deployment and (b) a neighbor relationship diagram.

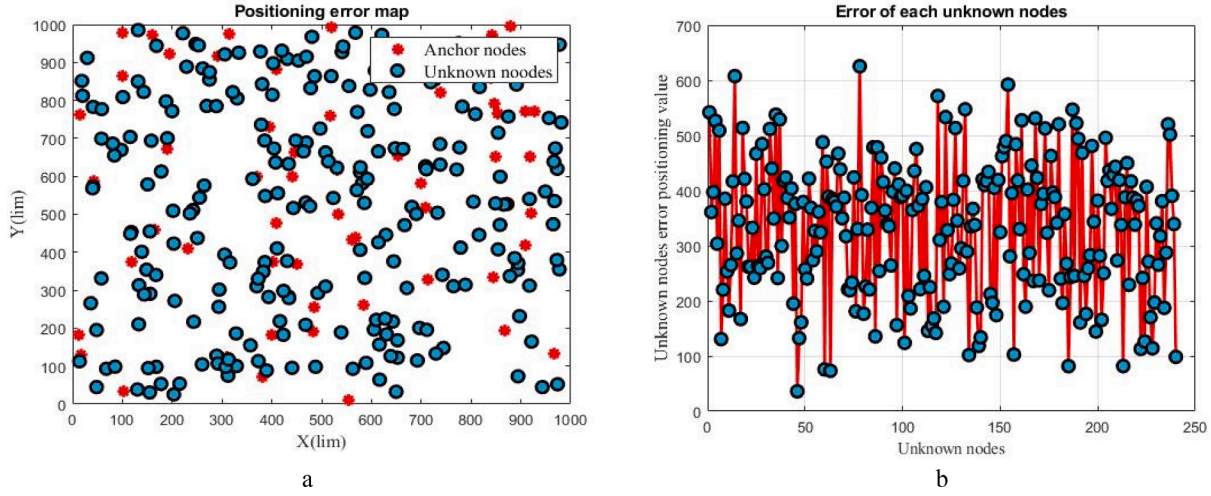


Fig. 11. Computing the error positioning map (a) and error values for each unknown node (b) for random deployment and localization WSNs.

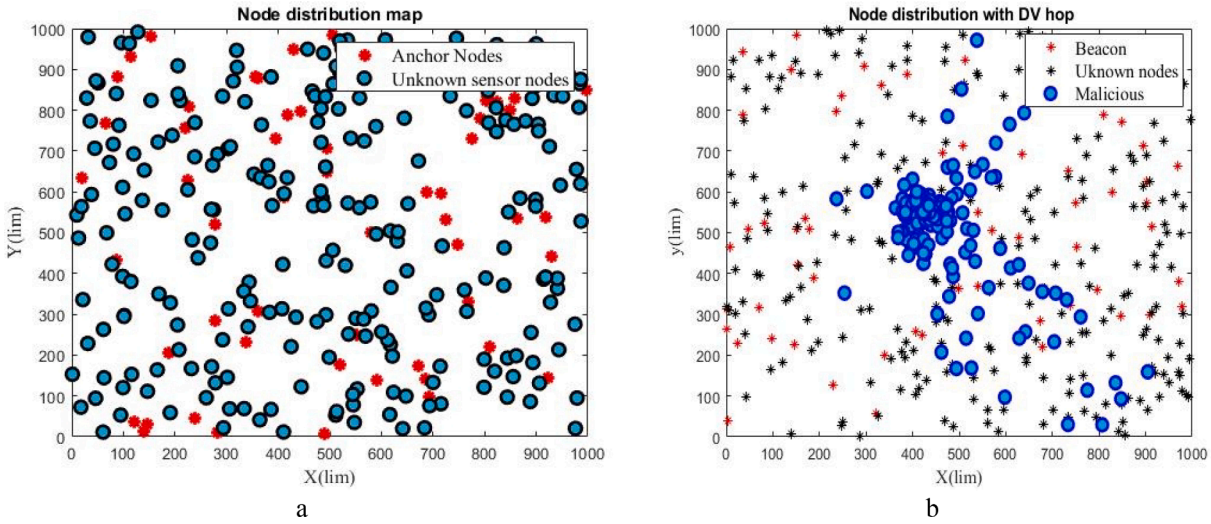
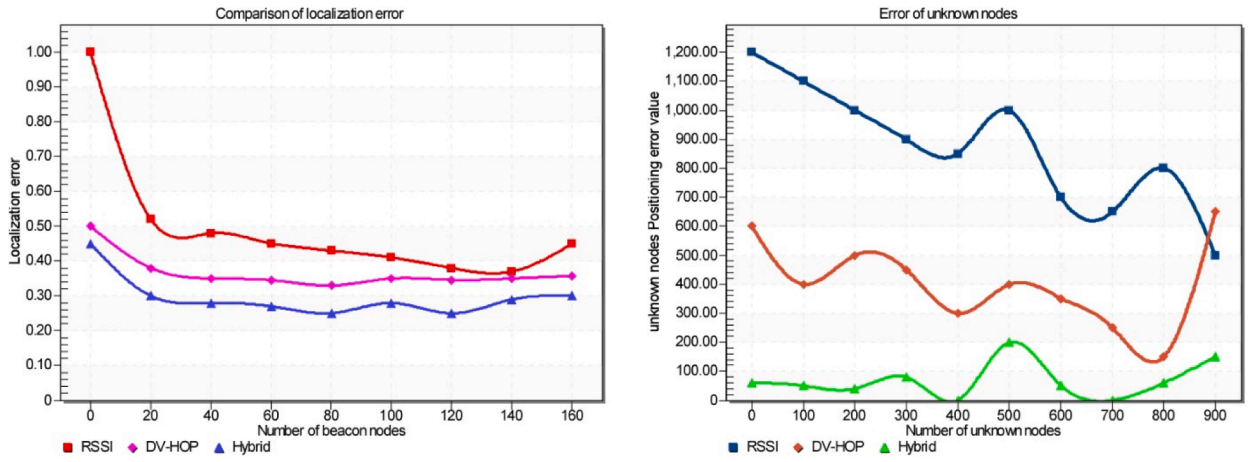
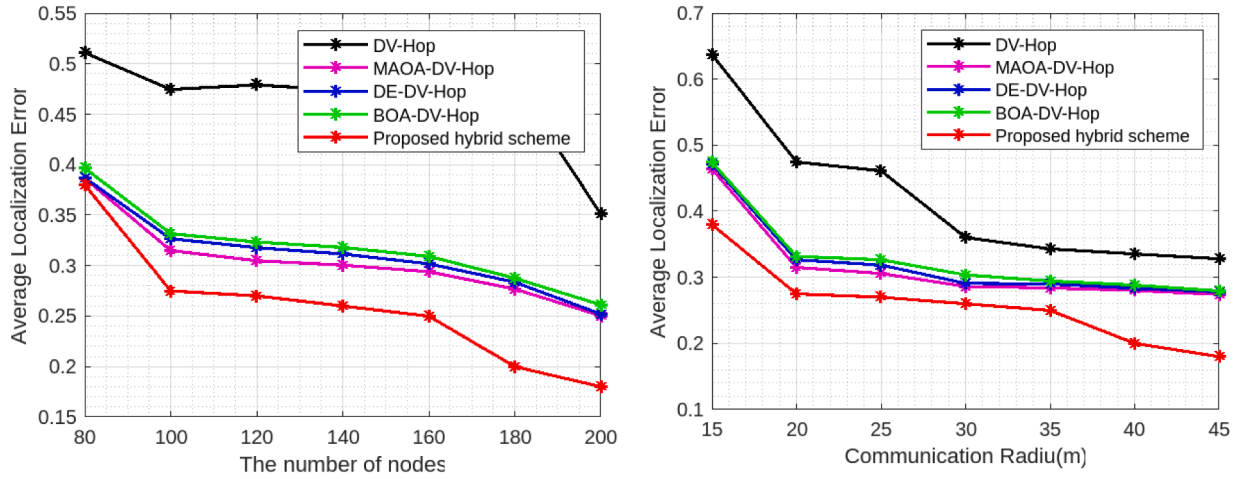


Fig. 12. Node distribution map without an attack(a) and with an attack in wireless sensor networks.



(a) Error analysis of the localization process beacon nodes.

(b) Localization and positioning accuracy unknown nodes.



(c) The ALE with respect to the node density

(d) ALE with respect to communication radius

Fig. 13. Improving the localization and positional accuracy of wireless sensor nodes using a hybrid method.

across all five cases, indicating a high level of global optimization capabilities. Upon reaching the local optimal value, DE-DV-Hop quickly stabilized at 25 iterations. Therefore, RSSI-DE-DV-hop has superior localization performance compared to the other three localization methods, including MAOA-DV-hop. The results demonstrate that the differential evolution method helps the hybrid of range based and range free localization schemes improve localization accuracy and convergence rate.

The hybrid DE-RSSI-DV-hop localization algorithm combines three different techniques, namely Differential Evolution (DE), Received Signal Strength Indicator (RSSI), and Distance Vector (DV)-hop, to achieve more accurate and secure localization in Wireless Sensor Networks (WSNs). The superior performance of this algorithm compared to DV- hop, DE-DV-hop, BOA-DV-hop, and MAOA-DV-hop can be attributed to the following reasons:

DE algorithm: The DE algorithm is used to optimize the location estimation of the malicious nodes. This technique enables the algorithm to converge faster and provides more accurate estimates of the malicious node locations.

RSSI technique: The RSSI technique uses the received signal strength of the nodes to calculate the distance between the nodes. This technique provides more accurate distance estimates, especially in environments with a high level of interference or noise.

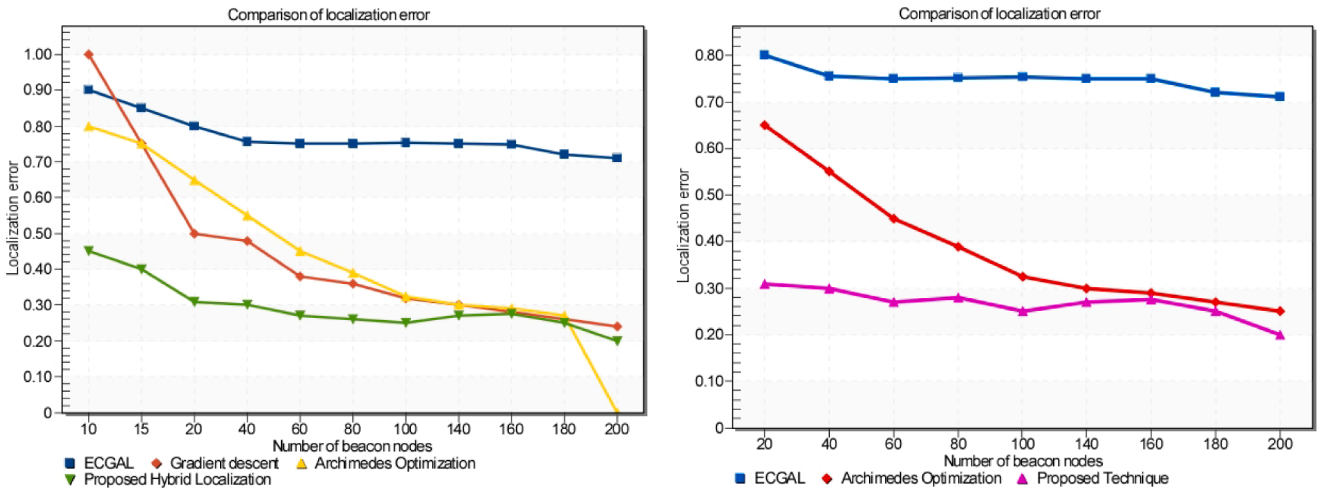
DV-hop technique: The DV-hop technique is used to estimate the distance between two nodes that are not directly connected. This technique is more secure than other techniques because it relies on multiple hops, which makes it difficult for an attacker to intercept or manipulate the messages.

Hybrid approach: By combining these three techniques, the hybrid DE-RSSI-DV-hop algorithm can achieve better accuracy and security in localization. This hybrid approach ensures that the algorithm is more robust and less vulnerable to attacks, making it more suitable for use in WSNs with malicious nodes.

In summary, the hybrid DE-RSSI-DV-hop algorithm has better results than other localization algorithms because it combines multiple techniques to achieve better accuracy and security in WSNs with malicious nodes. The results of the experiments measuring the location error in relation to the number of beacons are shown in Fig. 14 (a) and (b). In addition, when the number of operational sensor nodes increases, the localization error lowers for all algorithms [55]. In terms of localization error, the proposed hybrid approach achieves the best results of any of the strategies we've tried. More reference points are found when there are 200 beacon nodes, bringing the localization error closer to zero.

provides undeniable evidence that the novel approach is superior to standard location-based algorithms in identifying the cause of an error. Almost all of the strategies that have been tried and evaluated have produced positive results when applied to the same setting. The proposed strategy permits a steady decrease since it provides more reference points for the target nodes. Anchor nodes, on the other hand, fortify the network by bringing the unknown nodes closer to the anchors.

As may be shown in Fig. 14. (a) and (b), As more beacon nodes are added, the Average Localization Error (ALE) of four distinct localization



(a) Comparison of localization schemes (b) Comparison with gradient descent

Fig. 14. Evaluation of the suggested strategy by comparing its localization error to that of existing methods by changing the number of nodes in the network.

methods goes down. With more anchor nodes available, we can more accurately estimate the typical hop length. Anchor nodes provide more reliable distance estimates when further nodes are added [54,56]. This demonstrates that as the number of anchors increases, the suggested method becomes more accurate in its estimates of the locations of unknown nodes. Since some of the nodes can be used as anchors for node localization, the proposed method is more accurate than earlier approaches.

6.3. Attack detection analysis

The samples from the reference datasets have been put through training and testing processes [57]. First, we randomly assign each sample to one of two groups: the training set and the test set. Step two involves using the whole training set for both training and testing. Finally, cross-validation was utilized to test how well the proposed model actually worked. Area under the curve, false alarm rate, precision, and classification accuracy are used to evaluate performance. Machine learning models are used to assess the effectiveness of the proposed method using a benchmark dataset that represents a category of assaults on wireless sensor networks. To assess the efficacy of the proposed routing attack localization and detection in wireless sensor networks, the hybrid optimized machine learning also makes use of the same benchmark dataset. Table 5 displays the results of many machine learning algorithms side by side. Cluster labelling (CL) k-means binary classification methods are applied to further enhance the suggested system's performance. Table 5 shows comparative performance of the various hybrid machine learning techniques.

Together, the hyperparameter and Bayesian optimization (BO) techniques and the tree-based Parzen estimation (BO-PTE) are used to boost the performance of hybrid machine learning models for the proposed system. Table 6 displays the results of an analysis of the suggested scheme's effectiveness on the basis of the performance of various machine learning models applied to the UNSW_NB15 benchmark dataset. Using the benchmark dataset, the binary classification method based on hybrid cluster labelling K-means obtains a classification accuracy of 100 %. Table 6 shows how merging different hybrid machine learning models and moving data frames from one machine learning to the other improves the proposed system's performance even further. The results demonstrate that hybrid ML models outperform their standalone counterparts in terms of Validation, Accuracy, Precision, Recall, F1-score and training time. when it comes to classification and detection. Based on the results and model validation, we can say that the created system has high classification and detection accuracy against DoS attacks in WSNs.

The results show that hybrid machine techniques perform better attack detection and classification of attacks using NSL-KDD benchmark dataset as shown in Fig. 15 (a) and (b). For attack detection and classification, the extreme gradient boosting (XGB)-enhanced hybrid of a random forest and a decision tree achieves better results than either the random forest or the decision tree alone in terms of validity, accuracy, precision, recall, and f1-score.

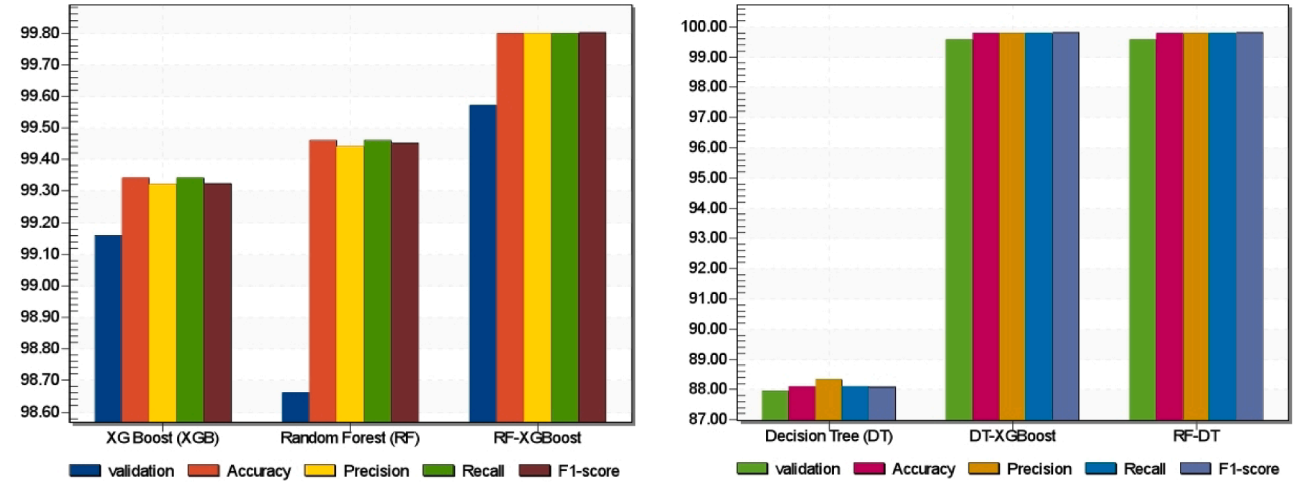
The performance of the proposed technique is effective compared to L. Yang et al. [42] developed multi-tiered hybrid intrusion detection systems (MTH-IDS) for secure vehicular networks using the benchmark dataset CICIDS2017 for known and unknown attacks and achieved average detection accuracy of 99.88 % using binary classification. P. Sun et al. [58] developed a hybrid deep learning-based intrusion detection system (DL-IDS) using a convolutional neural network and long short-term memory network (CNN-LSTM). The scheme achieved an average detection accuracy of 98.67 % by extracting the network traffic's. This proves that the proposed scheme effectively detects DoS attacks using the benchmark dataset in wireless sensor networks, as shown in Fig. 16 (b), using attack detection performance metrics. S. M. Kasongo [40] presented for an intrusion detection system for the internet of things using random forest based on a genetic algorithm (RF-GA) for feature selection as shown in Fig. 16. This achieved average detection accuracy

Table 5
Comparison performance of hybrid machine learning models using the benchmark datasets.

	Results against UNSW_NB15 dataset				Results against CICIDS2017 dataset			
Classifier	Accuracy	Precision	Recall	F1core	Accuracy	Precision	Recall	F1Score
XGB	99.78	99.74	99.78	99.75	99.82	99.86	99.82	99.83
RF	99.75	99.67	99.75	99.70	99.82	99.82	99.82	99.80
DT	99.68	99.63	99.68	99.66	99.82	99.91	99.82	99.85
ET	99.72	99.66	99.72	99.68	99.82	99.80	99.82	99.80
ES	99.78	99.74	99.78	99.75	99.82	99.91	99.82	99.85
CLK-M	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

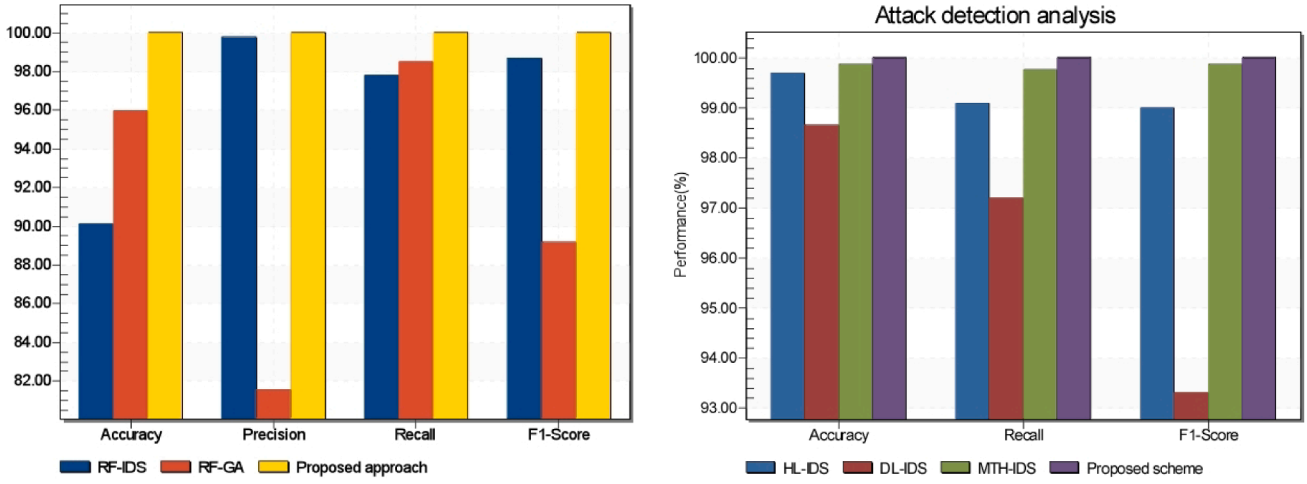
Table 6
Comparison of various hybrid machine learning models using NSL-KDD dataset for attack detection and classification.

	Performance evaluation metrics						
ML Classifiers	validation	Accuracy	Precision	Recall	F1-score	ROC	Running time
NB	83.94	83.71	90.35	83.71	85.68	99.90	0.025
DT	87.94	88.10	88.34	88.10	88.07	99.91	0.22
XGB	99.16	99.34	99.32	99.34	99.32	99.91	1.83
RF	98.66	99.46	99.44	99.46	99.45	99.93	0.17
DT-XGB	99.57	99.80	99.80	99.80	99.80	99.90	1.24
RF-XGB	99.57	99.80	99.80	99.80	99.80	99.85	1.44
RF-DT	99.57	99.79	99.79	99.79	99.79	99.90	0.327



(a) RF based comparative analysis (b) DT based comparative analysis

Fig. 15. Performance comparison of various machine learning models using NSL-KDD.



(a) Comparison based on UNSW_NB15 (b) Comparison based CICIDS2017

Fig. 16. Performance comparison of the proposed scheme based hybrid machine learning techniques using benchmark datasets.

of 87.61 %, which is less than compared to 100 % using hybrid binary classification. M. F. Suleiman and B. Issac [77] Evaluated six machine The size of the NSL-KDD dataset is decreased in the created In order to achieve high performance intrusion detection across a wide range of attack types, B. Media et al. [59] proposed a hybrid-layered IDS (HL-IDS) that employs a system by first performing data preprocessing on the dataset using various feature selection algorithms. Learning classifiers using UNSW_NB15, phishing and NSL-KD benchmark datasets for intrusion detection system. Random forest based intrusion detection system (RF-IDS) produced better detection accuracy using UNSW_NB15. Temporal and spatial features to enhance attack detection and classification. This confirms the proposed technique is effective for detection and localization of attacks as shown in number of distinct machine learning and feature selection approaches as shown in Fig. 16 (a) (b).

G. H. Lai [60] Proposed the detection of wormhole attacks in wireless sensor networks using low power and lossy network (LLN) routing protocol and achieving 100 % accuracy with fixed range and wormhole tunnel points. This confirms the proposed technique is effective for

localizing and detecting routing attacks in wireless sensor networks using a benchmark dataset. Y. Yuan et al. [61] Presented a novel lightweight method for Sybil attack detection in distributed WSNs using the approximate point in triangle (APIT) localization approach. They achieved an average detection rate of 90 %, which is less than the proposed work. D. Upadhyay et al. [48] proposed a framework for intrusion detection systems in smart grids using Gradient boosting feature selection (GBFS-IDS) by applying machine learning classification techniques. The scheme combines feature engineering with machine learning classifiers and achieves better performance as in Fig. 17 (a). This suggests that the proposed method is effective for DoS attacks in wireless sensor networks in various applications.

The proposed work is also practical compared to the MK-ELM model [62], which has an accuracy of 92.10 % using the UNSW-NB15 dataset. Fig. 17(b) shows the detection and localization for the proposed hybrid machine learning approach compared with other works for Sybil attack detection. The comparison performance is using sample experimental data set examined by V.Sujatha and Anita [18] using hybrid fuzzy extreme machines (H-FEM) techniques with an average detection rate of 97 %. Y. Kayode Saheed et al.[41] proposed a machine learning-based intrusion detection system (ML-IDS) for detecting internet of things (IoT) based network attacks. This method employed an ML-supervised algorithm-based IDS for IoT against a benchmark dataset, resulting in an average detection accuracy of 99.99 %. This demonstrates that the proposed technique is useful for detecting and localizing attacks in WSNs. The proposed attack detection and localization scheme achieve 100 % using the same dataset. B. Hasan et al.[9] The detection accuracy of the malicious node was determined by 91.66 % using the packet delivery and energy consumption evaluation metrics using an optimized artificial neural network. The various comparison performances conclude our proposed scheme is effective for the detection localization of attacks in WSNs. S. Karagol and D. Yildiz [63] proposed a novel path planning technique for statistically solving the problem of localization. Using statistical characterization, they deployed wireless sensor nodes with anchor nodes and unknown nodes using various evaluation criteria with a maximum localization error of 1.862. This suggests that the proposed scheme is more effective for the secure localization of unknown nodes with the help of the anchor nodes.

S. Abdulaziz AlRoomi et al. [64] presented a generalized likelihood ratio for finding the position and location of the compromised anchor node in wireless sensor networks for detecting and localizing malicious nodes and achieving a localization error of 0.66.

X. Liu et al. [65] proposed a range of based-secure localization techniques based on density based on spatial clustering for detecting malicious nodes and achieved a detection rate of 100 %. The localization accuracy and detection of the proposed scheme are practical compared to J. Won and E. Bertino [48] using a novel path planning techniqe for detection and localization of aligned beacon position attacks that exploit the beacon nodes inWSNs. B. Mukhopadhyay et al. [66] presented secure localization techniques and achieved a minimum localization error of 0.43 using a secure weighted least square using a reference point. This shows that the proposed system effectively detects and localizes routing attacks in WSNs.

The performance of the proposed system was assessed using network scalability, events, communication range, localization accuracy, communication failure, secure data aggregation ratio, and network load under a variety of simulation scenarios in order to detect and localise hostile nodes.

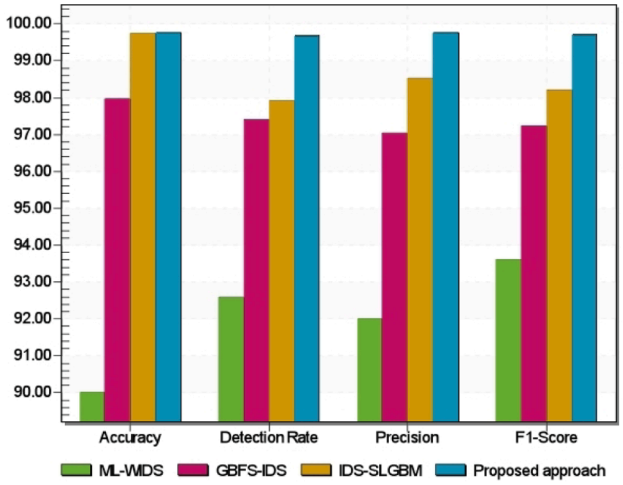
6.4. *Proposed system pros and cons*

Advantages:

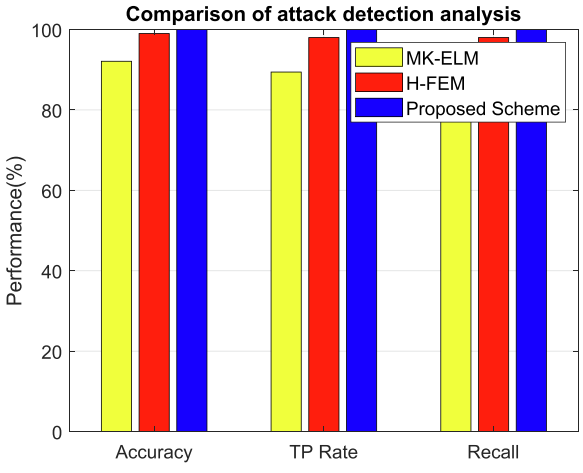
1. Improved security: Secure localization techniques can enhance the security of WSNs by preventing attackers from exploiting vulnerabilities in the routing protocol using machine learning-based attacks. This is because secure localization techniques ensure that nodes can accurately determine their locations, which is crucial for routing protocols to function properly.
2. Reduced energy consumption: Some secure localization techniques can reduce energy consumption in WSNs by minimizing the number of packets that are transmitted to determine node locations. This can help extend the lifetime of the WSN by conserving energy.
3. Scalability: Secure localization techniques can be scalable to large WSNs, as they can be implemented in a distributed manner across the network. This means that they can be easily applied to a large number of nodes, and do not require significant computational resources to operate.

Disadvantages:

1. Complexity: Secure localization techniques can be complex to implement and require significant computational resources to operate. This can make them difficult to implement in resource-constrained WSNs, where nodes may have limited processing power or memory.



(a) Comparative analysis of the proposed scheme



(b) Comparative analysis of the proposed scheme

Fig. 17. Performance comparison of attack detection and localization in WSNs.

2. Cost: Some secure localization techniques may require specialized hardware or software, which can increase the cost of deploying and maintaining the WSN.
3. False positives/negatives: Like any localization technique, secure localization techniques can produce false positives or negatives. This can lead to nodes being incorrectly identified as being in one location when they are actually in another, which can lead to routing errors or other issues.

The goal of secure localization techniques is to offer trustworthy location data to network nodes so they can identify and avoid routing attacks. Wormhole attacks, Sybil attacks, and Sinkhole attacks are some of the most popular types of routing attacks in WSNs. An attacker can launch a wormhole attack if he or she successfully constructs a connection between seemingly unrelated nodes in a network. This has the potential to reroute traffic and interfere with transmissions. A Sybil assault is a form of attack in which the attacker generates a number of dummy identities or nodes within the network in order to exert control over it. For the purposes of eavesdropping or disrupting communication, an attacker may use a sinkhole attack, in which they lure traffic to a certain node or point in the network. In order to protect WSNs from routing assaults, mixed machine learning models might be implemented. To identify and stop intrusion attempts, these models utilize several machine learning techniques.

7. Conclusion and future works

Secure networking planning and data routing for enhancing localization and the lifetime of wireless sensor networks is challenging in an unattended environment. Due to the random nature and deployment of the sensor, wireless sensor networks are vulnerable to routing attacks. Wireless Sensor Networks (WSNs) are prone to various security attacks, including malicious node attacks, due to their distributed nature and limited resources. Trust evaluation and beacon node selection are two important techniques that can be used to detect and prevent malicious nodes in WSNs. To select the beacon nodes, several factors should be considered, including the node's location, energy level, communication range, and its trust score. Nodes with high trust scores and adequate resources are preferred to act as beacon nodes. Furthermore, it is recommended to choose the beacon nodes randomly to prevent attackers from predicting their locations and disrupting the network. Both a range-free trust-based localization algorithm and a range-based secure localization algorithm are successfully implemented for WSNs. In a nutshell, the proposed approach effectively identifies rogue beacon nodes and enhances localization accuracy in conjunction with resistant network topology diversity and localization ratio.

This secure localization technique was designed using multiple stages. The first stage was deploying, clustering, routing selection and data aggregation. In this stage the beacons nodes are selected based on trust evaluation using distance computation. The second stage was data pre-processing using feature selection and feature engineering for training and testing the machine learning model. The last section is decision making for attack detection and classification for report generation. So it is essential to design and apply a secure localization approach in wireless sensor networks and analyze routing attacks using machine learning techniques to overcome the threats of the routing attacks in wireless sensor networks. The average localization error for the unknown nodes is 0.1908, which is immune, which leads to effective detection of malicious nodes. This could be computed with the application of machine learning techniques to analyze and evaluate the performance of the proposed system using the benchmark dataset CICIDS2017 with various classes of attacks. The hybrid optimized machine learning techniques are effective for detection and classification for enhancing the localization accuracy of malicious attacks in wireless sensor networks. The proposed approach achieves average detection accuracy of 100 % using a hybrid random Forest with cluster labeling K-means technique.

There is further work to be done on the localization problem in WSN. Given a wireless network of sensors and distance measurements from each sensor node to its neighbors, the question is how to accurately localize the unknown sensor nodes using the stated approach.

Declarations of Statements.

Funding.

there is no research funding for this work.

Code availability.

No application programs and algorithms are needed.

Authors' Contributions.

Compiling, writing, and conceptualizing are done by Gebrekiros. Prof. J. Panda and Prof. S. Indu did on the discussion, validation and evaluation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y.H. Robinson, S. Vimal, E.G. Julie, K. Lakshmi Narayanan, S. Rho, 3-Dimensional manifold and machine learning based localization algorithm for wireless sensor networks, Wireless Personal Communications 127 (1) (2022) 523–541.
[2] J. Chen, W. Zhang, Z. Liu, R. Wang, S. Zhang, CWDV-Hop: A hybrid localization algorithm with distance-Weight DV-Hop and CSO for wireless sensor networks, IEEE Access 9 (2021) 380–399, https://doi.org/10.1109/ACCESS.2020.3045555.
[3] A. Giri, S. Dutta, S. Neogy, Information-theoretic approach for secure localization against sybil attack in wireless sensor network, J Ambient Intell Human Comput 12 (10) (2021) 9491–9497.
[4] M.M. Singh, N. Dutta, T.R. Singh, U. Nandi, A technique to detect wormhole attack in wireless sensor network using artificial neural network, Springer, Singapore, 2021.
[5] M. Wazid, A.K. Das, An efficient hybrid anomaly detection scheme using K-means clustering for wireless sensor networks, Wireless Personal Communications 90 (4) (2016) 1971–2000, https://doi.org/10.1007/s11277-016-3433-3.
[6] B. Ahmad, W. Jian, Z.A. Ali, S. Tanvir, M.S.A. Khan, Hybrid anomaly detection by using clustering for wireless sensor network, Wireless Personal Communications 106 (4) (2019) 1841–1853, https://doi.org/10.1007/s11277-018-5721-6.
[7] G. Farjamnia, Y. Gasimov, C. Kazimov, Review of the techniques against the wormhole attacks on wireless sensor networks, Wireless Personal Communications 105 (4) (2019) 1561–1584, https://doi.org/10.1007/s11277-019-06160-0.
[8] H. Chen, W. Lou, Z. Wang, J. Wu, Z. Wang, A. Xi, Securing DV-Hop localization against wormhole attacks in wireless sensor networks, Pervasive and Mobile Computing vol. 16, no. PA (2015) 22–35, https://doi.org/10.1016/j.pmcj.2014.01.007.
[9] B. Hasan, S. Alani, M.A. Saad, Secured node detection technique based on artificial neural network for wireless sensor network, Int. J. Electr. Comput. Eng. 11 (1) (2021) 536–544, https://doi.org/10.11591/ijece.v11i1.pp536-544.
[10] G. Farjamnia, Y. Gasimov, and C. Kazimov, "An Improved DV-Hop for Detecting Wormhole Attacks in Wireless Sensor Networks," vol. 9, no. 1, pp. 1–24, 2020.
[11] R. Goyat, G. Kumar, M.K. Rai, R. Saha, R. Thomas, T.H. Kim, Blockchain powered secure range-free localization in wireless sensor networks, Arabian Journal for Science and Engineering 45 (8) (2020) 6139–6155, https://doi.org/10.1007/s13369-020-04493-8.
[12] X. Li, L. Yan, W. Pan, B. Luo, Secure and robust DV-Hop localization based on the vector refinement feedback method for wireless sensor networks, The Computer Journal 60 (6) (2017) 810–821, https://doi.org/10.1093/comjnl/bxx002.
[13] M. Beko, S. Tomic, Toward secure localization in randomly deployed wireless networks, IEEE Internet of Things Journal 8 (24) (2021) 17436–17448, https://doi.org/10.1109/JIOT.2021.3078216.
[14] V.P. Kavitha, J. Katiravan, Localization approach of FLC and ANFIS technique for critical applications in wireless sensor networks, Journal of Ambient Intelligence and Humanized Computing 12 (5) (2021) 4785–4795, https://doi.org/10.1007/s12652-020-01888-1.
[15] S. T. Patel and N. H. Mistry, "A review: Sybil attack detection techniques in WSN," *Proc. 2017 4th Int. Conf. Electron. Commun. Syst. ICECS 2017*, vol. 17, pp. 184–188, 2017, doi: 10.1109/ECS.2017.8067865.
[16] O. Cheikhrouhou, A. Koubaa, "BlockLoc: Secure localization in the internet of things using blockchain", 2019 15th Int Wirel. Commun. Mob. Comput. Conf. IWCMC 2019 (2019) 629–634, https://doi.org/10.1109/IWCMC.2019.8766440.
[17] F.Y. Yavuz, D. Ünal, E. Gül, Deep learning for detection of routing attacks in the internet of things, Int. J. Comput. Intell. Syst. 12 (1) (2018) 39–58, https://doi.org/10.2991/ijcis.2018.25905181.
[18] V. Sujatha, E.A.M. Anita, FEM-hybrid machine learning approach for the detection of sybil attacks in the wireless sensor networks, Int. J. Innov. Technol. Explor. Eng. 8 (7) (2019) 1171–1179.

[19] L. Wang, M.J. Er, S. Zhang, A kernel extreme learning machines algorithm for node localization in wireless sensor networks, IEEE Communications Letters 24 (7) (2020) 1433–1436, https://doi.org/10.1109/LCOMM.2020.2986676.
[20] A. Singh, V. Kotiyal, S. Sharma, J. Nagar, C.C. Lee, A machine learning approach to predict the average localization error with applications to wireless sensor networks, IEEE Access 8 (December) (2020) 208253–208263, https://doi.org/10.1109/ACCESS.2020.3038645.
[21] Q. Cheng, L. Zhang, B.o. Xue, F. Shu, X.u. Wang, A generalized thresholding algorithm with dimension reduction for device-free localization in IoT, Applied Intelligence 53 (8) (2023) 9089–9102.
[22] G. Bhatti, "Machine learning based localization in large-scale wireless sensor, Networks" 18 (12) (2018) 4179.
[23] T.K. Mohanta, D.K. Das, Improved wireless sensor network localization algorithm based on selective opposition class topper optimization, Wireless Personal Communications 128 (4) (2023) 2847–2868, https://doi.org/10.1007/s11277-022-10075-8.
[24] H. Sun, H. Li, Z. Meng, D. Wang, An improvement of DV - hop localization algorithm based on improved adaptive genetic algorithm for wireless sensor, Wireless Personal Communications (2023), https://doi.org/10.1007/s11277-023-10376-6.
[25] E. T. Fute, D. N. Pangop, and E. Tonye, "A new hybrid localization approach in wireless sensor networks based on particle swarm optimization and tabu search," pp. 7546–7561, 2023.
[26] Y. Jin, L. Zhou, L. Zhang, Z. Hu, J. Han, A novel range-free node localization method for wireless sensor networks, IEEE Wirel. Commun. Lett. 11 (4) (2022) 688–692, https://doi.org/10.1109/LWC.2021.3140063.
[27] A. Singh, V. Kotiyal, S. Sharma, J. Nagar, C.C. Lee, A machine learning approach to predict the average localization error with applications to wireless sensor networks, IEEE Access 8 (2020) 208253–208263, https://doi.org/10.1109/ACCESS.2020.3038645.
[28] S. Messous, H. Liouane, L. Bedogni, Online sequential DV-hop localization algorithm for wireless sensor networks, Mobile Information Systems 2020 (2020) 1–14.
[29] S. Dong, X.-G. Zhang, W.-G. Zhou, A security localization algorithm based on DV-hop against sybil attack in wireless sensor networks, Journal of Electrical Engineering and Technology 15 (2) (2020) 919–926.
[30] X. Qi, X. Liu, L. Liu, A combined localization algorithm for wireless sensor networks, Mathematical Problems in Engineering 2018 (2018) 1–10.
[31] P. Li, X. Yu, H.e. Xu, J. Qian, L.u. Dong, H. Nie, Research on secure localization model based on trust valuation in wireless sensor networks, Secur. Commun. Networks 2017 (2017) 1–12.
[32] L. Song, L. Zhao, J. Ye, DV-Hop node location algorithm based on GSO in wireless sensor networks, J. Sensors 2019 (2019) 1–9.
[33] A. Hadir, K. Zine-Dine, M. Bakhouya, J. El Kafi, An improved DV-Hop localization algorithm for wireless sensor networks, Int. Conf. next Gener. Networks Serv. NGNS (2014) 330–334, https://doi.org/10.1109/NGNS.2014.6990273.
[34] G. G. Gebremariam, J. Panada, S. Indu, and M. B. Road, "Localization and Detection of Multiple Attacks in Wireless Sensor Networks Using 1 Introduction".
[35] F. Khelifi, A. Bradai, A. Benslimane, P. Rawat, M. Atri, A survey of localization systems in internet of things, Mob. Networks Appl. 24 (3) (2019) 761–785, https://doi.org/10.1007/s11036-018-1090-3.
[36] O. Cheikhrouhou, G.M. Bhatti, R. Alroobaea, A hybrid DV-hop algorithm using RSSI for localization in large-scale wireless sensor networks, Sensors (switzerland) 18 (5) (2018) 1–14, https://doi.org/10.3390/s18051469.
[37] M. A. Tamtalini, A. E. B. El Alaoui, and A. El Fergougui, "ESLC-WSN: A Novel Energy Efficient Security Aware Localization and Clustering in Wireless Sensor Networks," *2020 1st Int. Conf. Innov. Res. Appl. Sci. Eng. Technol. IRASET 2020*, pp. 0–5, 2020, doi: 10.1109/IRASET48871.2020.9092203.
[38] S. Prithi, S. Sumathi, Automata based hybrid PSO–GWO algorithm for secured energy efficient optimal routing in wireless sensor network, Wireless Personal Communications 117 (2) (2021) 545–559, https://doi.org/10.1007/s11277-020-07882-2.
[39] D. B.d., F. Al-Turjman, A hybrid secure routing and monitoring mechanism in IoT-based wireless sensor networks, Ad Hoc Networks 97 (2020) 102022.
[40] S.M. Kasongo, An advanced intrusion detection system for IIoT Based on GA and tree based algorithms, IEEE Access 9 (2021) 113199–113212, https://doi.org/10.1109/ACCESS.2021.3104113.
[41] Y. Kayode Saheed, A. Idris Abiodun, S. Misra, M. Kristiansen Holone, R. Colomo-Palacios, A machine learning-based intrusion detection for detecting internet of things network attacks, Alexandria Eng. J. 61 (12) (2022) 9395–9409, https://doi.org/10.1016/j.aej.2022.02.063.
[42] L. Yang, A. Moubayed, A. Shami, MTH-IDS: A multitiered hybrid intrusion detection system for internet of vehicles, IEEE Internet of Things Journal 9 (1) (2022) 616–632, https://doi.org/10.1109/JIOT.2021.3084796.
[43] S. Sahu and B. M. Mehtre, "Network intrusion detection system using J48 Decision Tree," *2015 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2015*, pp. 2023–2026, 2015, doi: 10.1109/ICACCI.2015.7275914.
[44] P. Roy, C. Chowdhury, A Survey of machine learning techniques for indoor localization and navigation systems, J. Intell. Robot. Syst. Theory Appl. 101 (3) (2021) pp, https://doi.org/10.1007/s10846-021-01327-z.
[45] G. Kocher, G. Kumar, Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges, Soft Computing 25 (15) (2021) 9731–9763, https://doi.org/10.1007/s00500-021-05893-0.
[46] S.P.K. Gudla, S.K. Bhoi, S.R. Nayak, A. Verma, R. Kaluri, DI-ADS: A deep intelligent distributed denial of service attack detection scheme for fog-based IoT applications, Mathematical Problems in Engineering 2022 (2022) 1–17.
[47] S. Pande, A. Khamparia, D. Gupta, Feature selection and comparison of classification algorithms for wireless sensor networks, Journal of Ambient Intelligence and Humanized Computing 14 (3) (2023) 1977–1989.
[48] D. Upadhyay, J. Manero, M. Zaman, S. Sampalli, Learning Classifiers for Intrusion Detection on Power Grids, IEEE Transactions on Network and Service Management 18 (1) (2021) 1104–1116.
[49] L. Abhishek, "Optical character recognition using ensemble of SVM, MLP and extra trees classifier," *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 2020–2023, 2020, doi: 10.1109/INCET49848.2020.9154050.
[50] S. Taleb, A. Al Sallab, H. Hajj, Z. Dawy, R. Khanna, and A. Keshavamurthy, "Deep learning with ensemble classification method for sensor sampling decisions," *2016 Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2016*, pp. 114–119, 2016, doi: 10.1109/IWCMC.2016.7577043.
[51] M. Sri Vidya and G. R. Sakthidharan, "Accurate Anomaly Detection using various Machine Learning methods for IoT devices in Indoor Environment," *Proc. 5th Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2021*, pp. 308–316, 2021, doi: 10.1109/I-SMAC52330.2021.9640962.
[52] M. Feurer and F. Hutter, "Hyperparameter Optimization," pp. 3–33, 2019, doi: 10.1007/978-3-030-05318-5_1.
[53] M. Nivaashini, P. Thangaraj, Computational intelligence techniques for automatic detection of Wi-Fi attacks in wireless IoT networks, Wirel. Networks 27 (4) (2021) 2761–2784, https://doi.org/10.1007/s11276-021-02594-2.
[54] M. Cheng, T. Qin, J. Yang, C. Riziotis, Node Localization Algorithm Based on Modified Archimedes Optimization Algorithm in Wireless Sensor Networks, J. Sensors 2022 (2022) 1–18.
[55] J. Chen, S.H. Sackey, J.H. Anajemba, X. Zhang, Y. He, A.-B. Hassanien, Energy-Efficient Clustering and Localization Technique Using Genetic Algorithm in Wireless Sensor Networks, Complexity 2021 (2021) 1–12.
[56] Z. Ansari, R. Ghazizadeh, Z. Shokhmzan, "Gradient descent approach to secure localization for underwater wireless sensor networks", *2016 24th Iran*, Conf. Electr. Eng. ICEE 2016 (2016) 103–107, https://doi.org/10.1109/IranianCEE.2016.7585498.
[57] R. Panigrahi, S. Borah, M. Pramanik, A.K. Bhoi, P. Barsocchi, S.R. Nayak, W. Alnumay, Intrusion detection in cyber–physical environment using hybrid Naïve Bayes—Decision table and multi-objective evolutionary feature selection, Computer Communications 188 (2022) 133–144.
[58] P. Sun, P. Liu, Q.i. Li, C. Liu, X. Lu, R. Hao, J. Chen, DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system, Secur. Commun. Networks 2020 (2020) 1–11.
[59] A. Intelligence, S. Science, B. Media, and S. Nature, "A new hybrid approach for intrusion detection using machine learning," pp. 2735–2761, 2019.
[60] G.H. Lai, Detection of wormhole attacks on IPv6 mobility-based wireless sensor network, EURASIP Journal on Wireless Communications and Networking 1 (2016) 2016, https://doi.org/10.1186/s13638-016-0776-0.
[61] Y. Yuan, L. Huo, Z. Wang, D. Hogrefe, Secure APIT Localization Scheme Against Sybil Attacks in Distributed Wireless Sensor Networks, IEEE Access 6 (2018) 27629–27636, https://doi.org/10.1109/ACCESS.2018.2836898.
[62] W. Zhang, D. Han, K.C. Li, F.I. Massetto, Wireless sensor network intrusion detection system based on MK-ELM, Soft Computing 24 (16) (2020) 12361–12374, https://doi.org/10.1007/s00500-020-04678-1.
[63] S. Karagol, D. Yildiz, A Novel Path Planning Model Based on Nested Regular Hexagons for Mobile Anchor-Assisted Localization in Wireless Sensor Networks, Arabian Journal for Science and Engineering 47 (8) (2022) 9833–9848.
[64] S. Abdulaziz AlRoomi, I. Ahmad, T. Dimitriou, Secure localization using hypothesis testing in wireless networks, Ad Hoc Networks 74 (2018) 47–56, https://doi.org/10.1016/j.adhoc.2018.03.008.
[65] X. Liu, S. Su, F. Han, Y. Liu, Z. Pan, A Range-Based Secure Localization Algorithm for Wireless Sensor Networks, IEEE Sensors Journal 19 (2) (2019) 785–796, https://doi.org/10.1109/JSEN.2018.2877306.
[66] B. Mukhopadhyay, S. Srirangarajan, S. Kar, RSS-Based Localization in the Presence of Malicious Nodes in Sensor Networks, IEEE Transactions on Instrumentation and Measurement 70 (2021) 1–16.