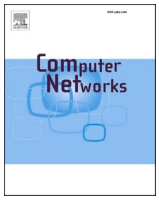
Computer Networks 259 (2025) 111086

Contents lists available at ScienceDirect

Computer Networks

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



Knowledge-based adaptive routing for energy efficiency and attack detection in ad hoc wireless sensor networks

M. Joselin Kavitha [a,*], M.R. Geetha [b], R. Isaac Sajan [b]

[a] Marthandam College of Engineering and Technology, India
[b] Ponjesly College of Engineering, Tamilnadu, India

ARTICLE INFO

Keywords:
Two-level reinforcement learning
route mutation
secure routing

ABSTRACT

Ad hoc wireless sensor networks (WSNs) are susceptible to active attacks due to their dynamic and self-organizing nature. Existing ad hoc WSN secure routing consumes excessive energy and cannot effectively counter active attacks. Thus, an energy-efficient, attack-detecting routing solution is required. This research introduces a novel Knowledge-Based Route Mutation (KBRM) mechanism for real-time security and adaptability in ad hoc WSNs. KBRM employs a two-level reinforcement learning process to provide immediate decision-making for attack detection and defense. It adapts to changing network conditions, reducing energy consumption, and enhancing security. The research presents a unified mathematical model for selecting attack strategies and imposes constraints on energy consumption, security, quality of service, and adaptability. A Predictive Context-Aware Defense mechanism utilizes custom context variables, dynamic accuracy estimation, and predictive threat assessment for improved attack detection. The extended Q-learning algorithm integrates node mobility and the state of neighboring nodes, enabling adaptive route mutation while balancing security and energy efficiency. The research offers a comprehensive approach to enhancing the security and adaptability of ad hoc WSNs. Eventually, comprehensive experimental outcomes demonstrate the efficiency of our approach compared to other solutions.

1. Introduction

Ad hoc wireless sensor networks (WSNs) are widely adopted due to their ability to collect real-time data in demanding environments, such as battlefield surveillance, environmental monitoring, and forest fire detection [1]. However, these networks are vulnerable to various active attacks, which can drain the battery of sensors and disrupt the communication process [2]. The active attacks in which, Node Replication Attacks involve the introduction of duplicate nodes with the same ID, compromising network integrity and facilitating further attacks. Selective Forwarding Attacks occur when compromised nodes selectively drop packets, disrupting data transmission. Data Tampering Attacks involve altering the data being transmitted, which can mislead decision-making processes reliant on accurate information [3–5]. These attacks can compromise the security of the network and drain the battery of sensors, resulting in a loss of data. Therefore, it is necessary to develop an energy-efficient secure routing mechanism that can defend against active attacks in ad hoc WSNs [6]. Existing mechanisms for secure routing in ad hoc WSNs are energy-intensive and cannot effectively defend against active attacks [7–11]. Therefore, there is a need for an energy-efficient secure routing mechanism that can effectively detect and defend against active attacks in ad hoc WSNs.

A major challenge in securing wireless ad hoc networks is their dynamic and rapidly changing nature [12–13]. Existing security methods often rely on static, pre-configured rules and policies, which may not be sufficient to detect and respond to new and evolving security threats [14–15]. Furthermore, some security methods may require significant amounts of computing resources or network bandwidth to operate, which can be a challenge in resource-constrained ad hoc networks [16]. This can result in higher network overheads, which can impact the network's performance and reduce its ability to detect and respond to security threats in a timely manner [17–20]. The proposed energy-efficient secure routing mechanism is designed to counter these specific threats by incorporating dynamic, adaptive responses that evolve with the network.

The key contribution of our work is given as follows:

* Corresponding author.
E-mail address: joselinkavitham@gmail.com (M.J. Kavitha).

https://doi.org/10.1016/j.comnet.2025.111086
Received 22 November 2023; Received in revised form 20 December 2024; Accepted 24 January 2025
Available online 25 January 2025

- Knowledge-Based Route Mutation (KBRM): The proposed KBRM mechanism introduces a dynamic approach to safeguard ad hoc WSNs from active attacks such as selective node isolation attack, wormhole attack, black hole attack, and jamming attack. It combines route mutation with reinforcement learning to enable immediate decision-making and enhance network security.
- Adaptive Learning: KBRM employs a two-level reinforcement learning process that continuously monitors network activity to detect active attacks and dynamically adjusts routing strategies. It enhances the network's ability to defend against similar attacks in the future.
- Constraint Enforcement: Energy consumption, security, quality of service, and dynamic constraints are enforced to enhance network efficiency and adaptability.
- Predictive Context-Aware Defence: The research introduces a Predictive Context-Aware Defence mechanism that utilizes custom context variables, dynamic accuracy estimation, and predictive threat assessment to improve attack detection and mitigation.
- Extended Q-Learning Algorithm: An extended Q-learning algorithm is introduced for route mutation with adaptive node mobility and neighbour node state. This dynamic approach balances security and energy efficiency, ensuring adaptability to changing conditions.
- We undertake comprehensive simulations, and the findings demonstrate that, when compared to state-of-the-art methods, CQ-RM offers notable benefits in a number of areas, including defence, mutation overhead, network, and convergence performance.

2. Related work

Exponentially-Ant Lion Whale Optimisation (E-ALWO) was introduced by Suresh Kumar and Vimala [21] to solve issues with fault tolerance, energy-aware trust-based routing, and intrusion detection. Utilising a fitness function that takes delay and energy into account, the cluster head (CH) is selected based on the node with the shortest distance and the highest energy level. The CH then uses multi-path fading and free-space models to transport data packets with the least amount of energy loss. It also has a system for identifying and isolating black hole nodes by keeping monitoring records on how the nodes handle packet forwarding. It employs a threshold-based strategy to identify and stop the DoS attack by restricting the number of packets that may be delivered by a node in a specific amount of time. However, the limited resources available for computation and communication make it difficult to apply these optimization algorithms in energy-constrained environments like wireless ad hoc network. Due to the limited energy supply, the nodes in the network may not have enough energy to perform the necessary computations for these algorithms, which can lead to reduced network performance and shorter network lifetimes.

Prasad [22] introduced the Enhanced Energy Efficient-Secure Routing (EEE-SR) to enhance energy efficiency and network longevity. The protocol employs energy-efficient management techniques to establish a secure network data connection while determining the data transmission route with minimal energy consumption at the network nodes. The protocol further incorporates a trusted route mechanism to select nodes with higher energy levels at specific times for packet forwarding. Additionally, it limits the quantity of broadcast messages and identifies nodes with a security strategy to prevent attacks such as flooding, worm-hole, and grey-hole attacks.

In ad hoc sensor networks, striking a balance between energy use and security is difficult. It is essential for long-term network functioning to gather data in an energy-efficient manner. To solve these problems, Kumar et al. [23] suggested a Security based Data Aware Routing Protocol (SDARP). It permits extensive data collection while balancing energy and safety measurements. To ensure authentication and data integrity, the protocol combines energy-based decryption and encryption algorithms with fuzzy-based data collection techniques. In accordance with the number of hop distance and nodes, the cluster-based data collection method creates clusters, and the CH is chosen based on node votes. The best CH keeps tabs on cluster members' and its own behaviour. The strategy also decreases route flooding by retaining viable routes for extended periods of time.

Vignesh et al. [24] proposed the adjacency-based Energetic Association Factor Routing Protocol (EAFP) to enhance energy-efficient routing. This protocol introduces a new association parameter derived from the Energetic Association Factor (EAF) equation, which takes into account both the average number of nearby nodes and a node's remaining energy. By utilizing this equation, EAFP calculates an independent association factor for each node, thereby optimizing transmission power and managing associated overheads. This approach helps maintain the most energy-efficient routes, particularly in adaptive scenarios. The EAFP determines route specifications based on the total number of nearby nodes, aligning them with predetermined values. Despite its benefits, the EAFP suffers from high computational overhead, leading to increased energy consumption and a reduced network lifetime.

In order to strengthen network security by defending it against carousal, stretch, and fake data injection attacks, Jasper [25] introduced the Base Station Controlled Secure Routing Protocol (BSCSRP). By selecting a trustworthy and safe routing channel for data transmission, the safety-based trust process offers a straightforward and more robust solution for the network. By evaluating the nodes' dependability and choosing the best path for data transfer, it lessens the impact of attacks. While the indirect trust bases its assessment of the nodes' trustworthiness on the trustworthiness of their adjacent ones, the direct trust bases its assessment of the nodes' trustworthiness on their prior behaviour. The packet drop trust finds nodes that purposefully drop packets, whereas the attribute trust assesses the similarity of attributes across the nodes. The BSCSRP protocol analyses these variables and chooses a trustworthy and safe routing method for data transfer, which lessens threats and improves network security. Additionally, it uses less energy than other existing methods since it protects against attacks like stretch and carousal that force data packets to travel longer routes.

Kumaran and Chinnadurai [26] proposed competent adhoc sensor routing (CASeR) to provides energy efficiency in the network through several mechanisms. Firstly, it employs a reservation-based time division multiple access control (TDMA MAC) protocol that allows nodes to reserve future slots for transmission, minimizing the risk of data transmission collisions. This reduces the energy usage of the nodes and improves the efficiency. Secondly, it uses a cost-based metric for data forwarding, which helps to minimize energy consumption during packet forwarding. This is achieved by selecting the path with the lowest cost, which reduces the number of hops and thus the energy required for packet transmission. Thirdly, it makes use of a flat network topology, which guarantees that each node functions similarly and makes it possible for the protocol to operate simply and be quickly set up on a large number of nodes. This reduces the computational time and energy consumption required for network maintenance. Overall, the combination of these mechanisms in CASeR helps to minimize energy consumption and improve the energy efficiency of the network. However, it does not consider the security of the network, which can impact the energy consumption of nodes. Without adequate security measures, the network may be vulnerable to attacks that can lead to increased energy consumption and reduced network lifetime. The table 1 presents the summary of the related works.

Overall, the challenge lies in developing a routing protocol or optimization algorithm that effectively balances energy efficiency and security while addressing the resource constraints of ad hoc sensor networks. This protocol should optimize energy consumption, enhance network longevity, and mitigate security threats without imposing excessive computational overhead or communication requirements on network nodes.

Table 1
Overview of the reviewed works.

Author/ year	Technique	Routing	Metrics	Active attacks	Limitation	Energy Efficiency	Security	QoS
Suresh Kumar and Vimala [21], 2021	E-ALWO	Energy and trust based	Delay, trust, throughput and residual energy	black hole attack and the DoS attack	Not suitable for energy-constrained environments due to limited resources, which can lead to reduced performance and shorter network lifetimes.	High	Moderate	Moderate
Prasad [22], 2022	EEE-SR	Secure and energy efficient	E2E delay, overheads, MAC collision rate, energy usage, network association and packet delivery ratio (PDR)	Grey-hole, black-hole, worm-hole, spoofing, and flooding	Not suitable for dynamic changing environment.	High	Low	Moderate
Kumar et al. [23], 2020	SDARP	Secure and energy-efficient	Energy efficiency, end to end (E2E) delays, data gathering rate, and lifetime of network.	–	Not able to respond quickly to changes in attack strategies.	High	Moderate	Moderate
Vignesh et al. [24], 2021	EAFP	Energy efficient	Overheads in routing, PDR, average energy usage, average E2E delay, network association, and MAC collision rate	–	High computational overhead that can lead to increased energy consumption and reduced network lifetime.	Moderate	Low	Moderate
Jasper [25], 2021	BSCSRP	Secure and Energy-efficient	Energy usage, throughput, E2E delay, detection accuracy and rate, and False Positive Rate	Carousal, stretch, and false data injection attacks.	Requires a large amount of time and adapt to new attack scenarios.	Moderate	High	Moderate
Kumaran and Chinnadurai [26], 2021	SEEDRP	Energy-based	Throughput, E2E delay, Routing Overhead and PDR.	–	Does not consider network security, which can affect energy usage in the network.	Moderate	Low	Moderate
Proposed KBRM	Extended Q-Learning Algorithm	Route Mutation with Adaptive Node Mobility and Neighbor Node State	Energy efficiency, Security, QoS	Active attacks mitigation through dynamic adaptations	Efficient routing strategy with adaptive node mobility and neighbor state adjustments for improved network efficiency and security	High	High	High

3. System model

Consider a set of sensor nodes, denoted as $X = \{x_1, x_2, ..., x_n\}$ deployed in an ad hoc wireless sensor network (WSN), interconnected through wireless links to form a multi-hop network. These nodes collect environmental data and communicate with neighbors. The network topology T consists of an ad hoc wireless sensor networks, with sensor nodes X arranged in a multi-hop arrangement. This setup enables dynamic, self-organizing communication, facilitating real-time data collection and transmission in hazardous environments. The interference model accounts for potential interference factors affecting data transmission, such as signal strength, channel congestion, and external interference sources, aiming to mitigate these effects to ensure seamless data transmission and network performance. Let I_{ij}represent the interference between nodes i and j. The mobility model dynamically adjusts the mobility patterns of network nodes to optimize performance while minimizing security risks. It uses a random probability parameter P_r to trigger node mobility updates based on threat value T_v and calculates a cost C_{mob} for updating node mobility, considering the trade-off between adapting mobility and maintaining the current network state. By incorporating these mobility adjustments, the model aims to enhance both efficiency and security in ad hoc wireless sensor networks.

3.1. Threat model

In this section, we delineate the potential threats that the proposed system aims to mitigate, providing a comprehensive understanding of the adversarial scenarios in ad hoc WSNs. The threat model identifies the types of attacks, the capabilities of potential adversaries, and illustrative threat scenarios, establishing the context for the defensive mechanisms employed.

Ad hoc WSNs are susceptible to various active attacks that can compromise network security, data integrity, and communication reliability. The primary threats considered in this model include:

a) Selective Node Isolation Attack

In this type of attack, an adversary targets critical sensor nodes that act as routing hubs within the ad hoc wireless sensor network. By isolating these key nodes, the attacker disrupts the network's routing efficiency, leading to communication breakdowns and data transmission inefficiencies. The disruption of critical nodes can fragment the network, causing data loss and message delivery delays, ultimately compromising the network's ability to maintain reliable communication.

Let the set of critical nodes be denoted as$C = \{c_1, c_2, ..., c_n\}$. This attack is tracked by the connectivity status of these nodes by monitoring the number of active communication links $L_i$for each nodec_i. A sudden drop in the number of active links for a critical node $c_i$can be detected as:

$$\Delta L_i = L_i^t - L_i^{t-1} < \theta_{iso} \tag{1}$$

where θ_{iso} is a predefined threshold for detecting a potential isolation attempt.

b) Wormhole Attack

During a wormhole attack, malicious nodes collaborate to create a shortcut or tunnel between distant parts of the network. This tunnel is exploited to redirect and manipulate network traffic, allowing the attacker to intercept and alter data passing through it. Wormhole attacks compromise data integrity and confidentiality, potentially leading to unauthorized access to sensitive information and data tampering.

This attack is detected by analysing packet travel timesT_{travel} between nodes. Abnormally short paths that deviate significantly from the expected travel time can be identified as potential wormhole tunnels:

$$T_{travel} < \theta_{wh} . \overline{T_{travel}} \tag{2}$$

where $\theta_{wh}$is a threshold factor and $\overline{T_{travel}}$ is the average expected travel time between nodes.

c) Black Hole Attack

In a black hole attack, the adversary introduces a malicious node that functions as a black hole. This node absorbs or drops data packets instead of forwarding them, as a legitimate node would. The network nodes, trusting the malicious node for data relay, unknowingly contribute to significant communication disruption. As a result, data packets fail to reach their intended destinations, leading to data loss and impaired network communication.

This attack is detected by monitoring the packet delivery rates $P_{delivery}$for each node. Nodes that consistently exhibit low delivery rates compared to their neighbors can be identified as potential black hole nodes:

$$P^{i}_{delivery} < \theta_{bh}.\overline{P^{N(i)}_{delivery}} \tag{3}$$

where θ_{bh} is a threshold factor, $P^{N(i)}_{delivery}$ is the delivery rate of node i, and $\overline{P^{N(i)}_{delivery}}$ is the average delivery rate of node i's neighbors $N(i)$.

d) Jamming Attack

A jamming attack involves adversaries generating radio interference on the network's frequency, disrupting wireless communication. This interference results in packet loss and degraded network performance. Jamming attacks hinder the network's ability to transmit data reliably, causing increased latency and reduced overall network efficiency. The impact of such attacks includes substantial packet loss and impaired network operations.

This attack is identified by detecting sustained interference patterns and significant packet loss P_{loss} within specific frequency bandsf_{band}.

$$P^{f_{band}}_{loss} > \theta_{jam}.\overline{P^{f_{band}}_{loss}} \tag{4}$$

where $\theta_{jam}$is a threshold factor, $P^{f_{band}}_{loss}$ is the packet loss rate within the specific frequency band, and $\overline{P^{f_{band}}_{loss}}$is the average packet loss rate across all frequency bands.

3.2. Constraints

Here are the four selected constraints with associated formulas:

3.2.1. Energy consumption constraints

These constraints limit energy consumption to exstend the network's lifetime. Each node's energy consumption is monitored, and energy thresholds are defined to ensure optimal energy usage.$\sum (E_i - E_{\min}) \leq E_{\max}$for n nodes in the route, where E_i is the energy level of node i, $E_{\min}$ is the minimum energy threshold, and $E_{\max}$ is the maximum energy threshold.

3.2.2. Security constraints

Security measures, including trusted nodes and encryption, are enforced along the route. A constraint ensures that a certain percentage of nodes in the route are trusted.

$\sum T_i \geq T_{th}$for n nodes in the route, where $T_i$is a binary variable indicating node trustworthiness, and T_{th} is the minimum required number of trusted nodes.

3.2.3. Quality of service (QoS) constraints

Constraints related to QoS parameters, such as data transmission delays, are defined to ensure the total delay in the route does not exceed a specified threshold.$\sum \left(D_i + \sum D_{ij}\right) \leq Q_{th}$for n nodes in the route, where $D_i$is the transmission delay at node i, D_{ij} is the link transmission delay between nodes i and j, and Q_{th} is the maximum allowable delay.

3.2.4. Dynamic constraints

To adapt to changing network conditions, a constraint monitors link quality and avoids nodes with poor link conditions.$\sum L_i \leq L_{\max}$for n nodes in the route, where L_i is a binary variable indicating link condition acceptability, and $L_{\max}$ is the maximum allowable number of nodes with poor link conditions.

3.3. Problem formulation

We start by dividing time into equal slots, each with a length denoted asΔt. These slots are indexed by t, starting from 0 and incrementing by 1. This section describes the route mutation process as a Markov Decision Process (MDP), encompassing essential components such as a set of states, a set of actions, state transition probabilities, and a reward function. Consequently, the objective of finding the optimal mutated route is framed as determining the optimal policy for this MDP.

State (S): The state (S) signifies the current condition of the network at a specific time(t). The state is represented as a tuple of relevant variables describing the network's status. These variables include signal quality (SQ), node density (ND), energy levels (EL), and any other contextual features (CF)vital for effective attack detection and secure routing. For instance:

$$S(t) = (SQ(t), ND(t), EL(t), CF(t)) \tag{5}$$

Action (A): Actions (A)correspond to route selection or route mutation decisions within the network. The action space, denoted as $A(t)$ at time t, involves selecting the next hop node or path for data transmission.

Reward (R): The reward function (R)quantifies the immediate reward received by the agent for taking a specific action in a given state. This function is carefully designed to encourage actions that enhance both network security and energy efficiency. The reward function is formally represented as:

$$R(t) = f(s(t), A(t)) \tag{6}$$

Here, $f()$ is a function that evaluates network performance based on the state and action at time t

State Transition Probability (P): The state transition probability represents the likelihood of transitioning from one state to another when a particular action is taken. Given the inherent unpredictability in the environment, these transition probabilities are stochastically modelled. The specific equations for these transitions depend on the dynamics of the network environment.

4. Proposed energy-efficient secure routing

Introducing a pioneering solution for energy-efficient and secure routing in ad hoc wireless sensor networks (WSNs), the Knowledge-Based Route Mutation (KBRM) mechanism offers a dynamic approach to combat active threats. By leveraging advanced reinforcement learning and predictive context-aware defense, KBRM adapts routing paths in real-time to ensure efficient data transmission and mitigate attacks as shown in Fig. 1. This research explores the architecture, methodology, and components of KBRM, highlighting its potential to revolutionize secure routing in WSNs.

The proposed Knowledge-Based Route Mutation (KBRM) mechanism manages routing decisions, periodically changing paths to avoid predictable patterns and enhance security against active attacks. The reinforcement learning module, divided into high-level and low-level processes, monitors network activity, detects threats, and adapts routing strategies in real-time to ensure secure, efficient data transmission. The adversary detection module continuously analyzes network behavior to identify suspicious activities, signaling the route mutation controller to dynamically alter routing paths and mitigate threats. The predictive context-aware defense mechanism employs custom context variables, dynamic accuracy estimation, and predictive threat

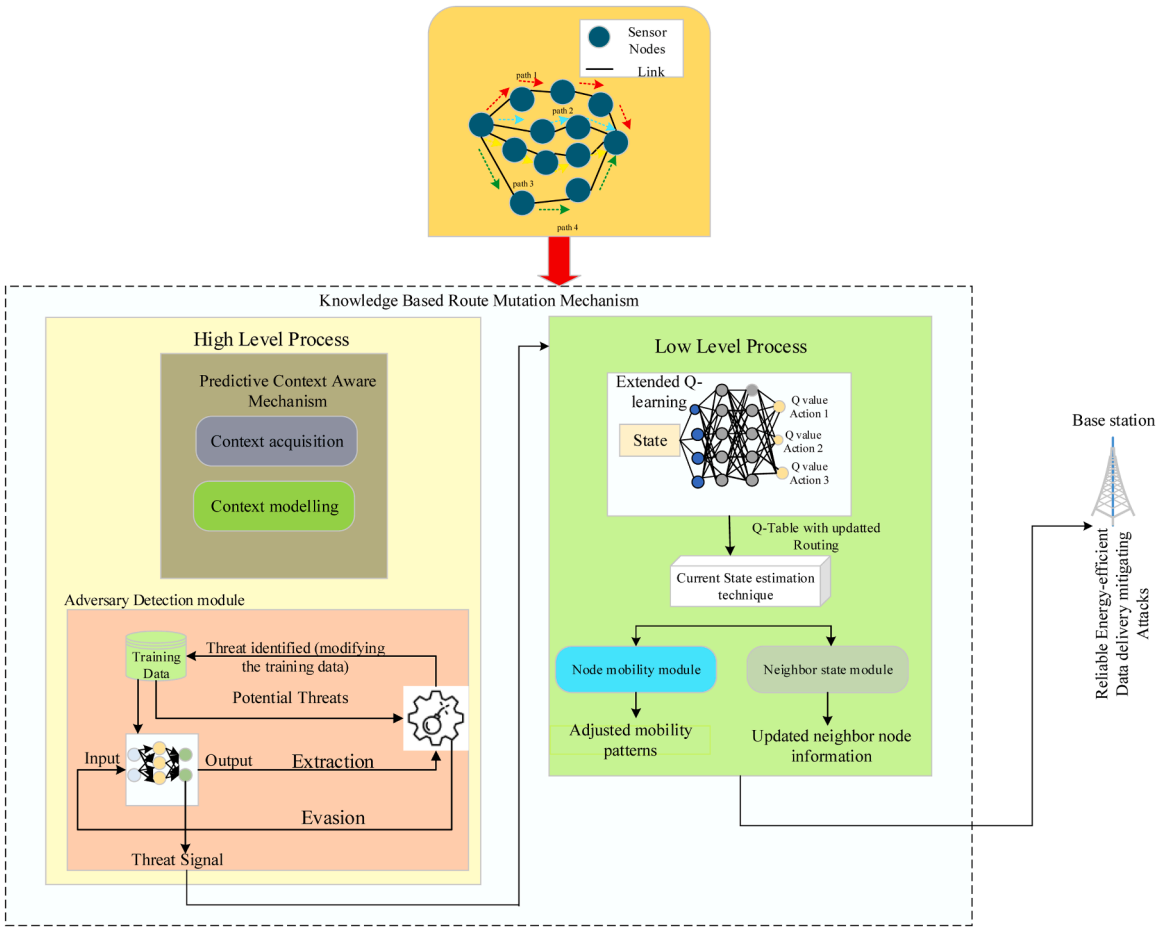


Fig. 1. Architecture of the proposed method.

assessment to enhance attack detection, while the extended Q-learning algorithm integrates node mobility and neighbor node states for adaptive route mutation, balancing security and energy efficiency. Data is aggregated and securely transmitted to the destination node or base station, mitigating routing attacks, selective node isolation attacks, and energy drain attacks, ensuring reliable data delivery, energy efficiency, and enhanced security.

4.1. Initialization phase

Sensor node deployment involves strategically placing sensor nodes throughout the designated area. These nodes have the capability to sense environmental data, denoted asD_e, such as temperature(T), humidity(H), and pressure(P), among others. Each node is denoted asN_i, where i represents the node's unique identifier. Additionally, sensor nodes are equipped with communication modules enabling them to establish connections with nearby nodes, forming a multi-hop network topology. The deployment process can be represented mathematically as$\{N_1, N_2, \ldots\ldots N_n\}$, where $N$is the total number of sensor nodes deployed.

Upon deployment, sensor nodes initialize and start searching for neighboring nodes within their communication range. Let $\mu(N_i)$ represent the set of neighboring nodes detected byN_i. The initial communication setup involves establishing links between neighboring nodes, forming the foundation of the multi-hop network. This process can be mathematically represented as:

Sensor nodes continuously collect environmental data from their surroundings. Let $D_e(N_i)$ denote the environmental data collected by nodeN_i. This data may include parameters such as temperature(T), humidity(H), and pressure(P). After collecting environmental data, sensor nodes utilize simple initial routing protocols to set up basic data transmission paths from nodes to the base station. This involves determining the next hop for each node based on proximity or other metrics.

Let d_{ij} represent the distance between node $i$and node$j$, and $NextHop(i)$denote the next hop for node $i$towards the base station. The basic routing decision can be made based on minimizing the distance to the base station:

$$NextHop(i) = \arg \min_{j \in Neighbors(i)} d_{ij} \tag{7}$$

where $Neighbors(i)$ represents the set of neighboring nodes.

This equation selects the neighbor node that minimizes the distance to the base station as the next hop for forwarding data. Alternatively, other metrics such as signal strength or available bandwidth can be considered for routing decisions.

4.2. Proposed KBRM mechanism

The KBRM mechanism represents a sophisticated approach to enhancing security in ad hocWSNs. Leveraging a multi-hop two-level reinforcement learning process, KBRM integrates route mutation defense technology to enable immediate decision-making aimed at countering attacks and periodically altering routing paths. This strategic combination of reinforcement learning and route mutation ensures the network's resilience to evolving threats while optimizing data transmission efficiency. The Multi-Hop Route Mutation approach forms the backbone of KBRM's routing strategy. This innovative technique allows the mechanism to dynamically navigate through the network by

traversing paths between nodes in an ad hoc manner. Unlike traditional routing protocols that rely on predefined paths or centralized control, the multi-hop route mutation approach adapts to the network's changing topology and environmental conditions.

Formally, the multi-hop route mutation approach can be represented as a sequence of routing decisions$\{r_{1,r}, r_2, \ldots\ldots, r_n\}$, where each r_i represents a routing path between a pair of nodes. The routing path

$r_i$can be expressed as a set of intermediate nodes $\{n_{1,r}, n_2, \ldots\ldots, n_k\}$that the data packet traverses from the source to the destination.

$$r_i = \{n_{1,r} n_2, \ldots\ldots, n_k\} \tag{8}$$

The route mutation process involves dynamically adjusting these routing paths in response to changes in the network topology, node mobility, or detected security threats. This can be represented as a function $f_{route}(.)$that selects the optimal routing path r_j from the set of available paths$R = \{r_{1,r} r_2, \ldots\ldots, r_m\}$:

$$r_j = f_{route}(R, C, T) \tag{9}$$

where $C = \{c_{1,r} c_2, \ldots\ldots, c_k\}$is the set of compromised nodes or regions detected by the adversary detection module. T is the current network topology and node mobility information.

By decentralizing the routing process, KBRM achieves efficient and robust data transmission even in highly dynamic network environments. This is accomplished through a multi-hop approach utilizing a two-level reinforcement learning process, which divides the responsibilities into high-level and low-level processes.

4.2.1. Two-level reinforcement learning module activation

By leveraging reinforcement learning techniques, the module enables the network to autonomously adjust its routing strategies based on real-time feedback and environmental cues. This proactive approach not only enhances the network's ability to detect and respond to potential threats but also lays the groundwork for the subsequent high-level process for attack detection and the low-level process for secure routing.

a) High-Level Process for Attack Detection:

The high-level reinforcement learning process continuously monitors network activity data to detect unusual patterns indicative of attacks. It incorporates the Predictive Threat Assessment mechanism to anticipate potential threats by analyzing trends in context variables and threat values over time. This advanced mechanism includes dynamic accuracy estimation and context representation, allowing for more accurate and proactive attack detection. Based on the observed network behavior and the output of the Predictive Threat Assessment, the high-level process adjusts learning rates $(\alpha)$and mutation periods(β). Mathematically, this can be represented as:

$$\alpha, \beta = f(\textit{Network Activity Data}) \tag{10}$$

where $f$represents the function mapping network activity data to learning rates and mutation periods.

Predictive Context-Aware Defence mechanism

In this section, we present an advanced high-level reinforcement learning process tailored for attack detection within ad hoc WSNs. Our mechanism incorporates custom context variables, dynamic accuracy estimation, and predictive threat assessment. This strategic approach harnesses a contextual framework to assess the reliability of the current network situation, enabling the detection of active attacks with improved accuracy and anticipation. Our context representation encompasses a four-tuple, featuring key elements.

Attack Cost (C_{attack})**:** This factor accounts for the financial impact incurred by potential attacks on the network. It allows for the assessment of the economic implications of security breaches. It can be formulated as:

$$C_{attack} = \sum_{i=1}^{N_a} (\alpha_i . \Delta T_i . L_i . P_{success}) \tag{11}$$

Here, N_a be the number of attack events detected,

α_i denotes the severity weight of the i^{th} attack,

ΔT_i denotes the time duration for the i^{th} attack persists,

L_i Loss incurred per unit time due to the i^{th} attack,

$P_{success}$ Probability of the attack being successful (based on detection rate and attack sophistication).

Attack profit (P_{attack})**:** Attack profit evaluates the potential gain for the attacker when malicious traffic successfully reaches its intended target. It provides a means to quantify the incentive for attackers to compromise the network.

$$P_{attack} = \sum_{i=1}^{N_a} (G_i . \Delta T_i . (1 - D_{detect})) \tag{12}$$

Here, G_i be the gain obtained by the attacker from the i^{th} attack,

ΔT_i denotes the time duration of the i^{th} attack,

D_{detect} be the detection probability (the probability that the system identifies and mitigates the attack).

Defence cost $(C_{defense})$**:** Defence cost is tied to the periodic mutation of routes. By considering this cost, we can gauge the trade-off between security measures and system throughput.

$$C_{defense} = \sum_{j=1}^{N_d} (\beta_j . \Delta T_j . M_j . R_j) \tag{13}$$

Here, N_d Number of defensive measures deployed,

β_j Cost weight for the j^{th} defense mechanism is active,

ΔT_j Time duration for which the j^{th} defense mechanism is active,

M_j Resource consumption rate,

R_j Route mutation frequency (number of route changes per unit time).

Defence profit $(P_{defense})$**:** Defence profit quantifies the advantage held by the defender when it successfully maintains secure routes. It reflects the defender's ability to protect the network. In an effort to adapt our context representation to the unique characteristics of ad hoc WSNs, we introduce custom context variables such as signal quality, node density, and energy levels.

$$P_{defense} = \sum_{k=1}^{N_a} (S_k . (1 - P_{success}) . U_k) \tag{14}$$

Here, S_k be the sensitivity of the system to the k^{th} attack (higher for critical attacks like black holes or wormholes),

$P_{success}$ be the probability of the attack succeeding,

U_k denotes the utility gained by successfully defending against the k^{th} attack.

Signal quality (SQ)**:** This variable takes into account the quality of wireless communication links between nodes. It is pivotal for understanding the strength and reliability of network connections.

$$SQ = \frac{\textit{Received Signal Strength}}{\textit{Required Minimum Signal Strength}} \tag{15}$$

If $SQ > 1$ be the high signal quality and $SQ \leq 1$ be the poor signal quality.

Node Density (ND)**:** The density of nodes in the network can significantly impact network performance. This variable allows us to consider the network's spatial distribution, which can influence the effectiveness of attacks and defences.

$$ND = \frac{N_{active}}{A} \tag{16}$$

Here, N_{active} be the number of active nodes in the region,

A denotes the area of the network region.

Energy Levels (EL)**:** Assessing the remaining energy of nodes is crucial for understanding the network's energy efficiency. This context variable provides insights into energy consumption and resource constraints.

$$EL = \frac{\sum_{i=1}^{N} E_i}{N} \tag{17}$$

Here, E_i be the energy remaining in the i^{th} node,

N denotes the total number of nodes.

These custom context variables enrich the context representation, allowing for a more comprehensive and tailored analysis of network conditions.

Dynamic Accuracy Estimation: Rather than relying on a static accuracy rate(ξ_a), we employ dynamic accuracy estimation to adapt to the specific context. This dynamic approach considers the accuracy rate for each custom context variable $(\xi_a)$and computes a composite accuracy rate. By doing so, we ensure that the accuracy estimation is flexible and can evolve based on the available information. This dynamic accuracy estimation is a critical component for refining the context awareness process. It is determined dynamically using the following formula:

$$\xi_a = \frac{\sum_{i=1}^{n} \xi_a}{n} \tag{18}$$

where ξ_a is the accuracy rate associated with each context variable (e.g., signal quality, node density, energy levels), and$n$is the total number of context variables.

Predictive Threat Assessment: To enhance our Predictive Context-Aware Defence mechanism, we introduce predictive threat assessment. The predictive threat assessment is integrated as follows.

$$K(t_o) = -\lim_{\Delta t \to 0} \frac{\Omega(t_0 + \Delta t) - \Omega(t_0)}{\Delta t} \tag{19}$$

This feature enables us to anticipate potential threats by analysing trends in the context value and threat value over time. The predictive threat assessment is built upon Eq. (3), which calculates the threat value based on the rate of change in the context value. When$K > 0$, it provides an early warning of lower context reliability, suggesting an increased profit for attackers. In contrast, when$K < 0$, it signifies higher context reliability as the defender outperforms the attacker. This predictive capability is invaluable for proactive threat mitigation.

The Predictive Threat Assessment algorithm evaluates the threat level within a system by considering various contextual factors and predictive capabilities. It takes into account inputs such as estimated attack costs, the system's current state represented by a context matrix, defense costs, signal quality, node density, and energy levels. Operating over a series of time slots, it updates the context matrix with environmental factors like defense costs, signal quality, and node density. The algorithm then calculates the dynamic accuracy rate by aggregating individual accuracy rates for each context variable, reflecting the reliability of the contextual information. With this information, it computes the context value, offering an overall assessment of the system's security status considering the reliability of the context. Leveraging predictive threat assessment techniques, it calculates the threat value based on the context value, estimated attack costs, and energy levels, enhancing the algorithm's ability to identify potential threats and enabling proactive mitigation measures.

b) Low-Level Process for Secure Routing:

The low-level process focuses on the real-time routing of data packets, ensuring that they are transmitted through secure and efficient paths. It adapts to changes in network topology and node mobility, maintaining optimal performance and security despite the dynamic nature of WSNs. This process leverages the Q-learning algorithm to dynamically adjust routing paths based on evolving network conditions and security threats.

(i) Q-Learning Algorithm:

The Q-learning algorithm introduces novelty by integrating route mutation with adaptive node mobility and neighbour node state in ad hoc WSNs. Unlike traditional approaches, it dynamically adjusts routing paths based on evolving network conditions and security threats, enhancing network resilience while optimizing energy consumption. By considering node mobility and neighbour behaviour, it enables adaptive decision-making in real-time, improving the network's ability to defend against attacks and prolong its lifetime. This dynamic adaptation to the changing environment of ad hoc WSNs, along with its real-time learning capabilities, sets our algorithm apart from existing approaches, making it innovative and effective for securing wireless sensor networks.

The Q-learning algorithm improve network performance and security by dynamically selecting and mutating routes in response to real-time network conditions. The algorithm consists of the following key steps, each contributing to its learning process:

Initialization: The algorithm begins by defining the state-action pairs that represent the possible conditions of the network and the decisions it can take. The state space (S) includes custom metrics such as signal quality (SQ), node density (ND), and energy levels (EL), while the action space (A) includes options such as route selection and route mutation. A Q-table $(Q(s,a))$ is initialized with small random values, representing the estimated cumulative reward for taking action a in state s. The learning parameters are also set, including: $\alpha = learningrate$, $\gamma = discountfactor$, and $\varepsilon = explorationrate$. Here α controls the weight of newly acquired knowledge over previously stored data. γ determines the balance between immediate and future rewards. ε governs the trade-off between exploring new actions and exploiting learned ones

Threat Detection: The system uses a predictive context-aware defense mechanism to detect potential threats in real time. By analyzing metrics such as SQ, ND, and EL, it calculates a threat value K to assess the likelihood of attacks:

$$K = f(SQ, ND, EL) \tag{20}$$

Here, f is a function capturing abnormal behavior, such as sudden drops in SQ or high ND. Based on K, the algorithm dynamically adjusts the mutation periods (ΔT) to alter routing paths more frequently in high-threat scenarios and mobility thresholds to ensure secure routing actions are taken under varying conditions.

Route Mutation: At each time step t, the algorithm selects a routing action A_t based on the current state S_t. An ε-greedy strategy is employed to balance exploration and exploitation:

$$A_t = \begin{cases} Random\ Action, \ with\ probability\ \varepsilon, \\ \underset{a}{\text{argmax}} Q(S_t, a), with\ probability\ 1-\varepsilon. \end{cases} \tag{21}$$

This approach allows the system to discover new routes while leveraging the learned Q-values to prioritize actions with the highest expected rewards. The selected action is executed, such as initiating a route mutation or selecting an alternative path, and the resulting next state (S_{t+1}) is observed.

Reward Calculation: After executing the action A_t, the system calculates an immediate reward R_t to evaluate the effectiveness of the action. The reward function is designed to encourage actions that enhance network performance and security:

$$R_t = w_1.PDR - w_2.EnergyConsumption - w_3.Latency \tag{22}$$

Here, PDR is the Packet Delivery Ratio (%),

$EnergyConsumption$ reflects the energy used by the nodes during the action,

$Latency$ is the delay incurred in data transmission,

w_1, w_2, w_3 are weights reflecting the relative importance of each metric.

Positive rewards are assigned for actions that increase PDR or conserve energy, while penalties are imposed for high latency, excessive

energy consumption, or failure to mitigate threats.

Q-Table Update: The Q-value for the current state-action pair (S_t, A_t) is updated using the observed reward R_t and the expected maximum future reward from the next state (S_{t+1}). The update rule is:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_t + \gamma \max_{a'} Q(S_{t+1}, a') - Q(S_t, A_t) \right] \quad (23)$$

Here, R_t represents the immediate reward,

$\max_{a'} Q(S_{t+1}, a')$ is the maximum Q-value of the next state S_{t+1}, representing the best possible future reward,

α adjusts the influence of the new information on the existing Q-value.

This iterative update refines the Q-values, enabling the algorithm to improve its routing decisions over time by learning from past actions.

Policy Extraction: After multiple iterations, the Q-table converges to an optimal policy π^*, which maps each state to the action with the highest expected cumulative reward:

$$\pi^*(S) = \arg\max_{a} Q(S, a) \quad (24)$$

This policy is then used to guide the routing decisions, ensuring secure and energy-efficient paths are consistently chosen.

(ii) Extended Q-learning Algorithm for Route Mutation with Node Mobility and Neighbour Node State

In this research, we introduce an extended Q-learning algorithm designed to enhance route mutation with adaptive node mobility and neighbour node state in ad hoc WSNs. The algorithm begins with an initialization phase, configuring the Q-table with zero or small random values, setting discount and exploration factors, and preparing parameters for adaptive node mobility. Two crucial thresholds, $th_{l\_mobility}$ and $th_{h\_mobility}$, alongside mobility intervals and cost coefficients, are defined to facilitate adaptive mobility and neighbour state adjustments. The primary Q-learning loop involves episodes where random initial states are selected, and within each episode, a time slot loop dynamically updates node mobility and neighbour node state based on cost evaluation. The algorithm selects actions using a Q-learning strategy, executes them, observes the outcome reward based on energy efficiency and security objectives, and proceeds to the next state. Once all episodes are complete, the Q-table reflects the learned policy, offering a mutated route selection policy with node mobility and neighbour node state adaptations. This dynamic approach mitigates attacks, prolongs network lifetime, and ensures adaptability in the ever-changing ad hoc WSN environment.

Node mobility module: The Node Mobility Module is designed to dynamically adjust the mobility patterns of network nodes to optimize the network's performance while minimizing security risks. The decision to update node mobility is based on a random probability p, where $p \leq p_{mobility}$. When an update is triggered, the module assesses the current network state, which includes factors like the threat value K, to calculate a cost C_m. This costs a weighted sum of defence cost and attack cost, determined by coefficients $\xi d_{mobility}$ and $\xi d_{mobility}$ respectively. It represents the trade-off between adapting node mobility and maintaining the current state. The cost calculation ensures that the decision to update node mobility considers the potential impact of security threats.

$$C_m = \xi d_{mobility} \times DefenceCost + \xi a_{mobility} \times AttackCost \quad (25)$$

The module then adjusts the adaptive node mobility period N_m based on the calculated cost. The adaptive mobility period is determined as follows.

If the cost C_m is higher than a predefined high threshold $th_{h\_mobility}$, it indicates a significant security threat or the need for rapid adaptation. In this case, the mobility period is shortened to $N_m = Cl_{mobility} \times \tau n$ allow nodes to adapt quickly. Here $Cl_{mobility}$ represents a low constant.

If the cost C_m is below $th_{h\_mobility}$ but above a low threshold $th_{l\_mobility}$, it signifies a moderate threat level. The mobility period is adjusted to $N_m = Ch_{mobility} \times \tau n$ which is longer than the low-threshold case, but not as short as the high-threshold scenario. Here, represents a higher constant.

When the cost C_m is below $th_{l\_mobility}$, it implies a low threat level, and the mobility period remains at the basic duration $N_m = \tau n$. In this case, there is no urgency to update node mobility, and the network maintains its current state to conserve resources.

Neighbour Node State Module: The Neighbour Node State Module focuses on updating the state information of neighbouring nodes, which is crucial for effective decision-making in dynamic networks. Similar to the Node Mobility Module, it initiates updates based on a random probability p. When an update is triggered, the module calculates a cost using the same coefficients. (Eq. (7)). This cost reflects the trade-off between updating neighbour node states and maintaining the current state, considering defence and attack costs. By ensuring that neighbour node information is accurate and up-to-date, this module enhances the network's ability to adapt to changing conditions and make efficient routing decisions. In summary, these modules work in tandem to balance the need for adaptation in response to security threats with the need to maintain network stability and energy efficiency. They provide an adaptive mechanism that takes into account the network's context and threat level to make informed decisions about mobility and neighbour node state updates.

Algorithm 1 outlines the process of Predictive Threat Assessment. Algorithm 2 implements Route Mutation by considering Node Mobility and Neighbour Node State, utilizing the Extended Q-learning Algorithm. This extended Q-learning algorithm establishes a dynamic and secure routing strategy for a network. It utilizes route mutation (changing paths), node mobility (strategic repositioning), and neighbour analysis to adapt to attacks. The algorithm learns from experience (Q-table) and considers factors like network topology, node status, and threats. It balances exploiting known good routes with exploring new ones (exploration factor) to find optimal secure paths. Node mobility can be triggered based on security benefits, costs, and minimum intervals. This approach enhances network resilience against evolving threats.

This novel approach enables the system to defend against active attacks, optimize energy consumption, and maintain network reliability in dynamic environments. By dynamically adapting node mobility and neighbor node state through the extended Q-learning algorithm, the system learns to mitigate attacks, making it resilient against evolving threats while conserving energy. Overall, the Q-learning mechanism provides a powerful tool for enhancing route mutation in ad hoc WSNs, ensuring adaptability and security in challenging network conditions.

5. Experimental evaluation

To evaluate and demonstrate the effectiveness of our proposed method, a series of simulations are conducted comparing the proposed KBRM scheme with two state-of-the-art solutions, namely I-RRM (Integrated Random Route Mutation) [27] and Two-way Multi-path [28]. In experimental setup, Z3 Solver [29], a cutting-edge theorem prover developed by Microsoft Research, capable of handling complex constraints, featuring tens of thousands of constraints and millions of variables, are employed [30]. Our proposed scheme is implemented using Python in conjunction with the Z3 solver.

The table 2 details the configuration and parameters for training and testing a reinforcement learning model to enhance security and efficiency in ad hoc Wireless Sensor Networks (WSNs). The simulation uses 100 nodes deployed randomly within a 1000 m x 1000 m area, each with a 100 m communication range and initial energy of 1000 units. Data packets are 512 bytes, and nodes communicate via dynamically established multi-hop routes. Various attack scenarios are simulated, including black hole attacks (nodes drop packets), selective forwarding attacks (specific packets dropped), sinkhole attacks (nodes disrupt

Algorithm 1
Predictive Threat Assessment.

Input: Estimated attack cost $(bca_{t,1}, ..., bca_{t,n})$
Context matrix $(\varphi_{t,1}, ..., \varphi_{t,n})$
Defence cost (C_{td})
Signal quality (SQ)
Node Density (ND)
Energy Levels(EL)
Output: Threat value (K)
initialize K = 0
for each time slot t from 1 to num time slots
update the context matrix Ω(t) using Eq. (2), considering D, S, and N:
$\Omega(t) = f(D, S, N)$
calculate the dynamic accuracy rate ξabased on $\xi\{a, i\}$ for each context variable
$\xi_i = 0$
for i from 1 to n:
$\xi a = \xi a + \xi\{a, i\}$
compute the context value Ω(t) using Eq. (2) with the dynamic accuracy rate ξa
$\Omega(t) = g(\xi a)$
use the predictive threat assessment in Eq. (3) to calculate the threat value K
$K = h(\Omega(t), C, E)$
rseturn K.
End for

Algorithm 2
Route Mutation with Node Mobility and Neighbour Node State using Extended Q-learning Algorithm.

Input: State space: S (set of all possible states)
Action space: A (set of all possible actions)
Output: Q: Q-table (learned policy for route selection)
initialize Q-table (all entries to zeros or small random values)
for each iteration of episode do
select a random initial state S(t) from S
loop through each time slot t = 1, 2, ..., T:
for time slot t = 1, 2, …, T do
Fm = Fm - 1
High-level Process: Attack Detection and Adaptation
if Fm == 0 then
if Attack Detected(S(t)) then
update defense strategies(S(t), current node mobility, neighbour node state)
Fm = Nm
end if
Low-level Process: Secure Routing with Route Mutation Defence
select an action Af(t) based on Q(S(t), A) and ε (using your Q-learning strategy)
exec Action(Af(t))
calculate Reward(R(t))
S(t + 1) = update State(S(t), Af(t))
Q-learning update:
(Replace with your specific Q-learning update formula considering reward R(t), next state S(t + 1), discount factor γ, etc.)
end for
episode-end update:
update defense strategies based on feedback() (adapt based on episode experience)
end for

Table 2
Summary of hyperparameters.

Parameter	Values
Quantity of network nodes	$N = 100$
Deployment Area	$1000m \times 1000m$
Communication Range	$100m$
Initial Energy per Node	1000 *units*
Data Packet Size	512 bytes
Reinforcement Learning	
State Space	Node energy levels, packet delivery success rate, detection of abnormal behavior
Action Space	Route selection, route mutation, packet forwarding
Reward Function	Positive reward for successful packet delivery, negative reward for detected attacks and packet drops, energy efficiency reward based on conserved energy
Q-Learning Configuration	
Learning Rate (α)	0.1
Discount Factor (γ)	0.9
Exploration Rate $(\in)$	Initially high (0.9) and decays over time

traffic), hello flood attacks (fake hello messages drain resources), and wormhole attacks (creating tunnels to disrupt routing). Reinforcement learning components include a state space monitoring node energy level, packet delivery success rates, and abnormal behavior detection, while the action space involves route selection, route mutation, and packet forwarding. The reward function incentivizes successful packet delivery and energy efficiency, penalizing detected attacks and packet drops. Using Q-learning, the training process involves a learning rate (α) of 0.1, a discount factor (γ) of 0.9, and an exploration rate (ε) starting high at 0.9 and decaying over time. Training consists of 1000 episodes of 1000 steps each, with network resets and random node placement per episode. Q-values are updated based on actions and rewards, with exploration gradually reduced to favor learned behaviors.

For testing, the environment is configured with different topologies (grid, clustered, and random), varying node densities (25, 50, and 100), and different attack intensities (low, medium, and high), ensuring a comprehensive evaluation of the model's performance under diverse conditions. Historical training data showed node energy levels ranging

from 500 to 1000 units, packet delivery success rates between 70 % and 95 %, and attack detection rates between 60 % and 90 %. Rewards ranged from −50 for severe packet drops to +100 for successful mitigations and efficient energy use. Historical testing data indicated that grid topologies with high node densities achieved packet delivery rates around 92 % under low attack intensities, while clustered topologies under high attack intensities saw delivery rates drop to around 75 %. Clustered topologies conserved more energy compared to random topologies, especially under low to medium attack intensities, demonstrating the adaptability and robustness of the proposed methodology in diverse network conditions. As shown in Fig. 2, the simulated network consists of 100 nodes, with the red lines indicating possible mutated paths from Node 0 to Node 50. All simulations are performed on a machine featuring an Intel Core i7–8750H processor running at 2.2 GHz and 16GB of RAM. The performance of the proposed KBRM scheme is evaluated from five key perspectives.

5.1. Performance metrics and equations

To evaluate the performance of the proposed KBRM mechanism in ad hoc wireless sensor networks (WSNs), several performance metrics are crucial. Below are the key metrics along with their corresponding equations:

(i) Energy Consumption

Energy consumption is a critical metric as it directly impacts the operational lifetime and efficiency of the network. Total Energy Consumption (TEC) is the cumulative energy consumed by all active nodes in the network during the simulation. It is calculated as:

$$\text{TEC} = \sum_{i=1}^{N} E_i \tag{26}$$

Here E_i is the energy consumed by node i, and N is the total number of nodes.

Energy consumption by each node (E_i) is further computed as:

$$E_i = T_t.P_t + T_r.\mathrm{P}_r + T_{idle}.P_{idle} + T_{comp}.P_{comp}$$

Here, T_t, T_r, T_{idle}, and T_{comp} be the time durations for transmission, reception, idle mode, and computation tasks, respectively.

P_t, P_r, P_{idle}, and P_{comp} be the power consumption rates for each corresponding activity.

(ii) Network Lifetime

Network lifetime refers to the duration until the first node depletes its energy or the network becomes non-functional. The metric ensures that energy-efficient routing strategies are prioritized to extend the operational period of the network. It is a crucial indicator of the network's sustainability and is expressed as:

$$NL = \min_{i \in N}(T_i) \tag{27}$$

where $T_i$is the time at which node i depletes its energy and becomes non-operational,

N be the total number of nodes in the network.

(iii) Detection Rate

Detection rate measures the ability of the system to identify and mitigate attacks within the network. It quantifies the proportion of successfully detected attacks relative to the total number of attacks simulated or attempted. This metric reflects the robustness of the system's security mechanisms, such as the predictive context-aware defense and reinforcement learning-based route mutation strategies. It is denoted as,

$$DR = \frac{Number\ of\ attacks\ \text{detected}}{Total\ number\ of\ attacks} \tag{28}$$

A high detection rate indicates the effectiveness of the system's security mechanisms, such as the predictive context-aware defense.

(iv) Route Change Frequency (RCF)

The RCF evaluates the adaptability and unpredictability of routing strategies in a network. It measures how frequently the routing paths in the network are dynamically changed to prevent attackers from exploiting static or predictable routes. This metric is especially important in enhancing security, as static routes can become vulnerable to attacks. The RCF is denoted as,

$$RCF = \frac{Number\ of\ route\ changes}{Total\ Time} \tag{29}$$

(v) Quality of Service (QoS)

QoS is a comprehensive metric used to evaluate the overall performance of a network under varying operational conditions. It combines multiple key parameters such as bandwidth, average latency, jitter, and

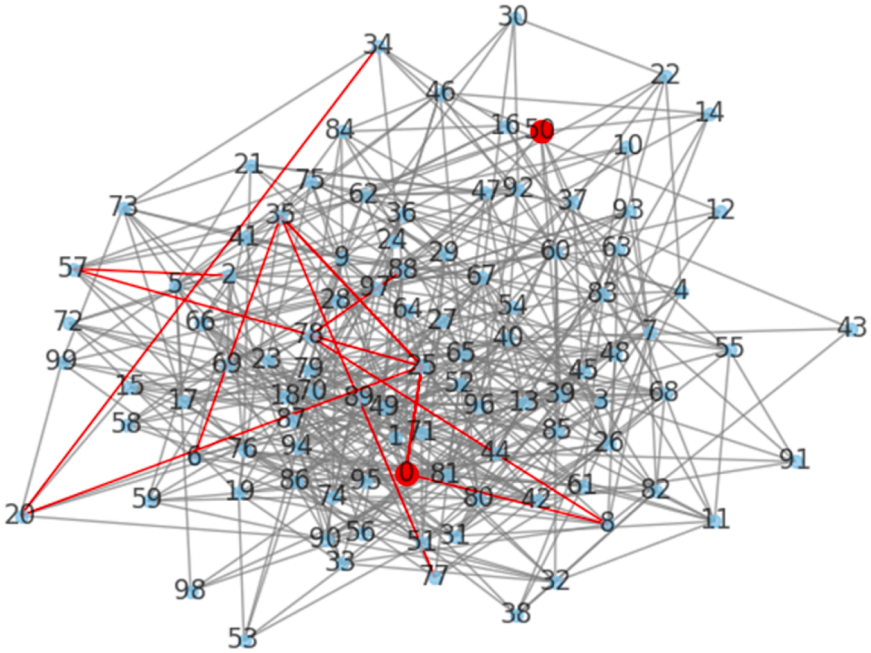


Fig. 2. Network topology with N = 100.

packet loss to provide an aggregated measure of the network's ability to deliver reliable, efficient, and uninterrupted service. In the WSNs, QoS is critical to ensuring the successful transmission of data, especially in applications that demand high reliability and low latency, such as real-time monitoring or security-sensitive communications. It is denoted as,

$$QoS = \omega_i \times PDR - \omega_2 \times AL + \omega_3 \times Bandwidth - \omega_4 \times Jitter \quad (30)$$

Here, *PDR* is the packet delivery ratio (%),
AL be the average latency (ms),
Bandwidth is the available bandwidth (Mbps),
Jitter is the variation in delay (ms),
ω_1, ω_2, ω_3, and ω_4 are weights assigned to each parameter based on their importance.

5.2. Performance comparison of proposed with the existing methods

In this study, we evaluate the performance of the proposed KBRM mechanism against existing state-of-the-art secure routing protocols in ad hoc WSNs. The comparison focuses on key metrics such as Detection Rate, Route Change Frequency, QoS, Energy Consumption, and Network Lifetime. The performance of KBRM is assessed and contrasted with SEECR, E-ALWO, and SDARP using synthetic values to illustrate the advantages of the proposed method.

The table 3 above compares the proposed KBRM mechanism with SEECR, E-ALWO, and SDARP across several key metrics. KBRM demonstrates a superior detection rate of 97.25 %, significantly outpacing SEECR at 88.25 %, E-ALWO at 78.25 %, and SDARP at 75.25 %, highlighting its advanced capability to detect and respond to active attacks in real-time. With a route change frequency of 20 changes per hour, KBRM far exceeds SEECR at 10 changes per hour, E-ALWO at 5 changes per hour, and SDARP at 12 changes per hour, making it more difficult for attackers to predict and target specific routes. KBRM also achieves a QoS score of 90, outperforming SEECR at 80, E-ALWO at 60, and SDARP at 75, demonstrating its ability to maintain reliable and secure data transmission while adapting to network conditions. Furthermore, KBRM's energy consumption is the lowest at 1300 Joules, compared to SEECR at 1500 Joules, E-ALWO at 2000 Joules, and SDARP at 1800 Joules, which contributes to its extended network lifetime of 110 h, surpassing SEECR at 100 h, E-ALWO at 75 h, and SDARP at 80 h. Overall, KBRM's superior performance across these metrics underscores its effectiveness in enhancing the security, efficiency, and adaptability of ad hoc wireless sensor networks.

5.3. Performance comparison of the proposed extended Q-Learning algorithm

In this section, we evaluate the performance of Deep Q-Networks [4] versus the proposed Extended Q-Learning Algorithm across multiple metrics within the context of ad hoc Wireless Sensor Networks (WSNs). The table below summarizes the key performance indicators for both algorithms, highlighting their effectiveness in terms of network resilience, efficiency, and energy consumption.

The comparison table 4 contrasts performance metrics between Deep Q-Networks and the Proposed Extended Q-Learning Algorithm in ad hoc Wireless Sensor Networks. The Extended Q-Learning Algorithm notably reduces dead sensors from 15 to 3 per time unit, enhancing network robustness. However, it shows a higher average hop count of 7.5 compared to 3.23 in Deep Q-Networks, indicating potential issues in routing efficiency. The proposed algorithm also notably decreases network lifetime from 34,560 to 6043 time units, suggesting longevity challenges. Additionally, it exhibits higher energy consumption (8.2 units) than Deep Q-Networks (3.3 units), indicating potential inefficiencies. Yet, it boasts lower routing overhead (80–100 bytes) compared to Deep Q-Networks (100–240 bytes), implying more optimized routing decisions.

Table 4
Performance Comparison of the proposed Extended Q-Learning Algorithm.

Metric	Deep Q-Networks	Proposed Extended Q-Learning Algorithm
Number of Dead Sensors at Time Unit	15	3
Average Hop Count per Time Unit (tu)	3.23	7.5
Network Lifetime	34,560	6043
Energy Consumption (10^{-3})	3.3	8.2
Routing Overhead (Bytes)	100–240	80–100

5.4. Defence performance

One of the key parameters for evaluating defence performance is the attack success rate. In our experiments, we conducted an extensive analysis spanning2.3×10^5time slots, assessing the attack success rate against various attack strategies while deploying three different routing mutation schemes such as Two-way Multipath, I-RRM (Integrated Random Route Mutation), and our proposed method.

As depicted in Fig. 3, the attack success rate in the I-RRM scheme exhibits minimal variation over time slots. The wormhole attack consistently attained the highest success rate, peaking at roughly 26 %. Success rates for other attack strategies hovered around 25 %, 24 %, and 17.7 %, respectively.

Our proposed KBRM mechanism effectively mitigates various attacks, including SNI (Selective Node Isolation), wormhole, and black hole attacks, by dynamically adjusting routing strategies based on

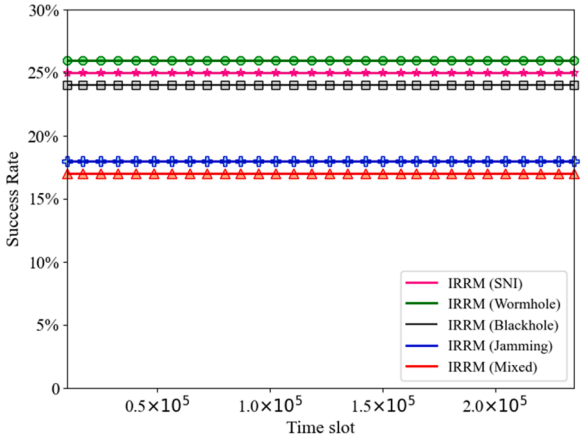


Fig. 3. Comparison of defence performance against attack strategies with I-RRM deployed.

Table 3
Performance of the Proposed Method.

Method	Platform	Detection Rate (%)	Route Change Frequency	Quality of Service (QoS score)	Energy Consumption (Joules)	Network Lifetime (Hours)
SEECR [6]	Linux	88.25	10 changes/hour	80	1500	100
E-ALWO [21]	Windows	78.25	5 changes/hour	60	2000	75
SDARP [23]	MacOS	75.25	12 changes/hour	75	1800	80
Proposed KBRM	Python	97.25	20 changes/hour	90	1300	110

learned attack patterns and network conditions. For instance, KBRM employs anomaly detection techniques such as custom context variables, dynamic accuracy estimation, and predictive threat assessment to identify and reroute traffic away from malicious nodes or compromised paths associated with the SNI attack, resulting in a significant decrease in its success rate from 21 % to 8 %. Similarly, KBRM addresses wormhole attacks by utilizing these anomaly detection techniques, validating routing information integrity, and dynamically adjusting routing paths to avoid potential wormhole tunnels. This comprehensive approach, as demonstrated in Fig.4& 5, leads to a notable decrease in the wormhole attack success rate from 27.6 % to 14.8 % over time slots. Additionally, KBRM counters black hole attacks by leveraging knowledge about network topology and historical attack patterns obtained through these anomaly detection techniques, resulting in decreased success rates from 20 % to 10 % and from 22 % to 5 %, respectively.

The performance evaluation of our defense model, measured through the reduction in attack success rates across various attacks, is obtained through extensive experimentation and analysis in simulated network environments. By subjecting KBRM to diverse attack scenarios, such as mixed attacks comprising jamming, black hole, wormhole, and SNI attacks, we observe significant reductions in attack success rates compared to traditional schemes like I-RRM and Two-way Multipath. For instance, KBRM deployment results in a reduction in success rates of 24 % for mixed attacks, 22 % for jamming attacks, 20 % for black hole attacks, 21 % for wormhole attacks, and 23 % for SNI attacks. These results highlight the effectiveness of KBRM in enhancing network security and resilience across various attack scenarios

Fig. 6 provides a comparative analysis of attack success rates across the three route mutation schemes. KBRM notably reduces the success rate of the attacks by approximately 12 % when compared with I-RRM and around 8 % in comparison to Two-way Multipath. Although the difference in success rates for mixed attack strategies among the three RM schemes is minimal, KBRM remains effective in lowering the success rate of these mixed attacks

5.5. Context-aware analysis

As previously explained, the context value signifies the overall balance of gains and expenses within the continuous interplay of attack and defense dynamics. To gauge the accuracy of Predictive Context-Aware Defence mechanism, we observed the trends in context values when KBRM is deployed against various attack strategies as given in Fig. 7. These trends serve as indicators of the defender's situational awareness. In our findings, context values for different attack strategies initially decrease and then gradually rise. This is primarily due to the declining success rate of attacks, which shifts the balance in favour of the

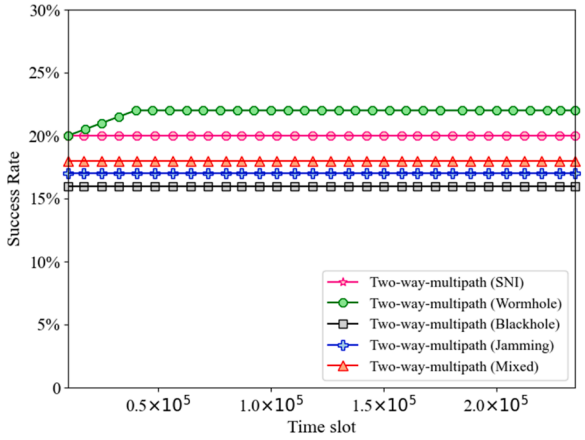


Fig. 4. Comparison of defense performance against attack strategies with Two-way multipath deployment.

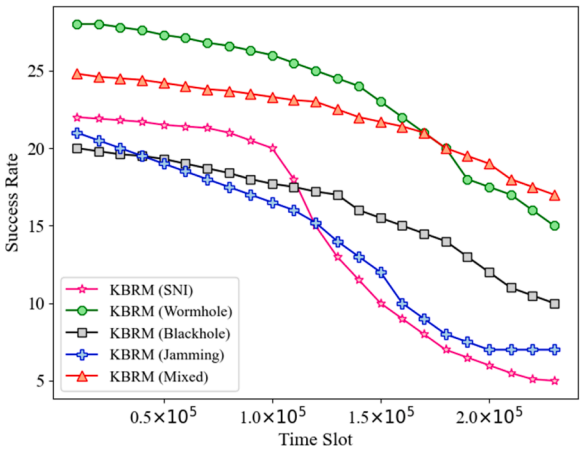


Fig. 5. Comparing defensive performance against attack strategies with the deployment of KBRM.

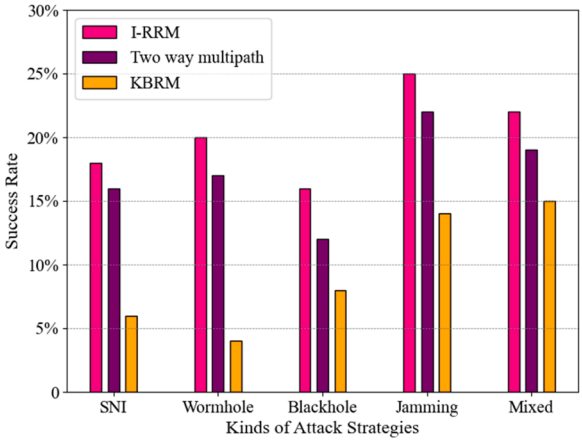


Fig. 6. Comparison of Attack Strategy Success Rates.

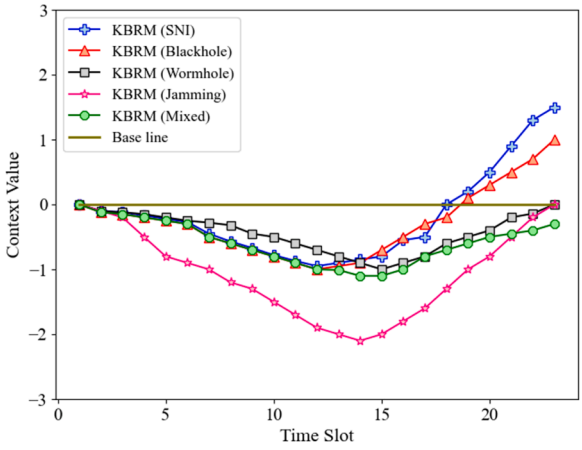


Fig. 7. Context values for five attack strategies.

defender. For instance, context values under SNI attack exhibit a slight initial decrease followed by a rapid increase, showcasing KBRM's effectiveness in guiding routes away from attacks. In contrast, mixed attack context values demonstrate a prolonged and gradual decline, reflecting the challenge of defending against dynamically changing attack strategies.

Comparable patterns are observed in other attack approaches as

well. In conclusion, the analysis of context value trends is vital for determining the threat value, which is inversely proportional to the context value. This real-time threat assessment capability empowers precise evaluations of the network's security situation, ensuring timely and informed decision-making.

5.6. Performance of mutation overhead

This study revealed that Route Mutation (RM) can incur significant costs, particularly in terms of network and management overhead. Specifically, the repeated mutations in each time slot can result in substantial network performance overhead, leading to increased resource consumption. To address this challenge, we introduced a Node mobility module as a key component of our approach, which aims to mitigate this resource consumption during the learning process. As illustrated in Fig. 8, we conducted a comprehensive evaluation by considering different pairs of parameters denoted as$[C_1, C_2]$, where we examined scenarios of [1, 1], [2, 4], and [3, 5]. The [1, 1] scenario represents a constant mutation period throughout the learning process. Our results demonstrated that the node mobility module introduction had a minimal impact on the defence performance of our KBRM scheme. Instead, its influence on the convergence time of KBRM is marginal. This is largely attributed to our PCD mechanism, which ensures that mutations are initiated only in relatively insecure network environments.

Furthermore, we observed a noticeable reduction in the number of mutations, as highlighted in Fig. 9. This reduction signified a substantial decrease in mutation overhead. Specifically, across five distinct attack strategies, we observed varying numbers of non-mutation events, with counts of1.85×10^5,1.93×10^5,1.95×10^5,1.4×10^5, and1.6×10^5, respectively. It's important to note that the most significant reduction in the number of mutations iss observed in response to the SNI attack strategy. In contrast, the jamming attack strategy exhibited the least decrease in mutation events.

5.7. Network performance

Two key metrics used to evaluate network performance in the routing mutation scheme are delay and mutation distance. Delay is crucial for Quality of Service (QoS) and is typically associated with hop count in relatively uniform networks. Fig. 10 shows that delays generally decrease by about 47 %, except when they increase by 32 % under wormhole attack. The reason for this phenomenon is associated with the nature of the wormhole attack itself. Wormhole attacks involve malicious nodes colluding to create shortcuts or tunnels in the network. These shortcuts bypass regular network paths and can lead to an increased number of hops (or longer paths) for data to reach its destination. As a result, the delay metric is adversely affected, leading to higher delays. Overall, KBRM has minimal impact on delay. Fig. 11 reveals that mutation distance gradually decreases by approximately 44 % in KBRM under all attack strategies. The decrease in attack success rates in the RL process results from the elongation of the mutation period, leading to this reduction. Therefore, KBRM reduces network overhead and increases RM feasibility.

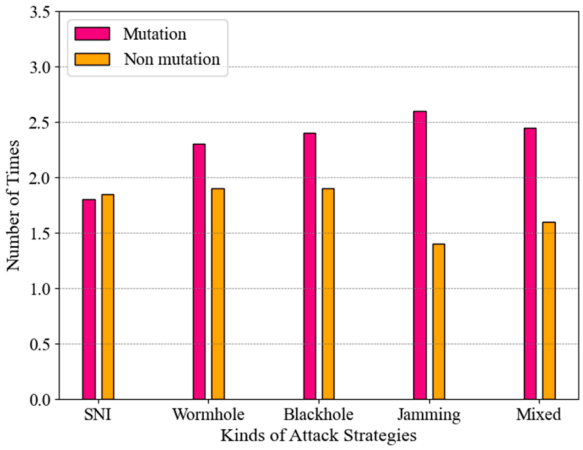


Fig. 9. Mutation performance.

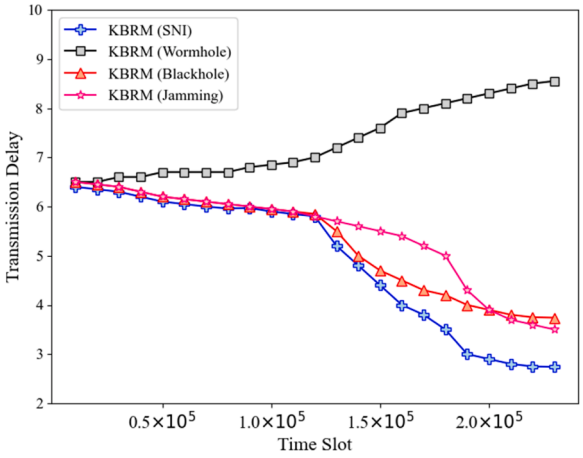


Fig. 10. Transmission delay comparison.

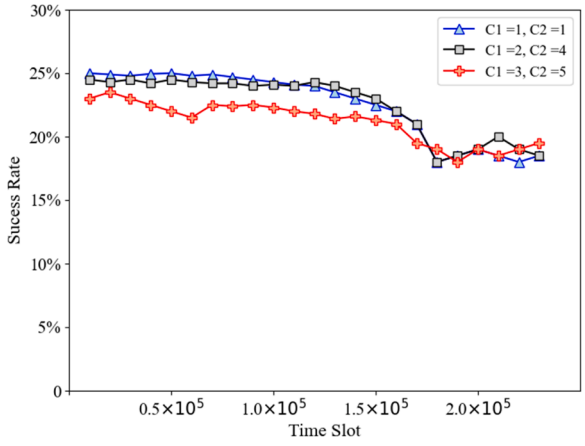


Fig. 8. Performance of defense against mixed attacks with varying mutation periods.

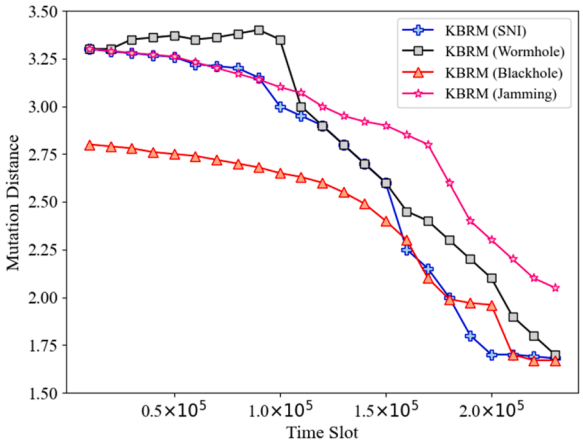


Fig. 11. Mutation distance comparison.

5.8. Convergence performance

In our assessment, we utilize "fitness" to signify the cumulative and progressive value update of an incomplete history. Fitness is represented as $F(\tau)$, where τ denotes the sequence of states and actions produced in the process of reinforcement learning (RL in episode k, and $F(\tau)$ serves as an evaluated measure for this sequence. In practice, fitness can be estimated as $F(\tau) \approx \sum_{t=1}^{T} f_t$, where $|T|$ is the number of time slots in sequence τ, and the calculation of f_t is done.

Contrasting dynamic learning rates with a fixed constant learning rate of 0.9 reveals significant performance differences. As shown in Fig. 12, fitness increases more rapidly at the outset when using dynamic learning rates compared to a constant learning rate of 0.9. However, over 0.74×10^5 time slots, the fitness achieved with the constant learning rate surpasses that attained with dynamic learning rates, with the gap between them progressively widening. Eventually, while the fitness of dynamic learning rates reaches convergence, that of the constant learning rate persists in growing. As the attack success rate converges, dynamic learning rates approach zero, hastening their fitness convergence. This implies that dynamic learning rates have the potential to expedite the convergence process of reinforcement learning (RL).

6. Conclusion

This research offers a comprehensive solution to enhance the security and adaptability of ad hoc Wireless Sensor Networks (WSNs) against active attacks. The Knowledge-Based Route Mutation (KBRM) mechanism combines multi-hop route mutation and reinforcement learning, enabling real-time decision-making for attack detection and defence. The research presents a unified mathematical model, enforces critical constraints, and introduces a Predictive Context-Aware Defence mechanism that harnesses custom context variables and predictive threat assessment. Moreover, the extended Q-learning algorithm enhances route mutation by incorporating adaptive node mobility and neighbour node state. By providing an immediate response to active attacks, adapting to changing network conditions, and reducing energy consumption through route mutation, this research extends the lifetime and efficiency of ad hoc WSNs. The contributions made in this research pave the way for more robust and adaptable security mechanisms in dynamic network environments, ensuring reliable communication and protection against a variety of threats. However, the KBRM mechanism may face limitations in countering specific target attacks, such as sophisticated attacks that exploit vulnerabilities in the learning algorithm or the route mutation process.

Future work can explore advanced machine learning techniques, real-world deployment validation, scalability testing, energy-efficient routing strategies, addressing diverse attack scenarios, and contributing to standardization and integration efforts in WSN security. These future research directions will continue to advance the state-of-the-art in securing ad hoc WSNs and ensuring their long-term viability in dynamic and challenging environments. It is our hope that this work serves as a stepping stone for researchers and practitioners seeking to enhance the security and adaptability of wireless sensor networks.

Funding

There is no funding for this study.

Ethical approval

This article does not contain any studies with human participants and/or animals performed by any of the authors.

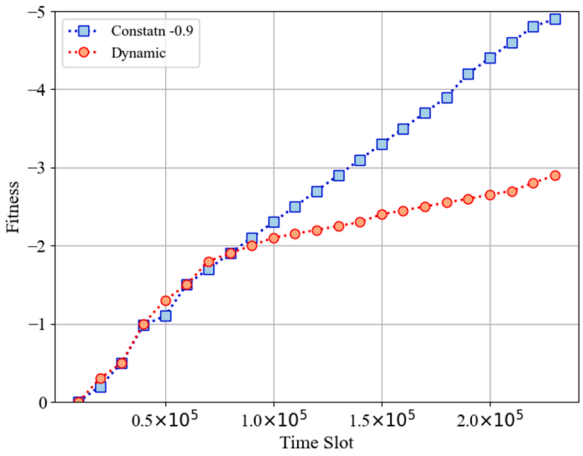


Fig. 12. Comparison of convergence performance.

Informed consent

There is no informed consent for this study.

CRediT authorship contribution statement

M. Joselin Kavitha: Writing – original draft, Visualization, Supervision, Project administration, Formal analysis. **M.R. Geetha:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **R. Isaac Sajan:** Validation, Software, Resources, Methodology, Data curation.

Declaration of competing interest

Authors declares that they have no conflict of interest.

Data availability

Data will be made available on request.

References

[1] D.S. Lakew, U. Sa'ad, N.N. Dao, W. Na, S. Cho, Routing in flying ad hoc networks: a comprehensive survey, IEEE Commun. Survey. Tutorials 22 (2) (2020) 1071–1120.
[2] M. Keerthika, D. Shanmugapriya, Wireless sensor networks: active and passive attacks-vulnerabilities and countermeasures, Global Transit. Proc 2 (2) (2021) 362–367.
[3] H. Hu, Y. Han, M. Yao, X. Song, Trust based secure and energy efficient routing protocol for wireless sensor networks, IEEE Access 10 (2021) 10585–10596.
[4] S. Lata, S. Mehfuz, S. Urooj, Secure and reliable WSN for internet of Things: challenges and enabling technologies, IEEE Access 9 (2021) 161103–161128.
[5] V. Gomathy, N. Padhy, D. Samanta, M. Sivaram, V. Jain, I.S. Amiri, Malicious node detection using heterogeneous cluster based secure routing protocol (HCBS) in wireless adhoc sensor networks, J. Ambient. Intell. Humaniz. Comput 11 (2020) 4995–5001.
[6] K. Saeed, W. Khalil, S. Ahmed, I. Ahmad, M.N. Khattak, SEECR: secure energy efficient and cooperative routing protocol for underwater wireless sensor networks, IEEE Access 8 (2020) 107419–107433.
[7] V. Bhardwaj, N. Kaur, SEEDRP: a secure energy efficient dynamic routing protocol in fanets, Wireless Personal Commun 120 (2) (2021) 1251–1277.
[8] S. Prithi, S. Sumathi, LD, 2FA-PSO: a novel learning dynamic deterministic finite automata with PSO algorithm for secured energy efficient routing in wireless sensor network, Ad Hoc Netw 97 (2020) 102024.
[9] R.I. Sajan, V.B. Christopher, M.J. Kavitha, T.S. Akhila, An energy aware secure three-level weighted trust evaluation and grey wolf optimization based routing in wireless ad hoc sensor network, Wireless Netw 28 (4) (2022) 1439–1455.
[10] R.S. de Sousa, A. Boukerche, A.A. Loureiro, A distributed and low-overhead traffic congestion control protocol for vehicular ad hoc networks, Comput. Commun 159 (2020) 258–270.
[11] Reeya Agrawal, Neetu Faujdar, Carlos Andres Tavera Romero, Oshin Sharma, Ghadia Muttashar Abdulsahib, Osama Ibrahim Khalaf, Romany F. Mansoor, Osama A. Ghoneim, Classification and comparison of ad hoc networks: a review, Egypt. Informa. J 24 (1) (2023) 1–25.

[12] G.Vidhya Lakshmi, P. Vaishnavi, A trusted security approach to detect and isolate routing attacks in mobile ad hoc networks, J. Eng. Res. 12 (3) (2024) 379–386.
[13] Hasanien Ali Talib, Raya Basil Alothman, Mazin S. Mohammed, Malicious attacks modelling: a prevention approach for ad hoc network security, Indones. J. Electr. Eng. Comput. Sci. 30 (3) (2023) 1856–1865.
[14] Joshua Reginald Pullagura, Venkata Rao Dhulipalla, Black-hole attack and counter measure in ad hoc networks using traditional routing optimization, Concurr. Comput. 35 (9) (2023) e7643.
[15] Thabiso N. Khosa, Topside E. Mathonsi, Deon P.Du Plessis, A model to prevent gray hole attack in mobile ad-hoc networks, J. Adv. Inform. Technol. 14 (3) (2023).
[16] Ala Mughaid, Ibrahim Obaidat, Ashraf Aljammal, Shadi AlZu'bi, Fatima Quiam, D. Laila, Aseel Al-zou'bi, Laith Abualigah, Simulation and analysis performance of ad-hoc routing protocols under ddos attack and proposed solution, Int. J. Data Netw. Sci. 7 (2) (2023) 757–764.
[17] Virendra. Dani, Detection of denial-of-service attack using weight based Trust aware routing approach, J. Inform. Assur. Secur. 18 (3) (2023).
[18] Olatayo Moses Olaniyan, Ayobami Taiwo Olusesi, Bolaji Abigail Omodunbi, Wajeed Bolanle Wahab, Olusogo Julius Adetunji, Bamidele Musiliu Olukoya, A data security model for mobile ad hoc network using linear function Mayfly advanced encryption standard, Int. J. Emerg. Technol. Adv. Eng. 13 (3) (2023) 101–120.
[19] Chitra Sabapathy Ranganathan, Rajeshkumar Sampathrajan, Wicked node detection in wireless ad-hoc network by applying supervised learning, Int. J. Electr. Comput. Eng 14 (4) (2024) 2088–8708.
[20] Rasha Hameed Khudhur Al-Rubaye, AYÇA KURNAZ TÜRKBEN, Using artificial intelligence to evaluating detection of cybersecurity threats in ad hoc networks, Babylonian J. Netw. 2024 (2024) 45–56.
[21] K. SureshKumar, P. Vimala, Energy efficient routing protocol using exponentially-ant lion whale optimization algorithm in wireless sensor networks, Comput. Netw 197 (2021) 108250.
[22] R. Prasad, Enhanced energy efficient secure routing protocol for mobile ad-hoc network, Global Transit. Proceed 3 (2) (2022) 412–423.
[23] K.V. Kumar, T. Jayasankar, V. Eswaramoorthy, V. Nivedhitha, SDARP: security based Data Aware Routing Protocol for ad hoc sensor networks, Int. J. Intellig. Netw 1 (2020) 36–42.
[24] C.C. Vignesh, C.B. Sivaparthipan, J.A. Daniel, G. Jeon, M.B. Anand, Adjacent node based energetic association factor routing protocol in wireless sensor networks, Wireless Personal Commun 119 (2021) 3255–3270.
[25] J. Jasper, A secure routing scheme to mitigate attack in wireless adhoc sensor network, Comput. Secur 103 (2021) 102197.
[26] K.M. Kumaran, M. Chinnadurai, A competent ad-hoc sensor routing protocol for energy efficiency in mobile wireless sensor networks, Wireless Personal Commun 116 (1) (2021) 829–844.
[27] Q. Duan, E. Al-Shaer, S. Chatterjee, M. Halappanavar, C. Oehmen, Proactive routing mutation against stealthy Distributed denial of service attacks: metrics, modelling, and analysis, J. Defence Modell. Simul 15 (2) (2018) 219–230.
[28] A. Aseeri, N. Netjinda, R. Hewett, Alleviating eavesdropping attacks in software-defined networking data plane, in: Proceedings of the 12th Annual Conference on Cyber and Information Security Research, 2017, pp. 1–8.
[29] L. De Moura, N. Bjørner, Z3: an efficient SMT solver. In International conference On Tools and Algorithms for the Construction and Analysis of Systems. Berlin, Heidelberg: Springer Berlin Heidelberg. (2008) 337–340.
[30] L. De Moura, N. Bjørner, Satisfiability modulo theories: introduction and applications, Commun. ACM 54 (9) (2011) 69–77.