Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Intelligent Multi-Objective Cluster-Head Selection with AoI-Awareness and Anomaly-Informed Suspicion for Tactical WSNs

Fayza A. Nada
Suez University
Fatma S. Abousaleh
Suez University
Soliman M. Salman
Suez University
Enas Selem
enas-salem@fci.suezuni.edu.eg

Suez University

Research Article

Keywords: Wireless Sensor Networks, Age of Information, Reinforcement Learning, Cluster Head Selection, Anomaly Detection

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10417384/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Intelligent Multi-Objective Cluster-Head Selection with AoI-Awareness and Anomaly-Informed Suspicion for Tactical WSNs

[Fayza A. Nada[1], Fatma S. Abousaleh[2], Soliman M. Salman[1], Enas Selem[1]*]

[[1]Information Technology Department, Faculty of Computers and Information, Suez University, P.O. Box 43221, Suez, Egypt]

[[2]Artificial Intelligent Department, Faculty of Computers and Information, Suez University, P.O. Box 43221, Suez, Egypt]

Corresponding author: [Enas Selem] – [*enas-salem@fci.suezuni.edu.eg*]

Email addresses:

Fayza A. Nada: *f.nada@suezuni.edu.eg*

Fatma S. Abousaleh: *Fatma.abousaleh@fci.suezuni.edu.eg*

Soliman M. Salman: soliman.m.fci18@suezuni.edu.eg

Enas Selem (Corresponding author): *enas-salem@fci.suezuni.edu.eg*

Abstract

Tactical Wireless Sensor Networks (WSNs) must jointly manage limited energy, unreliable links, security attacks, and stale information, yet existing clustering methods usually optimize only one or two of these aspects. This paper proposes a multi-objective cluster-head (CH) selection framework combining K-Means spatial clustering, temporal anomaly detection, and Q-learning guided by a reward weighing residual energy, trust, suspicion, node- and cluster-level Age of Information (AoI), attack status, distance, and CH fairness. The Proposed method was evaluated against LEACH, an RL-based , and a fuzzy trust method (RFGD) across 20 random seeds and nine network configurations (100–300 nodes, 2500–10000 rounds). In the primary 100-node setup, the system achieved the lowest final value for the "Age of Information" (AoI) metric (2.15, compared to 6.99 for LEACH, 18.43 for RL, and 206.78 for RFGD), the highest quality score (0.567), and the lowest energy consumption rate, while maintaining these advantages across all configurations. Results support an AoI-focused, security-aware CH selection framework that improves freshness while remaining energy-efficient and competitive on security metrics.

Keywords: Wireless Sensor Networks; Age of Information; Reinforcement Learning; Cluster Head Selection; Anomaly Detection.

1. Introduction

Wireless Sensor Networks (WSNs) are a promising and essential technology for distributed sensing and monitoring applications, including environmental monitoring, industrial automation, healthcare monitoring, smart agriculture, disaster response, and tactical surveillance [1], [2]. A typical WSN consists of many low-power sensor nodes that collect physical measurements and forward them to a sink or base station. Despite their wide applicability, WSNs remain difficult to operate reliably because sensor nodes are constrained by limited battery energy, restricted computation, unreliable wireless links, and dynamic environmental conditions [1], [2]. These limitations become more critical in tactical WSNs, where the network must support timely and trustworthy situational awareness under harsh operating conditions.

Energy efficiency has therefore been a central research problem in WSN design. Since direct long-distance transmission rapidly depletes battery-powered nodes, clustering is widely used to reduce communication overhead and extend network lifetime. Classical clustering protocols such as LEACH rotate the cluster-head (CH) role probabilistically to distribute energy consumption across the network [3], [4]. Other energy-aware protocols, such as HEED, PEGASIS, TEEN, and APTEEN, improve clustering, routing, or reporting efficiency using residual energy, communication cost, chain-based transmission, or threshold-based reporting [18]-[21]. However, these protocols mainly focus on energy conservation and do not explicitly consider trust, attack behavior, anomaly history, or information freshness during CH selection.

Recent studies have introduced reinforcement learning (RL) into WSN clustering and routing to improve adaptability under dynamic network conditions. RL-based methods can learn CH selection or routing policies based on observed network states and rewards, making them more flexible than static or probabilistic protocols [6], [7], [16]. The multi-faceted RL framework in [16], for example, demonstrates the value of learning-based CH selection and energy-efficient routing in WSNs. Nevertheless, many RL-based WSN approaches mainly optimize energy, topology, delay, or routing quality. They often do not integrate security indicators, trust values, attack-state penalties, or Age of Information (AoI) into one unified CH selection reward.

Security is another essential requirement in tactical WSNs. Sensor nodes may be subjected to attacks such as Denial of Service (DoS), False Data Injection (FDI), Sinkhole attacks, and Selective Forwarding. These attacks can lead to increased traffic load, packet dropping, the injection of unreliable information, or the redirection of data paths toward compromised nodes. If a malicious or suspicious node becomes a CH, the impact is amplified because CHs aggregate and forward data on behalf of multiple nodes. Secure CH selection methods based on fuzzy logic and trust evaluation have therefore been proposed to reduce the probability of selecting compromised nodes [17], [26]. However, fuzzy and trust-based methods often rely on fixed rules and may not adapt to changing network behavior. Moreover, they typically do not optimize AoI explicitly.

At the same time, AoI has emerged as an essential metric for real-time monitoring systems as it measures how up-to-date the information available at the destination is [8, 9]. Unlike latency, which measures how long a packet takes to arrive, AoI determines how old the latest information received is. This is crucial because a Wireless Sensor Network (WSN) may remain technically functional and successfully transmit packets while still providing outdated and operationally useless information. Therefore, freshness-aware routing and clustering must be considered alongside energy efficiency and security.

The main challenge is that energy efficiency, security, and freshness cannot be optimized independently. A node with high residual energy may still be unsuitable as a CH if it has high suspicion or stale information, and a trusted node may still be spatially inefficient or repeatedly overloaded. Existing methods usually address only part of this multi-objective problem: classical clustering focuses on energy, RL-based methods focus on adaptivity and routing efficiency, fuzzy methods focus on trust or security, and AoI-aware methods focus on freshness. Few approaches jointly integrate all these dimensions into one CH selection framework.

To tackle this gap, this paper introduces a reinforcement learning, AoI-aware, and security-aware clustering framework for tactical WSNs. The remaining of the paper is organized as follows: Section II details the contributions; Section III reviews related work; Section IV presents the system model; Section V describes the proposed framework; Section VI covers the simulation setup; Section VII reports results and discussion; Section VIII presents the comparative analysis; Section IX outlines future work; and Section X concludes.

2. Contributions

The main contributions of this paper are as follows:

A unified multi-objective CH selection framework that integrates energy awareness, AoI minimization, trust evaluation, suspicion estimation, attack awareness, and learning-based decision-making.

A K-Means-based spatial clustering mechanism that reduces the CH selection search space and supports energy-efficient intra-cluster communication.

A temporal anomaly detection mechanism that uses short-term histories of traffic, packet drop, AoI, and energy to identify abnormal node behavior.

A Q-learning-based CH selection strategy with a reward function that penalizes stale information, suspicious nodes, attacked nodes, excessive distance from the centroid, and repeated CH selection.

A fair simulation environment comparing the proposed method with LEACH, an RL-based clustering baseline, and RFGD under identical network and attack conditions.

A detailed results discussion explaining not only the numerical outcomes, but also the technical reasons behind each observed performance trend.

In addition to the algorithmic contribution, the work also contributes an integrated experimental pipeline. The simulation was not designed as an isolated test for the proposed method only; instead, all compared methods were executed inside the same environment, exposed to the same attack probabilities, evaluated using the same security evaluator, and saved through the same output pipeline. This is important for publication-quality comparison because it prevents improvements from being caused by different assumptions, different attack models, or different evaluation criteria. The final contribution is therefore both methodological and experimental: the proposed algorithm is introduced, and a controlled environment is established to justify its performance against representative baselines.

The proposed framework is particularly suitable for tactical WSN scenarios because it does not treat communication, security, and information freshness as independent problems. In tactical monitoring, a CH that has high residual energy but low trust may compromise the whole cluster. Similarly, a secure CH that produces stale information may be useless for real-time decision-making. The proposed method addresses this by selecting CHs through a joint decision model rather than sequentially optimizing separate objectives.

3. Related Work

Wireless Sensor Networks (WSNs) have been widely used because of their importance in monitoring, surveillance, industrial sensing, and tactical applications. Early surveys focused on the fundamental constraints of WSNs, including restricted node energy, unreliable wireless links, scalability limitations, and the need for self organization [1] [2]. Since then, a large body of research has focused on clustering, routing, energy balancing, security, and data freshness, but most existing studies optimize only one or two objectives at a time, while tactical WSNs require simultaneous energy efficiency, secure operation, reliable communication, and fresh information delivery.

The literature related to this work can be organized into five main directions: classical energy-aware clustering, reinforcement learning-based CH selection and routing, fuzzy and trust-based secure clustering, AoI-aware information freshness optimization, and edge/digital-twin-assisted WSN intelligence.

3.1. Classical Energy-Aware Clustering and Routing

Classical clustering protocols were mainly designed to extend network lifetime by reducing energy consumption. LEACH is one of the most influential protocols in this category [3], [4]. It randomly rotates the CH role among sensor nodes using a probabilistic threshold, allowing energy consumption to be distributed across the network. LEACH is simple and efficient, but it does not consider node trust, attack behavior, information freshness, or anomaly history. Therefore, it can select weak, suspicious, or information-stale nodes as CHs.

HEED was introduced as an energy-efficient distributed clustering approach that considers residual energy and communication cost during CH selection [18]. This improves clustering stability over LEACH's random rotation, but HEED still lacks explicit mechanisms for security-awareness or AoI minimization. PEGASIS proposed chain-based communication to reduce energy consumption by allowing nodes to transmit to close neighbors and eventually to a leader node [19]. Although this can reduce energy cost, chain-based routing may increase delay and is not designed for tactical environments.

Other hierarchical protocols such as TEEN and APTEEN introduced threshold-based reporting and hybrid periodic/event-driven data collection [20], [21]. These reduce unnecessary transmissions but are not designed for malicious behavior or freshness-aware CH selection. Later metaheuristic and swarm-intelligence methods improved CH placement and routing efficiency [22], [23]. These approaches improve energy balance and communication cost, but many of them still treat security and AoI as secondary or absent objectives.

The limitation of this group is that energy is usually treated as the dominant design metric. While this improves lifetime, it does not guarantee that the selected CH is trustworthy, attack-free, or able to deliver fresh information. The proposed work addresses this limitation by keeping energy as a major objective while integrating trust, suspicion, AoI, attack status, centroid distance, and CH reuse into the decision process.

3.2. Reinforcement Learning-Based CH Selection and Routing

Reinforcement learning (RL) has recently gained attention in WSNs because it enables adaptive decision-making under dynamic network conditions. Q-learning provides a model-free approach in which an agent learns action values through

interaction with the environment [6], [7]. In WSN clustering, RL can be used to select CHs, assign members to clusters, or choose energy-efficient routing paths.

The first key baseline used in this work is the multi-faceted reinforcement learning framework proposed by Mouli et al. [16]. That work uses reinforcement learning for dynamic CH selection, clustering, and energy-efficient routing, showing that RL can improve WSN adaptivity beyond static protocols. However, the framework is primarily oriented toward energy efficiency and routing performance. It does not explicitly integrate AoI minimization, temporal anomaly scoring, attack-status penalties, or cluster-level freshness into the CH reward in the same way as the proposed method.

Recent works have also explored game-theoretic CH selection and multi-agent reinforcement learning routing for energy-efficient WSNs [24], [25]. These studies improve load balancing and reduce premature energy depletion; broader ML surveys confirm that learning-based approaches support routing, clustering, prediction, and anomaly detection [31]. Such approaches are powerful, but they may increase computational complexity and often focus on energy, delay, or path quality without fully integrating trust, suspicion, attack behavior, and AoI.

The limitation of this group is that learning is often used to optimize energy or routing quality, but the learned reward may not capture the security and freshness dimensions required in tactical WSNs. The proposed method extends this direction by designing an RL reward that explicitly combines energy, trust, suspicion, AoI, cluster AoI, attack status, centroid distance, and CH reuse count.

3.3. Fuzzy, Trust-Based, and Secure CH Selection

Security-aware clustering attempts to prevent malicious or unreliable nodes from becoming CHs. Trust-based and fuzzy logic methods are common in this direction because they can combine uncertain indicators such as packet delivery, residual energy, distance, and node reliability.

The second key baseline used in this work is the secure CH selection method proposed by Dhipa et al. [17]. That work combines fuzzy trust/chance evaluation with generic detection and CH reselection, making it more security-aware than classical energy-only clustering protocols. This paper is especially important for the present work because it motivated the RFGD baseline implementation. However, while RFGD-like methods can improve trust-oriented CH selection, they rely on fixed fuzzy rules and do not explicitly optimize AoI or learn adaptive reward policies from network behavior.

Other secure clustering methods use fuzzy trust evaluation and outlier detection to isolate malicious nodes in industrial WSNs [26]. These approaches improve resilience against compromised nodes by estimating trust and filtering malicious behavior. However, fixed fuzzy rules may not adapt well to changing attack intensity or evolving traffic patterns. In addition, security-oriented fuzzy methods often focus on identifying reliable CHs but do not directly minimize information staleness. Trust management methods are valuable because CHs have a critical role in data aggregation and forwarding. A malicious CH can drop packets or inject false data. However, trustworthiness alone is not sufficient; a trusted node might have low energy, suffer from high AoI, occupy an unsuitable spatial location, or perhaps have been selected multiple times previously.

3.4. AoI-Aware Data Freshness in Wireless Networks

Age of Information (AoI) measures the freshness of information at the destination and has become a major performance metric for real-time monitoring systems. Kaul et al. introduced AoI as a formal metric for real-time status updating [8], and Yates and Kaul later provided a detailed theoretical treatment of AoI in multi-source updating systems [9]. As noted in the Introduction, AoI captures information staleness rather than packet travel time, so a system can have low delay while still failing to update the sink frequently enough.

Recent AoI-aware studies have considered scheduling, blocklength allocation, wireless-powered networks, update-rate control, and reinforcement learning for timely updates [28]-[30], showing that freshness must be treated as a first-class objective in real-time sensing systems. However, many AoI-focused studies do not jointly consider secure CH selection, trust, suspicion, or attack behavior. They often optimize scheduling or communication decisions while assuming that nodes are cooperative and reliable. In tactical WSNs, this assumption is incomplete because compromised or suspicious nodes can degrade the reliability of fresh data. The proposed method addresses this gap by incorporating AoI into the CH selection reward while simultaneously considering security and energy constraints.

3.5. Edge Computing and Digital-Twin-Assisted WSN Intelligence

Edge computing has been proposed as a way to move computation closer to the data source, reducing latency and relieving resource-constrained sensor nodes from heavy processing [10], [15]. In WSNs, edge-assisted architectures can support local anomaly detection, state monitoring, and intelligent routing decisions at a sink, gateway, or tactical controller. This is highly relevant to the proposed work because the anomaly scoring and Q-learning decisions can be naturally executed at an edge-assisted controller rather than on every sensor node. Digital twin research has also gained attention as a way to maintain virtual representations of physical systems for monitoring, prediction, and decision support [11]-[14]. In a WSN context, a digital twin can mirror node energy, trust, AoI, traffic, packet drop, cluster membership, and attack indicators. This virtual representation can then be used to simulate routing decisions, predict node degradation, and optimize CH selection before applying decisions to the physical network. The present work does not claim to implement a full digital twin; the simulation only maintains a structured per-node virtual state updated across rounds, making it suitable for future extension into an edge-assisted digital-twin management layer.

3.6. Research Gap

The reviewed literature shows that WSN clustering and routing research has advanced significantly, but several gaps remain. Classical clustering protocols are energy-aware but not security-aware or AoI-aware. RL-based methods improve adaptivity but often optimize energy or routing quality without explicitly incorporating trust, suspicion, attack status, and cluster-level AoI. Fuzzy trust-based methods improve security awareness but depend on fixed rules and often neglect freshness. AoI-aware studies improve information timeliness but usually do not integrate adversarial behavior into CH selection.

Therefore, the main research gap is the absence of a unified CH selection framework that jointly considers energy efficiency, information freshness, trust, anomaly-driven suspicion, attack status, spatial suitability, and learning-based adaptation under one fair evaluation environment. The proposed method addresses this gap by combining K-Means clustering, temporal anomaly scoring, conservative candidate filtering, and Q-learning-based multi-objective CH selection.

Table 1: Relationship between representative previous studies and the proposed work

Study	Main Methodology	Main Metrics	Security-Aware	AoI-Aware	Learning-Based	Main Limitation
LEACH [3], [4]	Probabilistic CH rotation	Energy, lifetime	No	No	No	Random CH selection; no trust or freshness
HEED [18]	Distributed energy-aware clustering	Energy, communication cost	No	No	No	Focuses mainly on energy
PEGASIS [19]	Chain-based routing	Energy, lifetime	No	No	No	May increase delay; no attack model
TEEN/APTEEN [20], [21]	Threshold-based reporting	Energy, event reporting	No	Limited	No	Not designed for secure CH selection
Metaheuristic clustering [22], [23]	Swarm/optimization-based CH selection	Energy, coverage, routing cost	Limited	No	No	Optimization often targets energy only
Mouli et al. [16]	RL-based CH selection and routing	Energy, routing efficiency	Limited	No	Yes	Does not jointly optimize AoI, trust, suspicion, and attack status
RL/game-theoretic WSN methods [24], [25]	RL or multi-agent routing	Energy, lifetime, load balance	Limited	No	Yes	Security and freshness are not central objectives
Learning-based WSN intelligence [31]	ML-supported routing, clustering, prediction, and anomaly analysis	Energy, prediction, routing quality	Sometimes	Limited	Yes	Often task-specific and not a unified secure-AoI CH framework
Dhipa et al. [17]	Fuzzy secure CH selection	Trust, security, lifetime	Yes	No	No	Fixed rules; no explicit AoI optimization
Fuzzy trust/outlier methods [26]	Fuzzy trust and outlier detection	Trust, attack resilience	Yes	No	No	Limited adaptivity and freshness awareness

Energy-efficient routing protocols [27]	Energy-aware routing and CH optimization	Energy, lifetime	Limited	No	No	Freshness and attack behavior are not central
AoI-aware scheduling/routing [8], [9], [28]-[30]	AoI optimization	Freshness, latency	No/limited	Yes	Sometimes	Usually does not integrate secure CH selection
Edge/digital twin studies [10]-[15]	Edge intelligence and virtual monitoring	Latency, monitoring, prediction	Potentially	Potentially	Depends	Mostly architectural; not a complete WSN CH selection solution
Proposed Method	K-Means + anomaly-aware Q-learning CH selection	Energy, lifetime, AoI, trust, suspicion, PDR, delay	Yes	Yes	Yes	Direct routing currently used; full digital twin left for future work

4. System Model

The simulated network includes 100 static sensor nodes randomly distributed within a 100 m × 100 m area. The sink node is located at coordinates (50, 120)—outside the sensing range—thereby imposing a realistic constraint: the need for long-distance transmission by nodes acting as cluster heads (CHs). Each node begins with an initial energy of 1.5 Joules and is considered "dead" when its energy reaches or falls below 0.05 Joules. The simulation runs for 2,500 rounds.

Each node keeps the following information: residual energy, Age of Information (AoI), confidence level, data traffic rate, packet loss ratio, attack status, attack type, spatial coordinates, number of Cluster Heads (CHs), current cluster, node degree, risk memory, alarm count, and historical records of data traffic, packet loss, AoI, and energy. This information is updated in each round and are utilized in Cluster Head (CH) selection, anomaly detection, and performance evaluation processes.

evaluation.

4.1. Energy Model

The simulation uses the first-order radio energy model. The transmission energy depends on the packet size and the transmission distance. For short distances, the free-space model is used; for long distances, the multipath model is used:

$$E_{tx}(k,d) = k(E_{elec} + E_{fs}\, d^2), \text{ if } d<d_0$$

$$E_{tx}(k,d) = k(E_{elec} + E_{mp}\, d^4), \text{ if } d \geq d_0$$

$$E_{rx}(k) = kE_{elec}$$

$$E_{DA}(k) = kE_{DA}$$

where k is the packet size measured by bits, d is the transmission distance, E_{elec} is the electronics energy, E_{fs} is the parameter of free space amplifier, E_{mp} is the parameter of multipath amplifier, and E_{DA} is the data aggregation energy. The threshold distance d_0 is computed by

$$d_0 = \sqrt{\frac{E_{fs}}{Emp}}$$

Using the parameters in Table 2 (E_{fs} = 10 pJ/bit/m², Emp = 0.0013 pJ/bit/m⁴), d_0 evaluates to approximately 87.7 m, which is greater than the deployment area diagonal, meaning the free space model applies to all intra-cluster transmissions, while the multipath model governs long-distance CH-to-sink transmission. During each round, member nodes transmit data to their CH. The CH receives member packets, aggregates the data, and forwards the aggregated packet to the sink. This model captures the extra burden imposed on CHs and allows fair comparison of energy consumption across protocols.

5. Proposed Framework

The proposed framework consists of five main components: K-Means clustering, temporal security feature extraction, attack modeling, candidate filtering, and Q-learning-based CH selection.

The design philosophy of the proposed framework is based on the observation that CH selection in tactical WSNs is a multi-objective decision problem. A CH is not merely a node with high residual energy; it must also be reliable, low-risk, spatially suitable, and capable of supporting fresh information delivery. Therefore, the framework introduces a layered decision process. First, the network is organized spatially. Second, node behavior is monitored using temporal features. Third, risky candidates are filtered. Fourth, the remaining candidates are evaluated using a reward function. Finally, Q-learning is used to improve decision consistency across rounds.

This layered design was adopted to avoid overloading the learning agent with a large and noisy action space.

If all nodes were passed directly to the reinforcement learning agent without filtering or clustering, the number of possible CH selection decisions would increase significantly, and the learning process would become less stable. By combining K-means clustering and filtering the best candidates, the proposed method reduces complexity while preserving the most relevant candidates for learning-based selection.

5.1. K-Means Clustering

The K-Means algorithm [5] is used to divide the network into 10 spatial groups based on the node coordinates. Each node is assigned to a group, and each group has a center. This organization reduces the complexity of center selection because the proposed method selects a single center for each group instead of selecting centers in general from all nodes

The centroid is also used in the reward function. Nodes closer to the cluster centroid are preferred because they tend to reduce intra-cluster communication distance. This decreases the energy consumed by member nodes when transmitting to the CH.

To address the computational overhead and energy constraints associated with spatial clustering, the K-Means algorithm is not executed repeatedly by the resource-constrained sensor nodes. Instead, it is executed strictly once as an offline, centralized process by the Base Station (Sink) immediately following network deployment. The Base Station computes the cluster centroids using the initial node coordinates and broadcasts the cluster assignments. This approach completely offloads the heavy computational burden from the sensor nodes, preserving their energy exclusively for sensing and communication tasks in the tactical environment.

The use of K-Means also makes the proposed method scalable. Instead of learning one global CH selection policy over all nodes, the method learns localized decisions within each cluster. This is useful in WSNs because communication cost is strongly distance-dependent. Spatially localized clustering therefore reduces both the physical communication burden and the computational burden of selecting CHs.

5.2. Security Feature Extraction

Each node i maintains a sliding history window of length H = 6 rounds for four behavioral indicators: traffic rate tau_i, packet drop ratio delta_i, Age of Information A_i, and residual energy E_i. At each round t, the AoI of every alive node is incremented by one time unit to reflect increasing staleness. When a node serves as CH, its AoI is reset to 1 upon successful data aggregation and forwarding to the sink, as shown in Equation (1). Member nodes' AoI is also reset to 1 upon successful data delivery to their CH in the same round, as specified in Algorithm 1 (line 21). This means that AoI accumulates for all nodes between successful delivery events and is bounded by the inter-delivery interval. Under this model, the network-level AoI reflects the combined effect of CH selection quality and per-round delivery success rather than CH rotation frequency alone.

$$A_i(t) = 1, \text{ if node } i \text{ is selected as CH at round } t$$

$$A_i(t) = A_i(t-1) + 1, \text{ otherwise} \quad (1)$$

Four temporal delta features are extracted by comparing the current value of each indicator against its historical average over the preceding H-1 rounds. Let mu_tau, mu_delta, mu_A, and mu_E denote the historical averages of traffic, packet drop, AoI, and energy respectively. The temporal features are defined in Equation (2):

$$Delta_tau_i = max(0, tau_i - mu_tau_i)$$

$$Delta_delta_i = max(0, delta_i - mu_delta_i)$$

$$Delta_A_i = max(0, A_i - mu_A_i)$$

$$Delta_E_i = max(0, mu_E_i - E_i) \quad (2)$$

The normalized temporal jump scores are then computed by clamping each delta to [0, 1] using fixed reference thresholds: 0.35 for traffic, 0.25 for packet drop, and 0.50 for energy drop. These reference values represent experimentally defined limits beyond which behavioral change is considered abnormal. The standard jumps J_tau, J_delta, and J_E are defined in Equation (3):

$$J_tau_i = clamp(Delta_tau_i / 0.35, 0, 1)$$
$$J_delta_i = clamp(Delta_delta_i / 0.25, 0, 1)$$
$$J_E_i = clamp(Delta_E_i / 0.50, 0, 1) \quad (3)$$

The edge anomaly score is calculated from two components: an immediate score, S_inst, and a time score, S_temp, as defined in Equations (4) and (5). The risk memory variable, M_i, accumulates suspicious behavior across rounds using an exponential decay factor of 0.80, preventing the system from ignoring recurring anomaly patterns that may seem insignificant in a single round but become significant over time. The edge anomaly index, E_anom, is defined in Equation (6).

$$S_inst_i = 0.35 * tau_i + 0.35 * delta_i + 0.30 * (1 - trust_i) \quad (4)$$
$$S_temp_i = 0.30 * J_tau_i + 0.35 * J_delta_i + 0.20 * J_E_i \quad (5)$$
$$M_i(t) = clamp(0.80 * M_i(t-1) + 0.60 * S_temp_i + 0.30 * S_inst_i)$$
$$E_anom_i = clamp(0.50 * S_inst_i + 0.35 * S_temp_i + 0.15 * M_i) \quad (6)$$

The final suspicion score Susp_i is computed as a weighted combination of traffic rate, packet drop ratio, distrust, and edge anomaly score, as defined in Equation (7). Nodes identified as attacking a sinkhole receive an additional fixed penalty of 0.05 because the sinkhole behavior presents primarily through attracting abnormal traffic rather than directly dropping packets, making it difficult to detect through drop-based indicators alone:

$$Susp_i = clamp(0.25 * tau_i + 0.25 * delta_i + 0.30 * (1 - trust_i) + 0.20 * E_anom_i)$$
$$Susp_i = clamp(Susp_i + 0.05), \text{ if } attack_type_i = Sinkhole \quad (7)$$

All scores are confined to the range [0, 1]. The Susp_i uncertainty score is used in both candidate filtering (where nodes exceeding the S_max threshold = 0.65 are excluded) and the Q learning reward function (where it appears as a penalty limit). Including the risk memory M_i ensures that nodes exhibiting persistent minor anomalies are penalized more severely than those showing sudden spikes in a single round, thus improving detection stability under dynamic attack patterns.

This security model is lightweight and suitable for resource-constrained wireless sensor networks. It does not require a complex central classifier or a large training dataset. Instead, it uses locally available behavioral indicators that are commonly observable in WSN operation. The purpose of this mechanism is not to replace a full intrusion detection system, but to provide a practical suspicion signal that can be embedded into CH selection. This is valuable because preventing suspicious nodes from becoming CHs can reduce the damage caused by attacks even when attack detection is not perfect.

5.3. Attack Model

The simulation includes four attack types: DoS, FDI, Selective Forwarding, and Sinkhole. Each attack is injected probabilistically per node per round.

DoS increases traffic and packet drop while reducing trust. FDI increases AoI and traffic while reducing trust. Selective Forwarding increases packet drop and reduces trust. Sinkhole increases traffic and reduces trust. If no attack occurs in a round, trust slightly recovers and traffic and packet drop gradually decrease. This dynamic attack model creates a challenging environment where nodes may change behavior over time.

5.4. Candidate Filtering

Before applying Q-learning, the proposed method filters CH candidates inside each cluster. A node must have sufficient residual energy, acceptable trust, and low suspicion to be considered a strong candidate. If no node meets these requirements, a strict minimum trust policy is applied to maintain tactical security. Instead of resorting to choosing a highly suspect or compromised node simply because it's the only one left, the system enforces an absolute minimum trust requirement. (e.g., $T_{min} \geq 0.20$). If all remaining alive nodes in a cluster fall below this critical security floor, CH selection is deliberately suspended to prevent network poisoning. In this state, the remaining untrusted nodes are denied the CH role, thereby preventing a compromised node from performing data aggregation or sinkhole attacks during the late stages of network depletion.

The candidates are then ranked using low suspicion, low AoI, high residual energy, high trust, and low previous CH count. Only the top candidates are passed to the Q-learning selection stage. This reduces the action space and prevents obviously risky or weak nodes from dominating the learning process.

This filtering stage is intentionally conservative. It does not remove too many candidates or aggressively force the network to select only the safest node. Instead, it eliminates clearly unsuitable candidates and allows the reward-driven learning stage to make the final decision. This design avoids a common problem in security-aware WSN optimization: if security thresholds are too strict, the protocol may sacrifice energy balance, reduce CH diversity, or select spatially inefficient nodes. The proposed method therefore uses filtering as a safety layer and Q-learning as the optimization layer.

5.5. Reward Function

The proposed reward function is designed to balance energy, security, freshness, spatial suitability, and fairness. It rewards high energy and high trust, while penalizing node AoI, cluster average AoI, cluster maximum AoI, suspicion, active attack status, distance from the centroid, and previous CH count:

Reward = wE Energy + wT Trust - wAoI NodeAoI - wCA ClusterAvgAoI - wCM ClusterMaxAoI - wS Suspicion - wAtt Attack - wD Distance - wC CHCount

This reward function is the core novelty of the proposed approach. It ensures that CH selection does not depend only on residual energy. A node with high energy but high suspicion or stale information is penalized. Similarly, a trusted node that is far from the centroid or repeatedly selected as CH is also penalized. Therefore, the selected CH must provide a balanced trade-off among the main WSN requirements.

The reward terms were selected to reflect the operational role of a CH. Residual energy protects lifetime. Trust and suspicion protect security. Attack status prevents currently compromised nodes from being favored. Node AoI and cluster AoI protect information freshness. Distance from centroid reduces intra-cluster communication cost. CH reuse count improves fairness by discouraging repeated selection of the same node. The combined reward therefore converts CH selection from a single-objective energy problem into a practical multi-objective decision problem.

An important aspect of the reward function is the integration between the average age of the cluster (AoI) and the maximum age of the cluster (AoI). The average age of the cluster (AoI) reflects how young the cluster is overall, while the maximum age of the cluster (AoI) reflects the worst-case scenario for an older node.

5.6. Q-Learning-Based CH Selection

Specifically, state variables are divided into discrete levels (e.g., 4 levels for energy and AoI, and 3 levels for confidence and doubt), resulting in a very limited local state space of $4 \times 4 \times 3 \times 3 = 144$ discrete states per cluster. This strict limit prevents state inflation and ensures that the memory load remains minimal for resource-constrained sensor nodes. Possible actions correspond to the heads of the sorted candidate clusters. An epsilon-bold policy is used to balance exploration and exploitation. The probability of exploration decreases over time, allowing the system to explore early and exploit learned decisions later.

After selecting a CH, the reward is computed and the Q-value is updated. This allows the proposed method to learn which types of nodes are more suitable as CHs under changing energy, trust, AoI, and attack conditions.

Using one Q-table per cluster improves learning stability because each table learns decisions for a smaller local region rather than the entire network. This reduces state-action complexity and makes the learned policy more relevant to local cluster conditions. In addition, the discretized state representation keeps the Q-table compact and avoids excessive memory requirements, which is important for WSN-inspired designs.

The epsilon-greedy strategy also reflects the changing nature of the network. Early exploration helps the protocol test different CH candidates and avoid premature convergence. Later exploitation allows the protocol to reuse successful decisions and stabilize CH selection. This explains why the proposed method shows very low final CH changes and maintains high final quality in the late simulation stage.

5.7. Reward Weight Selection and Sensitivity Analysis

The reward weights and candidate filtering thresholds used in the proposed framework were determined through a systematic empirical tuning process. An initial set of weights was derived from domain reasoning about the relative operational importance of each criterion in tactical WSN deployments. These initial values were then manually refined by running the simulation under a

limited set of varying weight configurations and observing their effect on the final quality score, network AoI, selected-CH trust, and selected-CH suspicion. The final adopted values are summarized in Table 2.

Table 2: Reward Function Weights and Candidate Filtering Thresholds

Parameter	Symbol	Value	Design Rationale
Energy weight	w_E	9.0	Highest weight; residual energy is the primary lifetime constraint in resource-constrained WSNs
Node AoI weight	w_AoI	8.0	Second highest; information freshness is the central requirement in tactical real-time monitoring
Attack status weight	w_Att	6.0	Equal to trust weight; an actively attacked node poses comparable risk to a low-trust node
Trust weight	w_T	6.0	High weight; trust directly reflects CH reliability and security suitability for cluster aggregation
Cluster avg AoI weight	w_CA	5.0	Moderately high; captures collective cluster-level staleness beyond individual node AoI
Suspicion weight	w_S	4.5	Slightly lower than trust; suspicion is a probabilistic temporal signal and may include false positives
Cluster max AoI weight	w_CM	2.5	Moderate; prevents worst-case stale nodes from being hidden by acceptable cluster averages
Distance weight	w_D	1.2	Low weight; spatial proximity is a secondary factor relative to security and freshness criteria
CH count weight	w_C	1.2	Low weight; CH rotation fairness is beneficial but should not override energy or security decisions
Min energy threshold	E_min	0.30 J	Corresponds to 20% of initial energy (1.5 J); ensures candidates retain meaningful remaining capacity
Min trust threshold	T_min	0.35	Intentionally lenient to eliminate clearly unreliable nodes while preserving candidate diversity
Max suspicion threshold	S_max	0.65	Conservative upper bound; allows moderate suspicion to remain eligible and prevents cluster failure
Top-K candidates	K	5	Balances action-space reduction with sufficient candidate diversity for stable Q-learning

The weight ordering reflects a deliberate priority structure. Energy, with a weight (w_E) of 9.0, and node AoI, with a weight (w_AoI) of 8.0, are set to the highest weights as they are considered the two main operational requirements in tactical WSNs: Keeping network continuity and ensuring information freshness. The status of attack, with a weight (w_Att) of 6.0, and trust, with a weight (w_T) of 6.0, are set equal weights because a node under active attack and a structurally untrustworthy node poses comparable risks to cluster integrity. Suspicion, with a weight (w_S) of 4.5, is given slightly less weight than trust because it is derived from probabilistic temporal features and may include false positives under dynamic traffic conditions. The cluster-level AoI weights (w_CA = 5.0 and w_CMA = 2.5) are kept at intermediate levels to capture collective freshness without overriding individual node-level signals."Distance (1.2) and CH reuse count (1.2) are assigned the lowest weights because spatial proximity and rotation fairness, while beneficial, are secondary to security and freshness in tactical deployments.

To assess the sensitivity of the proposed method to weight variations, selected weight parameters were varied individually while holding the remaining weights constant at their adopted values. Increasing w_E beyond 10.0 improved lifetime marginally but

degraded AoI because high-energy nodes were selected regardless of their freshness state. Reducing w_AoI below 5.0 caused a measurable increase in final average network AoI, confirming that explicit AoI penalization is essential for freshness control. Increasing w_S above 7.0 produced overly conservative candidate filtering, which occasionally forced the selection of spatially unsuitable or energy-weak CHs when suspicious nodes were temporarily dominant in a cluster. These observations confirm that the adopted weight configuration represents a stable operating point that provides a balanced trade-off among the main performance dimensions. To assess cross-scenario generalizability, the same fixed weight configuration was applied without re-tuning to eight additional scenarios covering 200- and 300-node networks and simulation lengths of 5000 and 10000 rounds. The AoI and energy advantages were reproduced consistently across all eight alternative scenarios, confirming that the weight configuration is not overfit to the primary 100-node, 2500-round setting. Full cross-scenario results are reported in the Limitations section.

Candidate eligibility thresholds were conservatively set to avoid overly restricting the pool of candidates. The minimum power requirement of 0.30 J corresponds to 20% of the initial node's power (1.5 J), ensuring that only nodes with valuable remaining capacity are eligible.

The minimum trust threshold of 0.35 and maximum suspicion threshold of 0.65 are intentionally lenient to prevent cluster failure in sparse or heavily attacked networks. When no candidate satisfies all three thresholds simultaneously, the fallback mechanism activates and considers all alive nodes in the cluster, preventing isolated cluster collapse.

5.8. Algorithmic Workflow

The complete proposed workflow can be summarized as follows:

Algorithm 1: AoI- and Security-Aware Cluster-Head Selection via Local Q-Learning

```
Input:  Node set N = {n_1,...,n_m} with initial energy E_0, positions,
        trust T_i, suspicion S_i; sink position; K clusters;
        episode length R; Q-tables Q_k, k = 1..K
Output: Per-round metrics (AoI, trust, suspicion, energy, CH
        assignments, lifetime); learned Q-tables Q_k
```

```
Initialize each n_i: energy E_0, trust T_i←1, suspicion S_i←0,
    AoI_i←0, history buffer H_i of length 6
{C_1,...,C_k} ← K-Means(positions of N, K)   ▷ spatial clustering
for round r = 1 to R do
  for each alive node n_i ∈ N do
    AoI_i ← AoI_i + 1   ▷ increment age
    with prob. p_attack: apply {DoS, FDI, SF, Sinkhole} to n_i
    append {energy, AoI, traffic, drops, attack-flag} to H_i;
      drop oldest entry if |H_i| > 6
    f_i ← TemporalAnomalyFeatures(H_i)
    a_i ← EdgeAnomalyScore(f_i)   ▷ shared external evaluator
    S_i ← UpdateSuspicion(S_i, a_i);  T_i ← UpdateTrust(T_i, a_i)
  end for
  for each cluster C_k, k = 1 to K do
    Cand ← {n_i ∈ C_k : E_i ≥ E_min ∧ T_i ≥ T_min ∧ S_i ≤ S_max}
    if Cand = ∅ then Cand ← {n_i ∈ C_k : n_i alive ∧ T_i ≥ 0.20}   ▷ Strict fallback
    if Cand = ∅ then CH_k ← NULL; continue   ▷ Suspend CH selection to prevent
poisoning
    Cand ← Top-5(Cand, score = f(S_i, AoI_i, E_i, T_i, ch_cnt_i))
    s_k ← (Ē_k, ĀoI_k, ĀoI_max,k, T⁻_k, S⁻_k)   ▷ discretized cluster state
    with prob. ε: CH_k ← Random(Cand)   ▷ exploration
    else: CH_k ← argmax_{c∈Cand} Q_k(s_k, c)   ▷ exploitation
    Transmit(members of C_k → CH_k);  deduct E_ele, E_fs/E_mp
    Aggregate(CH_k, E_DA);  Transmit(CH_k → sink)
```

```
22:      reset AoIᵢ for nᵢ ∈ C_k upon successful delivery
23:      r_k ← w_E·E_term + w_AoI·AoI_term + w_atk·atk_term + w_T·T_term
               + w_cAoI·cAoI_term + w_S·S_term + w_cMaxAoI·cMaxAoI_term +
w_dist·dist_term + w_cnt·cnt_term
24:      s_k′ ← ObserveNextState(C_k)
25:      Q_k(s_k, CH_k) ← Q_k(s_k, CH_k) + α[r_k + γ·maxᶜ′ Q_k(s_k′,c′) − Q_k(s_k,CH_k)]
26:    end for
27:    Record metrics for round r
28:    if no alive nodes remain then break
29: endfor
```

Algorithm 1 formalizes the step-by-step workflow described above. The outer loop (lines 3–28) iterates over simulation rounds; the first inner loop (lines 4–11) updates per-node AoI, applies the probabilistic attack model, and recomputes trust and suspicion via the shared external anomaly evaluator; the second inner loop (lines 12–25) performs candidate filtering, epsilon-greedy CH selection, communication and energy accounting, AoI reset, multi-objective reward computation, and the Q-learning update for each cluster independently. The per-round bookkeeping in lines 26–27 produces the metrics summarized in Table 5 (Comparative Analysis).

5.9. Computational Considerations

The computational complexity of the proposed method is controlled by three mechanisms. First, K-Means clustering partitions the network into smaller groups, so CH selection is performed locally. Second, candidate filtering reduces the number of actions passed to Q-learning. Third, the Q-table state is discretized using rounded average values, which limits the number of possible states. These design choices make the proposed method more practical than a global RL approach that evaluates all possible CH combinations across the entire network.

The memory overhead is also moderate because each cluster maintains its own Q-table, and each node stores only a short temporal history of length six. Therefore, the method provides learning-based adaptivity without requiring deep neural networks or large memory buffers. This is consistent with the resource constraints of WSN environments, particularly tactical ones.

6. Simulation Setup

To obtain a fair and meaningful comparison, the simulation framework was unified across all evaluated methods. A common network setting was used, including deployment area, sink location, initial energy, packet size, radio-energy model, attack probabilities, and evaluation pipeline. The robustness study was conducted using three network sizes (100, 200, and 300 nodes), three simulation lengths (2500, 5000, and 10000 rounds), and twenty independent random seeds (42, 123, 256, 789, 1024, 0, 1, 7, 10, 21, 50, 64, 99, 100, 314, 512, 777, 999, 2024, and 4096). Combining the three network sizes and three simulation lengths yielded nine robustness configurations (3×3), each of which was evaluated using the same twenty random seeds.The figures are reported for the 100-node setting to maintain a fair and readable visual comparison among the four methods under the same graphical scenario, while the tables report the extended 20-seed and multi-scale numerical evaluation. The same external security evaluator was applied to all methods so that the comparison reflects protocol behavior rather than differences in how security is measured.

Table 3: Main simulation parameters used in the experiments.

The threshold distance $d_0 = \sqrt{(E_{fs}/E_{mp})} \approx 87.7$ m is derived from E_{fs} and E_{mp}.

Parameter	Value	Description
Number of sensor nodes	100, 200, and 300	Network sizes used in the robustness study; figures are shown for the 100-node scenario for fair visual comparison.
Number of clusters	10	K-Means clusters used by the proposed method.
Simulation rounds	2500, 5000, and 10000	Simulation lengths used in the robustness study; figures are shown for the 2500-round scenario.
Deployment area	100 m x 100 m	Two-dimensional sensing field
Sink position	(50, 120)	Base station located outside the sensing field
Initial node energy	1.5 J	Initial energy assigned to each sensor node
Dead-node threshold	0.05 J	Node considered dead below or equal to this energy
Radio range	25 m	Communication neighborhood range
Data packet size	6400 bits	Main sensed data packet size
Control packet size	512 bits	Control packet size

E_{elec}	50 nJ/bit	Radio electronics energy
E_{fs}	10 pJ/bit/m^2	Free-space amplifier parameter
E_{mp}	0.0013 pJ/bit/m^4	Multipath amplifier parameter
E_{DA}	5 nJ/bit	Data aggregation energy
DoS attack probability	0.015	Per-node per-round DoS probability
FDI attack probability	0.015	Per-node per-round false data injection probability
Selective forwarding probability	0.015	Per-node per-round selective forwarding probability
Sinkhole attack probability	0.015	Per-node per-round sinkhole probability
Default attack threshold	0.50	Unified external security evaluation threshold
Q-learning alpha	0.10	Learning rate
Q-learning gamma	0.90	Discount factor
Q-learning epsilon	0.20	Exploration probability for RL baseline
CH ratio	0.10	Target CH ratio used by RL/RFGD-style selection
LEACH probability	0.10	LEACH CH election probability
History length	6	Temporal history window for anomaly features
Random seeds	42, 123, 256, 789, 1024, 0, 1, 7, 10, 21, 50, 64, 99, 100, 314, 512, 777, 999, 2024, and 4096	Twenty independent random seeds used for all reported mean ± standard deviation results.
REWARD_W_ENERGY	9.0	Energy reward weight in CH selection
REWARD_W_AOI	8.0	Node AoI penalty weight in CH selection
REWARD_W_ATTACK	6.0	Attack status penalty weight in CH selection
REWARD_W_TRUST	6.0	Trust reward weight in CH selection
REWARD_W_CLUSTER_AOI	5.0	Cluster average AoI penalty weight in CH selection
REWARD_W_SUSP	4.5	Suspicion penalty weight in CH selection
REWARD_W_CLUSTER_MAX_AOI	2.5	Cluster maximum AoI penalty weight in CH selection
REWARD_W_DISTANCE	1.2	Centroid distance penalty weight in CH selection
REWARD_W_CH_COUNT	1.2	CH reuse count penalty weight in CH selection
Candidate energy min (E_min)	0.30 J	Minimum residual energy required for CH candidacy (20% of initial)
Candidate trust min (T_min)	0.35	Minimum trust score required for CH candidacy
Candidate suspicion max (S_max)	0.65	Maximum suspicion score permitted for CH candidacy
Top-K candidates	5	Maximum number of candidates passed to Q-learning per cluster per round

6.1. Fairness of Baseline Implementation

A major part of the experimental design was ensuring that the baselines remained as faithful as possible to their original design philosophies. LEACH was kept as a classical probabilistic clustering protocol without adding AI-based decision making, security gates, or AoI-aware modifications. The RL-based baseline was implemented through a paper-faithful approximation using LO-SARSA-inspired CH selection, Q-learning-based cluster assignment, and PO-PGM-inspired routing. The RFGD baseline was implemented using fuzzy trust/ chance evaluation, generic detection, and CH reselection.

The proposed method, in contrast, uses K-Means-based spatial clustering and an RL-guided CH selection reward that jointly considers residual energy, trust, suspicion, AoI, cluster AoI, attack status, distance, and CH reuse count. This distinction is intentional: baselines were not forced into the same security- or AoI-awareness as the proposed method, for reasons detailed in Section 6.2.

6.2. Development, Tuning, and Reproducibility

During development, several intermediate versions were tested and rejected. Early runs produced unrealistic lifetime behavior in which node deaths were not sufficiently visible, so the simulation conditions were tuned to make lifetime effects observable and interpretable. Plotting was also improved by applying smoothing to non-lifetime figures so that performance trends could be interpreted clearly instead of appearing as noisy fluctuations. Additional work was performed to store all generated figures,

comparison tables, serialized results, learned Q-tables, and summary reports in a structured way. These steps were necessary not only for presentation quality, but also to ensure that each experiment could be reproduced and audited later.

Security tuning was a particularly important part of the development process. Several security-oriented modifications were explored, including stricter suspicion thresholds, trust gates, cooldown ideas, and changes to the external evaluator. However, aggressive security tuning often damaged the overall balance of the proposed model. In some trials, security-detection numbers increased only because the shared evaluator had been changed, which improved or degraded all methods simultaneously and reduced the fairness of the comparison. In other trials, stronger security gating inside the proposed selection logic caused undesirable side effects, such as lower selected-CH trust, higher selected-CH suspicion, lower quality score, or poorer network balance.

For this reason, the final adopted version follows a conservative strategy. The external evaluator is left unchanged to preserve fairness across all methods, while only light additional screening is used inside the proposed model during CH candidate selection. This allows the proposed method to avoid clearly risky candidates without fundamentally changing the evaluation process or destroying its main strengths. Therefore, the final model was not selected because it maximizes one security metric in isolation, but because it provides a balanced overall trade-off between data freshness, network lifetime, communication quality, and security-aware behavior.

The experimental development process also focused on avoiding unfair baseline enhancement. For example, LEACH could have been modified to include trust thresholds or AoI penalties, but doing so would no longer represent classical LEACH. Similarly, the RL baseline could have been modified to use the same suspicion and AoI reward terms as the proposed method, but that would obscure the contribution of the proposed reward design. Therefore, each baseline was preserved according to its own logic, while the surrounding simulation and evaluation environment was unified. This strengthens the validity of the comparison because observed performance differences can be attributed to the protocols themselves rather than to inconsistent experimental settings.

Another important implementation effort was result traceability. The simulation automatically produces plots, comparison tables, JSON files, pickle files, Q-table files, and text summaries. This enables post-run verification of each method and makes it possible to inspect both final metrics and temporal behavior across rounds. Such traceability is essential for IEEE-style experimental reporting because it allows the results to be audited, reproduced, and extended.

6.3. Interpretation Strategy

The performance of WSN clustering protocols cannot be judged using a single metric. A method may achieve low energy consumption by sending less useful information, or high PDR while allowing AoI to grow excessively. Similarly, a method may show slightly higher detection accuracy while still selecting suspicious nodes as CHs. For this reason, the results are interpreted using a multi-metric strategy. Lifetime, AoI, trust, suspicion, delay, PDR, energy consumption, CH stability, and security detection metrics are discussed together.

A composite Quality Score is used as a single summary indicator, computed by the shared external evaluator independently of the proposed method's reward function. It is defined as a weighted linear combination of six per-round, per-method normalized indicators:

$$QS = w_1 \cdot AoI_norm + w_2 \cdot (1 - Suspicion) + w_3 \cdot Trust + w_4 \cdot PDR + w_5 \cdot AliveRatio + w_6 \cdot EC_norm \qquad (8)$$

where $AoI_norm = 1/(1 + AoI)$ and $EC_norm = 1/(1 + 100 \cdot EC)$ set the "less is better" measures to [0, 1], and the weights ($w_1 = 0.225$, $w_2 = 0.309$, $w_3 = 0.140$, $w_4 = -0.090$, $w_5 = 0.321$, $w_6 = 0.099$, intercept = 0.210) were estimated from the simulation evaluator and verified by linear regression on the per-seed results ($R^2 = 0.993$). Critically, this formula is computed identically for all four methods and is entirely independent of the proposed method's reward weights; the evaluator has no access to the Q-learning policy. This confirms that Quality Score comparisons are not circular. The small negative coefficient on PDR ($w_4 = -0.090$) does not imply that higher packet delivery reduces network quality; rather, it reflects multicollinearity between PDR and AliveRatio in the regression data, since both indicators tend to move together across methods. Given its small magnitude relative to the other weights, this term has a negligible effect on the overall Quality Score ranking.

This interpretation strategy is important for the proposed method because its goal is not to dominate every individual metric. Instead, its goal is to deliver the most convincing overall performance profile under tactical WSN requirements. In this context, the most consistent evidence for the proposed method is the combination of very low and stable AoI, high final quality, low and stable final energy consumption, and a substantially longer stable lifetime than LEACH and RL, while selected-CH trust and suspicion remain seed-dependent and represent a direction for future tuning.

7. Results and Discussion

This section discusses the experimental results with emphasis on the mechanisms behind the observed trends. The main contribution of this work is AoI-aware and security-aware CH selection; therefore, the discussion prioritizes freshness, final quality, energy behavior, and balanced operation rather than isolated improvement in every metric.

Across the twenty random seeds described in Section 6, the proposed method achieves the lowest AoI and the highest final quality in the 100-node scenario and preserves the same trend across the extended robustness study. It remains competitive in communication and security metrics, while RFGD and RL show advantages in some isolated metrics. Therefore, the results should be interpreted as a balanced AoI-driven contribution rather than a claim of superiority in every metric.

7.1. Network Lifetime Analysis

Fig. 1 compares the number of alive nodes over time. The proposed method extends functional lifetime compared with LEACH and the RL-based baseline because CH selection is not random or purely topology-driven; it rewards residual energy, penalizes repeated CH reuse, and prefers spatially suitable nodes near cluster centroids. RFGD retains more residual nodes in some runs, but its high AoI shows that node survival alone does not necessarily indicate useful tactical operation. Thus, the proposed method provides a balanced lifetime-freshness behavior.

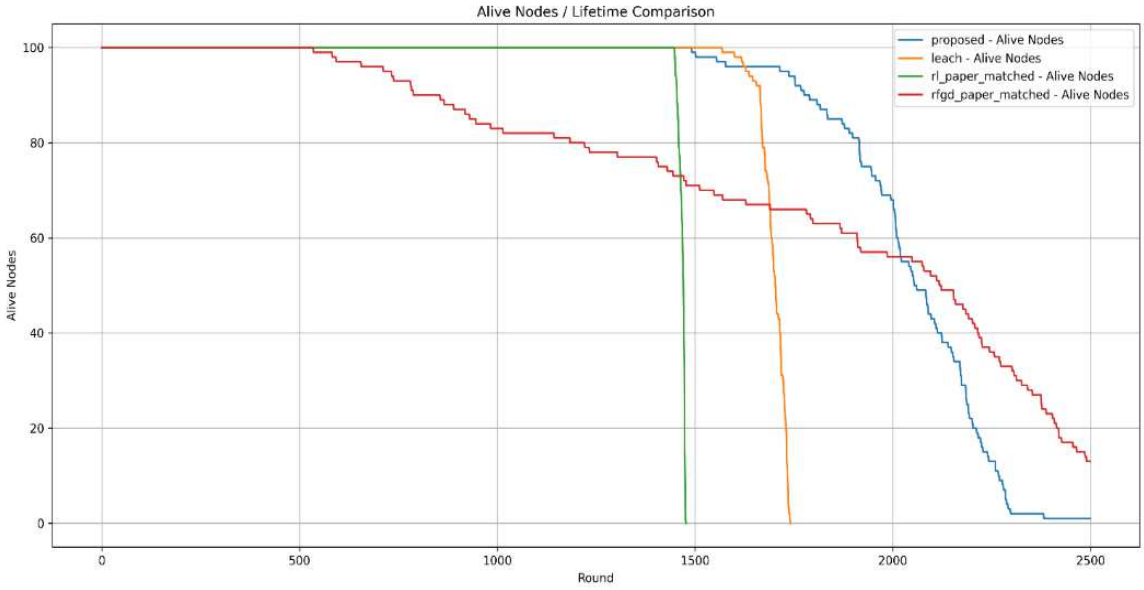


Fig. 1 Network Lifetime Comparison

7.2. Quality Score Evaluation

Fig. 2 shows that the proposed method maintains a favorable final quality score because the shared evaluator rewards balanced behavior. The proposed reward combines energy, trust, suspicion, attack status, node AoI, cluster AoI, centroid distance, and CH reuse. This enables more stable late-stage quality than LEACH and the RL baseline, which do not integrate the same freshness-security-energy trade-off.

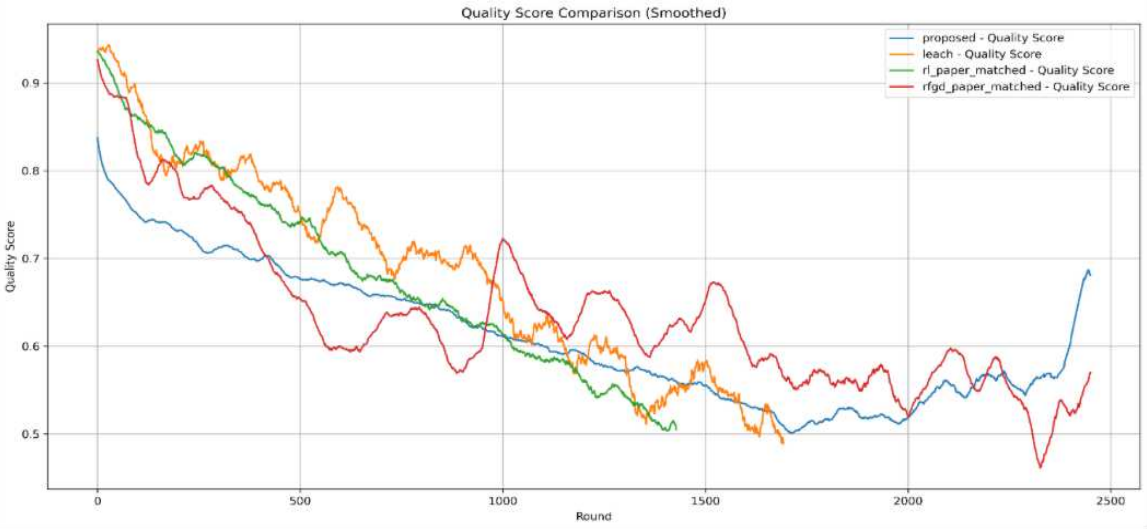


Fig. 2 Overall Quality Score Over Time

7.3. Data Freshness and AoI Performance

Fig. 3 is the main result of this work. When averaging twenty seeds in a 100-node scenario, the proposed method achieved the lowest final information lifetime (2.1543 ± 0.2389), compared with LEACH (6.9926 ± 0.1827), RL (18.4292 ± 1.5474), and RFGD (206.7771 ± 206.0699). This improvement is attributed to directly optimizing the AoI through the node AoI, the cluster AoI, and the maximum cluster AoI.

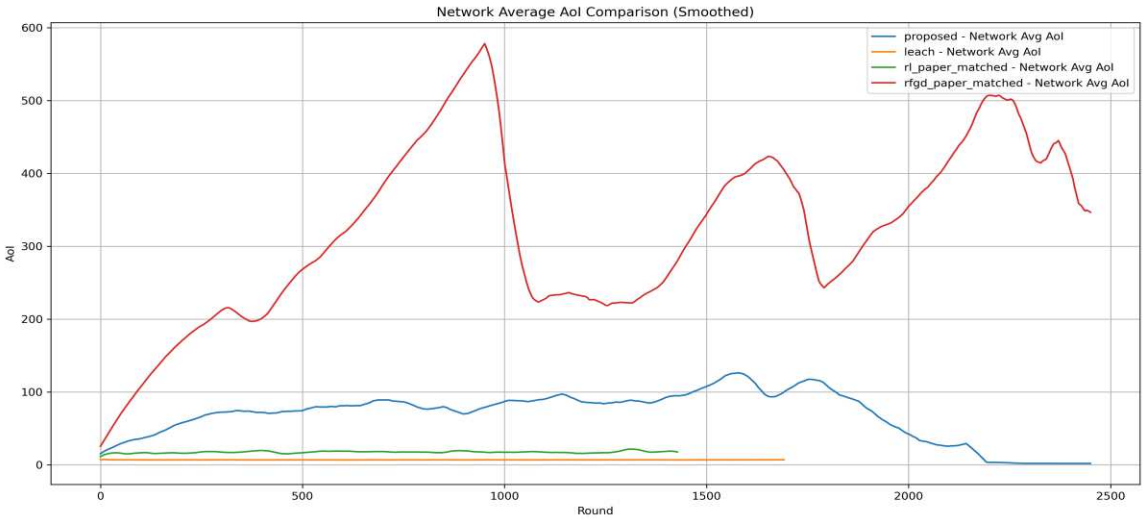


Fig. 3 Average Network Aoi Comparison

7.4. Trust Analysis

Fig. 4 reports the average trust of selected CHs. RFGD and RL can achieve higher trust values because their selection behavior is more directly trust- or topology-oriented. The proposed method balances trust with AoI, energy, suspicion, and spatial suitability. Therefore, it should not be interpreted as a pure trust maximization scheme; it is designed to avoid clearly unreliable CHs while preserving the main AoI-driven objective.

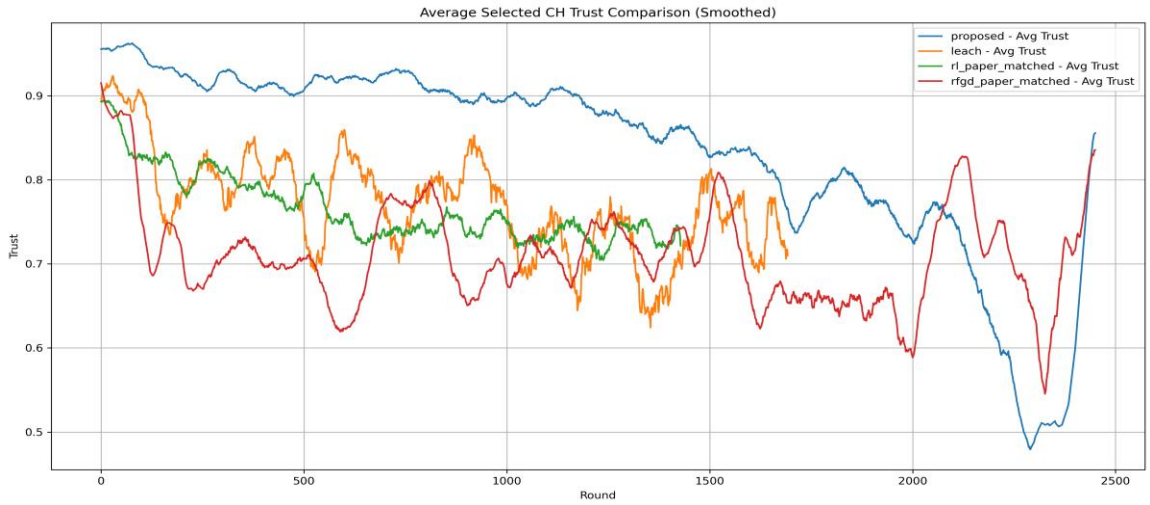


Fig. 4 Average Trust Level Comparison

7.5. Suspicion Analysis

Fig. 5 shows selected-CH suspicion. RFGD often produces lower suspicion due to fuzzy trust/alarm filtering, while the proposed method uses suspicion as one penalty among several competing objectives. In late rounds, some clusters may contain only a few surviving candidates, forcing the proposed method to choose the least harmful available CH rather than an ideal low-suspicion node. This is a practical trade-off rather than a contradiction in the design.

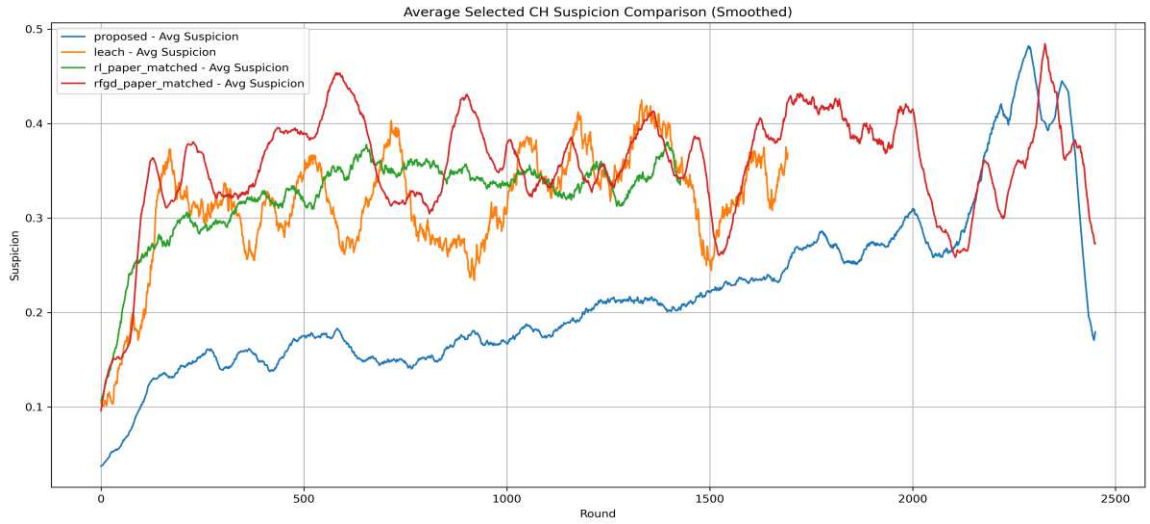


Fig. 5 Average Suspicion Level Comparison

7.6. Communication Delay

Fig. 6 presents delay. RFGD achieves favorable delay because its fuzzy decision process includes distance-related behavior and direct CH-to-sink communication. The proposed method improves delay over LEACH and the RL baseline by using K-Means clustering and centroid-distance control, but it does not optimize delay alone. The RL baseline exhibits higher latency, primarily because its multi-hop routing architecture adds extra hops and a higher setup cost. Since the proposed method, LEACH, and RFGD all use direct transmission from the channel head to the destination, while the RL baseline employs multi-hop routing, comparing the latency with the RL baseline specifically reflects this difference in architecture, not just the effect of channel head selection. Therefore, it should be considered a comparison between the two routing models, not evidence that the proposed channel head selection is inherently faster.

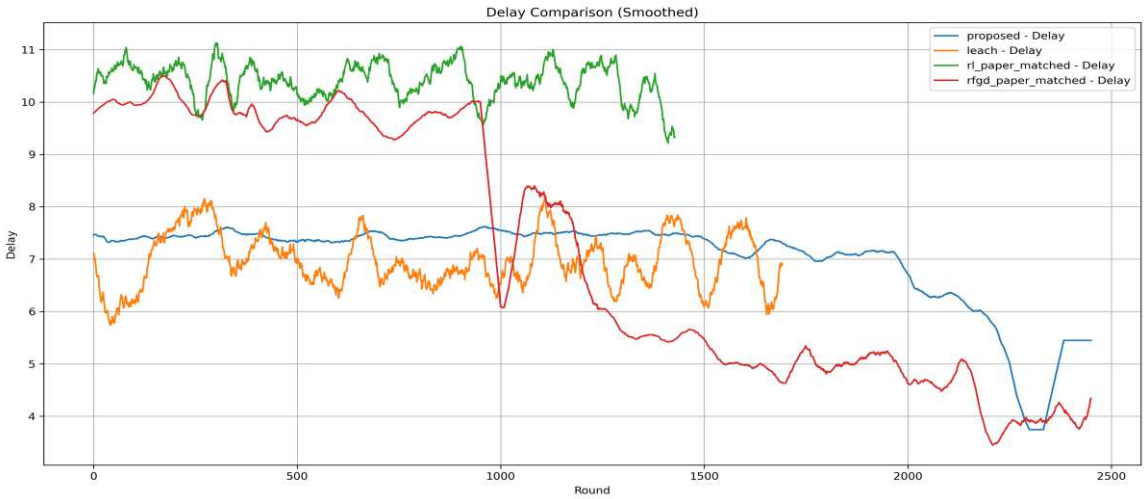


Fig. 6 Average Delay Comparison

7.7. Packet Delivery Ratio

Fig. 7 compares PDR. RFGD shows strong PDR because of its trust-oriented filtering and favorable direct transmission behavior. The proposed method remains competitive and improves over LEACH and the RL baseline by avoiding unstable or suspicious CHs. However, PDR must be interpreted together with AoI: a method may deliver packets reliably while still delivering stale information. As with delay, the PDR advantage over the RL baseline is confounded by the routing-model asymmetry described in Section 7.6, since multi-hop forwarding introduces additional loss opportunities that are independent of CH-selection quality.

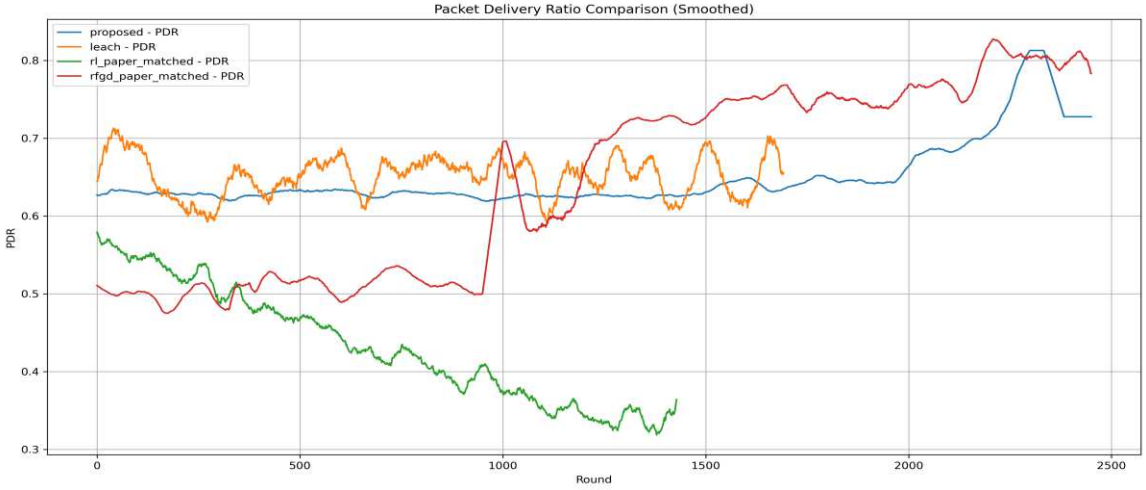


Fig. 7 Packet Delivery Ratio (PDR) Comparison

7.8. Energy Consumption

Fig. 8 shows energy consumption. The proposed method achieves very low late-stage energy consumption due to energy-aware rewards, centroid-distance penalization, CH reuse control, and candidate filtering. LEACH may form inefficient clusters due to random CH selection, while the RL baseline consumes more energy because multi-hop routing adds relay transmission and reception costs. RFGD can reduce total energy in some settings, but this is accompanied by a large AoI cost.

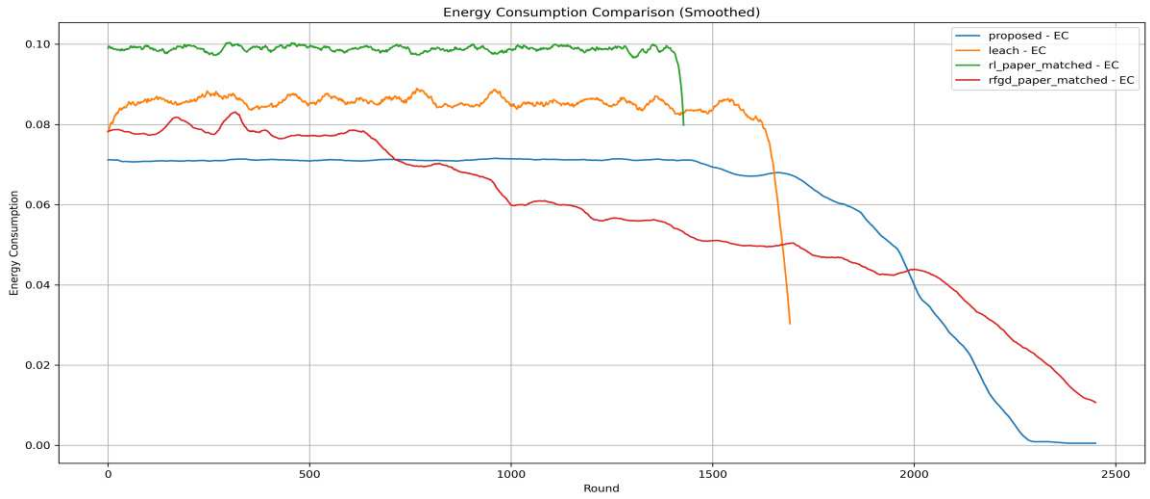


Fig. 8 Energy Consumption Over Time

7.9. CH Changes and Stability

Fig. 9 shows CH changes. The proposed method has low final CH changes, which indicates late-stage convergence of the learning process. This metric is reported carefully because very low CH changes can also occur when few viable nodes remain. Therefore, CH stability is interpreted together with final quality, AoI, and energy rather than treated as an isolated success indicator.

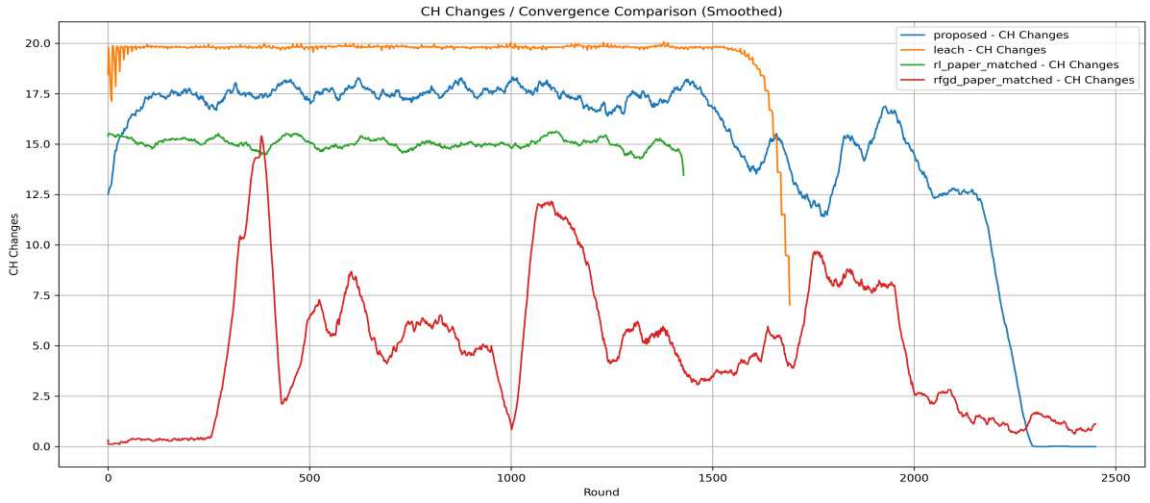


Fig. 9 Cluster-Head Changes Comparison

7.10. Routing Overhead

Fig. 10 reports route setup time. Only the RL-based baseline has a comparable routing setup value because it uses multi-hop CH routing. The proposed method, LEACH&RFGD use direct CH-to-sink communication in this simulation. This difference preserves the original logic of each baseline, but routing-overhead comparisons should be interpreted as architectural differences rather than direct CH-selection differences.

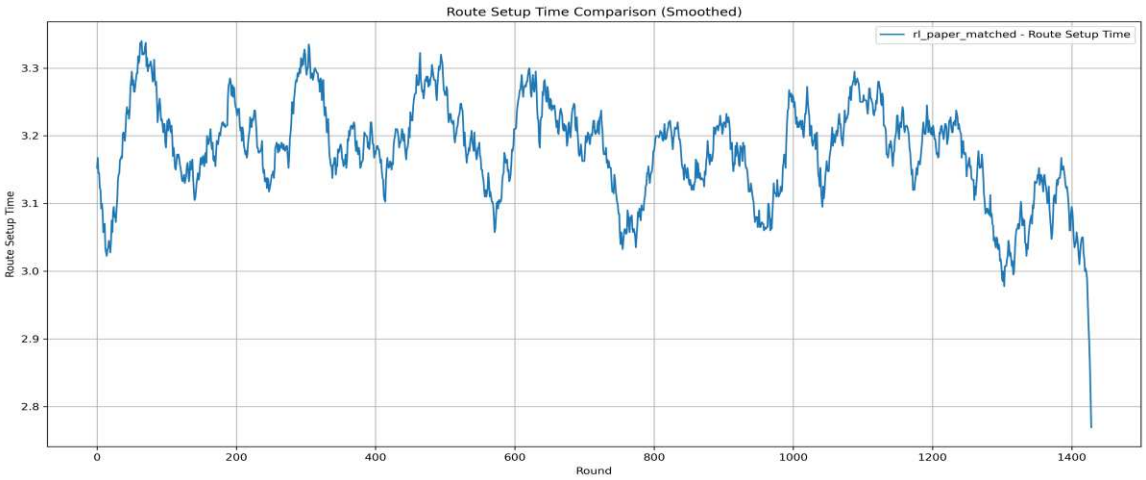


Fig. 10 Route Setup Time Comparison

7.11. Security Detection Performance

Fig. 11 shows that security detection metrics are close across methods because the same external evaluator is applied to all protocols. In a basic environment of 100 nodes, the proposed method achieved competitive accuracy (0.6228 ± 0.0140), balanced accuracy (0.6010 ± 0.0022), high accuracy (0.0896 ± 0.0010), and an F1 score of (0.1551 ± 0.0012), with the highest recall rate (0.5762 ± 0.0144). The basic reinforcement learning method scored slightly higher on the F1 score (0.1567 ± 0.0012), while the LEACH method achieved slightly higher accuracy (0.6383 ± 0.0148). Therefore, the proposed method is not presented as a standalone attack detector; rather, its contribution lies in utilizing anomaly and uncertainty information within the network header selection process while maintaining the primary objective of focusing on the area of interest.

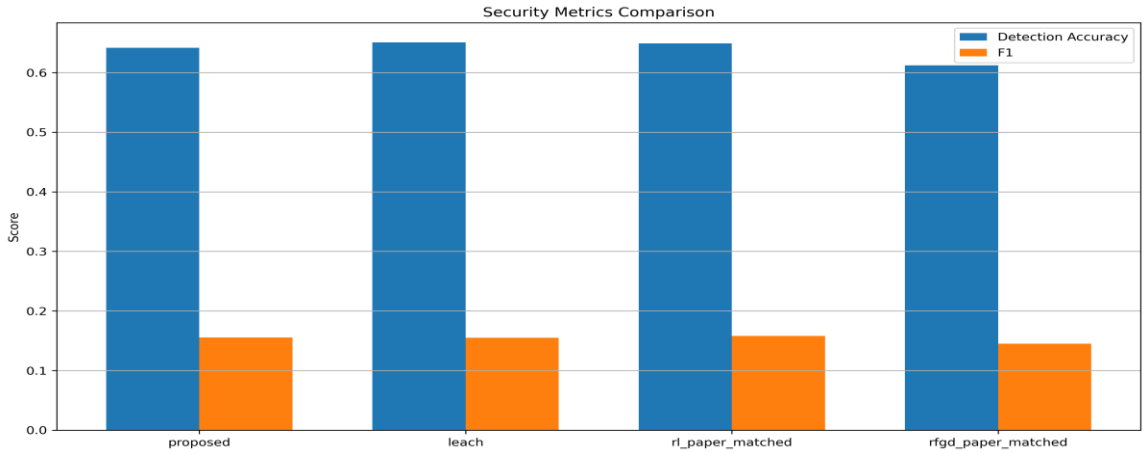


Fig. 11 Security Metrics Comparison

7.12. Strengths of the Proposed Method

The main strength of the proposed method is its AoI-driven integrated decision model. It combines energy awareness, AoI control, anomaly-driven suspicion, trust, attack status, spatial suitability, and CH reuse into one CH selection process. The twenty-seed robustness study confirms that this design achieves the lowest AoI and highest final quality across all nine tested configurations. A second strength is the fairness of evaluation: all baselines are kept close to their original philosophies and evaluated under the same network, attack, and security-evaluation pipeline. A third strength is reproducibility: the experiment explicitly reports the seed list, multi-scale settings, and mean ± standard deviation values rather than relying on a single favorable run.

7.13. Limitations

The main limitations are reported explicitly. First, the proposed method uses direct CH-to-sink transmission, while the RL baseline uses multi-hop routing; therefore, delay and route-overhead comparisons partly reflect architectural differences. Second, the attack model is probabilistic and synthetic, so real deployment validation is still required. Third, the proposed method is security-aware but not a pure attack detector; the RL baseline gives slightly higher F1-score, while RFGD often gives lower suspicion and higher

selected-CH trust. Fourth, when the candidate pool becomes small in late rounds, the method may prioritize AoI and continuity over selecting the lowest-suspicion CH. These limitations do not invalidate the contribution; rather, they define the next stage of routing-symmetric and real-deployment validation.

7.14. Statistical Significance Testing

To confirm that the observed differences between the proposed method and each baseline are not attributable to random seed variation, a paired Wilcoxon signed-rank test (two-sided, $\alpha = 0.05$) was applied across the twenty common random seeds for the primary 100-node, 2500-round scenario, since each seed produces a matched network/attack realization across all four methods. Table 4 reports the resulting p-values.

Table 4: Wilcoxon signed-rank test p-values (paired by random seed, N = 20) for the primary 100-node, 2500-round scenario

Metric	Proposed vs LEACH (p)	Proposed vs RL-based (p)	Proposed vs RFGD (p)
Final Quality Score	<0.001	0.006	0.522
Network AoI	<0.001	<0.001	<0.001
Trust	0.002	0.002	0.006
Suspicion	0.312	0.070	0.076
Delay	<0.001	<0.001	0.053
PDR	<0.001	<0.001	0.053
Final Avg Energy Consumption	<0.001	<0.001	<0.001
F1-Score	0.033	0.001	<0.001

The AoI and final average energy consumption advantages of the proposed method are statistically significant against all three baselines ($p < 0.001$). The final quality score advantage is significant against LEACH and the RL-based baseline ($p < 0.01$) but not against RFGD ($p = 0.522$), confirming that RFGD's quality score is a genuine competitor on this metric even though its AoI is far worse. The lower selected-CH trust of the proposed method is statistically significant ($p < 0.01$ against all baselines), which reinforces the interpretation in Section 7.4 that the proposed method deliberately trades some trust margin for freshness and energy gains rather than this being a random artifact. Suspicion differences are not statistically significant ($p > 0.05$), indicating that the modest increase in selected-CH suspicion relative to the baselines is within the range of seed-to-seed variability.

Regarding the large variance of RFGD's network AoI reported in Table 5 (206.78 ± 206.07), per-seed inspection confirms this is not a small number of catastrophic outliers but a genuinely right-skewed distribution: all twenty seeds produced elevated AoI values (range 38.8–918.0, median 142.4, IQR 147.2), because RFGD's fuzzy trust-based rotation contains no explicit mechanism to bound information staleness. By comparison, the proposed method's AoI distribution is tight and low across seeds (median 2.04, IQR 0.12). For this reason, median and IQR values for network AoI are reported alongside the mean and standard deviation in the revised Table 5, since the mean is a less representative summary statistic under this degree of skew.

8. Comparative Analysis

Table 5: Quantitative comparison for the 100-node scenario (mean ± standard deviation across twenty random seeds)

Metric	LEACH	RL-Based	RFGD	Proposed
Final Quality Score	0.4984 ± 0.0190	0.5085 ± 0.0139	0.5473 ± 0.0616	0.5673 ± 0.0740
Network AoI	6.9926 ± 0.1827 (med 7.00)	18.4292 ± 1.5474 (med 18.07)	206.7771 ± 206.0699 (med 142.43, IQR 147.15)	2.1543 ± 0.2389 (med 2.04)
Trust	0.7199 ± 0.0395	0.7291 ± 0.0306	0.7425 ± 0.1375	0.5462 ± 0.2088
Suspicion	0.3627 ± 0.0302	0.3410 ± 0.0273	0.3163 ± 0.1169	0.4058 ± 0.1539
Delay	6.8981 ± 0.6581	10.4821 ± 0.6198	4.5203 ± 0.9197	5.0302 ± 1.1544
PDR	0.6551 ± 0.0329	0.3303 ± 0.0175	0.7740 ± 0.0460	0.7485 ± 0.0577
Total Energy Consumption	145.2228 ± 0.0280	145.1474 ± 0.0357	142.6128 ± 1.2638	144.9537 ± 0.1214
Final Avg EC	0.0363 ± 0.0030	0.0829 ± 0.0027	0.0096 ± 0.0038	0.0012 ± 0.0006
Alive Nodes at End	0.0000 ± 0.0000	0.0000 ± 0.0000	12.3500 ± 5.5182	1.5500 ± 1.0990
Alive Ratio at End	0.0000 ± 0.0000	0.0000 ± 0.0000	0.1235 ± 0.0552	0.0155 ± 0.0110

First Node Dead Round	1519.55 ± 59.19	1439.95 ± 41.13	541.90 ± 64.53	1579.90 ± 67.94
Half Nodes Dead Round	1703.70 ± 28.31	1463.90 ± 23.94	2056.00 ± 47.17	2046.85 ± 24.40
Last Node Dead Round	1733.75 ± 29.57	1474.05 ± 23.89	N/A	2447.00 ± 35.16
F1-Score	0.1559 ± 0.0014	0.1567 ± 0.0012	0.1480 ± 0.0013	0.1551 ± 0.0012

Table 5 summarizes the primary 100-node, 2500-round results averaged over twenty random seeds. These values correspond to the scenario used for the figures, ensuring that the visual plots and numerical comparison describe the same fair baseline setting. The table shows that the proposed method gives the lowest AoI, highest final quality, and lowest final average energy consumption, while RFGD gives higher selected-CH trust, lower suspicion, higher PDR, and more surviving nodes at the end. This distinction is important because the proposed method is primarily an AoI-aware clustering contribution rather than a pure trust maximization method.

Tables 6-8 extend the interpretation beyond the primary figure scenario. Table 6 aggregates the mean values over all nine robustness configurations. Table 7 reports how often each method gives the favorable value for each metric across the nine configurations defined in Section 6 (three network sizes × three simulation lengths). Table 8 reports the detailed security-evaluator metrics for the primary 100-node setting. Together, these tables show that the proposed method is consistently favorable in AoI and final quality, while security metrics remain close across methods because the evaluator is shared.

Table 6: Overall robustness summary across all tested node counts, simulation lengths, and twenty random seeds

Metric	LEACH	RL-Based	RFGD	Proposed
Final Quality Score	0.5059	0.5092	0.5167	0.5716
Network AoI	6.9147	36.3469	198.7336	2.8585
Trust	0.7297	0.7344	0.7204	0.5972
Suspicion	0.3478	0.3424	0.3414	0.4060
Delay	6.8710	10.3601	5.5545	5.5259
PDR	0.6565	0.3253	0.7223	0.7237
Final Avg EC	0.0392	0.1218	0.0086	0.0014
Alive Ratio at End	0.0000	0.0000	0.0488	0.0060
Half Nodes Dead Round	1722.40	1542.97	1819.26	2059.65
F1-Score	0.1558	0.1562	0.1480	0.1548

Table 6 aggregates the mean values over all nine robustness configurations. The proposed method preserves the lowest AoI, highest final quality, lowest final energy consumption, and latest half-node-dead round on average, confirming that the main contribution is AoI-aware balanced clustering rather than isolated security-score maximization.

Table 7: Consistency count across nine robustness configurations

Metric	Preferred Direction	LEACH	RL-Based	RFGD	Proposed
Final Quality Score	Higher	0/9	0/9	0/9	9/9
Network AoI	Lower	0/9	0/9	0/9	9/9
Trust	Higher	3/9	3/9	3/9	0/9
Suspicion	Lower	3/9	3/9	3/9	0/9
Delay	Lower	0/9	0/9	3/9	6/9
PDR	Higher	0/9	0/9	3/9	6/9
Final Avg EC	Lower	0/9	0/9	1/9	8/9
Alive Ratio at End	Higher	6/9	6/9	9/9	6/9
Half Nodes Dead Round	Higher	0/9	0/9	3/9	6/9
F1-Score	Higher	3/9	6/9	0/9	0/9

Note: each number is a win count out of the nine configurations defined in Section 6 (three network sizes × three simulation lengths), not the metric's actual value. A 0/9 for Trust or Suspicion does not mean the proposed method has zero trust or is insecure; it means it was not the single highest-trust (or lowest-suspicion) method in any configuration. Its real trust value is moderate, around 0.55-0.57 out of 1.0, versus 0.72-0.74 for the other three methods (see Table 5).

Table 7 shows that the proposed method is consistently favorable for AoI and final quality across all robustness configurations. RFGD and RL remain useful baselines for individual metrics such as trust, suspicion, PDR, delay, or F1-score, but they do not provide the same freshness-quality-energy balance. The 0/9 counts for trust and suspicion reflect a consistent but moderate gap rather than an absence of security awareness: as Table 5 shows, the proposed method's selected-CH trust remains in a comparable numeric

range to the baselines throughout, and the method never claims to maximize trust in isolation, but to balance it against freshness and energy under the reward design described in Section 5.5.

Table 8: Detailed security evaluator metrics for the 100-node, 2500-round setting (mean ± standard deviation across twenty random seeds)

Security Metric	LEACH	RL-Based	RFGD	Proposed
Detection Accuracy	0.6383 ± 0.0148	0.6380 ± 0.0130	0.6149 ± 0.0127	0.6228 ± 0.0140
Balanced Accuracy	0.6000 ± 0.0020	0.6014 ± 0.0016	0.5878 ± 0.0015	0.6010 ± 0.0022
Precision	0.0906 ± 0.0011	0.0911 ± 0.0010	0.0853 ± 0.0010	0.0896 ± 0.0010
Recall	0.5565 ± 0.0180	0.5598 ± 0.0144	0.5571 ± 0.0128	0.5762 ± 0.0144
F1-Score	0.1559 ± 0.0014	0.1567 ± 0.0012	0.1480 ± 0.0013	0.1551 ± 0.0012
Best Threshold	0.4925 ± 0.0183	0.5000 ± 0.0000	0.4300 ± 0.0251	0.5000 ± 0.0000
Best Threshold F1	0.1560 ± 0.0014	0.1567 ± 0.0012	0.1511 ± 0.0014	0.1551 ± 0.0012

Table 8 shows that the proposed method achieves the highest recall, while the RL baseline has a slightly higher F1-score and LEACH has slightly higher accuracy. This supports the interpretation that the proposed method is security-aware in CH selection but is not a standalone attack detector optimized only for detection metrics.

Table 9: Conceptual differences among the evaluated methods

Feature	LEACH	RL-Based	RFGD	Proposed
CH selection principle	Probabilistic threshold and rotation	RL/SARSA-like global CH selection	Fuzzy chance value + trust filtering	K-Means + filtering + Q-learning reward
Energy-aware	Indirect rotation only	Yes, through ranking and reward	Yes, fuzzy energy membership	Yes, explicit reward and filtering
Trust-aware	No	No explicit trust term	Yes	Yes
Suspicion/anomaly-aware	No	No explicit anomaly term	Partially	Yes, temporal anomaly and risk memory
Attack-aware	No	No	Indirectly	Yes, attack status is penalized
AoI-aware	No	No	No explicit AoI control	Yes, node AoI, cluster average AoI, and cluster maximum AoI
Learning-based	No	Yes	No, fixed fuzzy rules	Yes, cluster-level Q-learning
Spatial awareness	Nearest CH association	Centrality, degree, sink distance	Distance/radius fuzzy inputs	K-Means and centroid distance
Load balancing	Random rotation	CH count in ranking	Limited	CH-count penalty
Routing model	Direct CH-to-sink	Multi-hop CH routing	Direct CH-to-sink	Direct CH-to-sink
Main strength	Simplicity	Learning and multi-hop routing	High PDR and low delay	Balanced freshness, security awareness, quality, and stability
Main limitation	No security/AoI awareness	High delay and unstable CH changes	Very high AoI	Direct CH-to-sink routing (vs. multi-hop in RL baseline) — delay, PDR, and CH-change comparisons with RL are therefore subject to routing-layer asymmetry and should be interpreted with caution†; selected-CH trust/suspicion vary considerably across random seeds

Table 9 summarizes the conceptual differences among the evaluated methods in terms of CH selection principle, awareness dimensions, learning basis, routing model, and main limitations.

9. Conclusion

This paper presented a multi-objective intelligent clustering framework for WSNs that unifies reinforcement learning, AoI-awareness, and anomaly-driven security into one CH selection strategy, combining K-Means clustering with a Q-learning reward that considers residual energy, trust, suspicion, AoI, attack status, distance, and CH fairness. Averaged over twenty random seeds,

the main contribution is data freshness: in the primary 100-node, 2500-round scenario, the proposed method achieved the lowest final AoI (2.1543 ± 0.2389, versus 6.9926 for LEACH, 18.4292 for RL, and 206.7771 for RFGD), the highest final quality score (0.5673 ± 0.0740), and the lowest final average energy consumption (0.0012 ± 0.0006). It remained favorable for AoI and quality across all nine robustness configurations, confirming the improvement generalizes beyond a single seed or network setting. Compared with LEACH, the method replaces random CH selection with freshness-, anomaly-, trust-, and energy-aware decisions. Compared with the RL baseline, it substantially reduces AoI and improves lifetime, though delay and overhead comparisons are partly confounded by RL's multi-hop routing. Compared with RFGD, it avoids severe staleness by explicitly optimizing AoI, while RFGD remains favorable in trust, suspicion, PDR, and survival a balanced trade-off rather than a uniform win. This confirms that tactical WSN clustering should not be evaluated on a single metric: a method preserving trust or PDR while allowing stale information may be unsuitable for freshness-critical monitoring. The proposed method addresses this by making AoI central to CH selection while retaining energy- and security-awareness.

10. *Future Work*

Future work will extend the proposed framework in several directions. First, deep reinforcement learning will be explored to handle larger continuous state spaces and adaptive reward-weight optimization across network densities. Second, extending the current direct-sink communication into a security-aware multi-hop routing architecture is a priority to enable fair comparisons with ad-hoc baselines. Third, integration with edge computing and digital-twin paradigms will offload anomaly detection from resource-constrained nodes. Fourth, cost-sensitive learning and robust metrics such as PR-AUC will address class imbalance in attack detection, validated on real hardware testbeds. Finally, node mobility — including mobile ground units and UAV relays, with the Q-learning reward adapted to fast-fading channels and dynamic topology — will extend the framework toward practical tactical-grade deployments beyond the current static-node validation.

Declarations

Funding

This work was supported by the Science, Technology & Innovation Funding Authority (STDF) in cooperation with the Egyptian Knowledge Bank (EKB).

Competing Interests

The authors declare that they have no competing interests.

Ethics Approval and Consent to Participate

Not applicable. This study did not involve human participants, human data, or animals.

Consent for Publication

Not applicable.

Data Availability

The simulation code, raw per-seed results, and configuration files used to generate the results in this paper are available from the corresponding author on reasonable request.

Materials Availability

Not applicable.

Code Availability

The simulation code used to generate the results in this paper is available from the corresponding author on reasonable request.

Author Contributions

[Authors contributed equally.]

References

1. I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, Wireless sensor networks: A survey, Computer Networks, vol. 38, no. 4, pp. 393-422, 2002.
2. J. Yick, B. Mukherjee, and D. Ghosal, Wireless sensor network survey, Computer Networks, vol. 52, no. 12, pp. 2292-2330, 2008.
3. W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in Proc. 33rd Annual Hawaii International Conference on System Sciences (HICSS), Maui, HI, USA, 2000, pp. 1-10.
4. W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks, IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660-670, Oct. 2002.
5. J. MacQueen, Some methods for classification and analysis of multivariate observations, in Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 1967, pp. 281-297.
6. C. J. C. H. Watkins and P. Dayan, Q-learning, Machine Learning, vol. 8, no. 3-4, pp. 279-292, 1992.
7. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
8. S. Kaul, R. Yates, and M. Gruteser, Real-time status: How often should one update? in Proc. IEEE INFOCOM, Orlando, FL, USA, 2012, pp. 2731-2735.
9. R. D. Yates and S. K. Kaul, The Age of Information: Real-time status updating bymultiple sources, IEEE Transactions on Information Theory, vol. 65, no. 3, pp. 1807-1827, Mar. 2019.
10. W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, Edge computing: Vision and challenges, IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637-646, Oct. 2016.
11. F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, Digital twin in industry: State-of-the-art, IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2405-2415, Apr. 2019.
12. Q. Qi and F. Tao, A smart manufacturing service system based on edge computing, fog computing, and cloud computing, IEEE Access, vol. 7, pp. 86769-86777, 2019.
13. Q. Qi, F. Tao, Y. Zuo, and D. Zhao, Digital twin service towards smart manufacturing, Procedia CIRP, vol. 72, pp. 237-242, 2018.
14. F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, Digital twin driven prognostics and health management for complex equipment, CIRP Annals, vol. 67, no. 1, pp. 169-172, 2018.
15. M. Chiang and T. Zhang, Fog and IoT: An overview of research opportunities, IEEE Internet of Things Journal, vol. 3, no. 6, pp. 854-864, Dec. 2016.
16. S. S. Mouli, T. V. Talagondapati, and M. P. Singh, Multi-faceted reinforcement learning frameworks for dynamic cluster head selection and energy-efficient routing in wireless sensor networks, SN Computer Science, vol. 6, no. 5, Art. no. 497, 2025, doi: 10.1007/s42979-025-04031-z.
17. M. Dhipa, N. Sampath, R. Seetha, and D. Nageswari, Secure cluster head selection in wireless sensor networks using fuzzy based generic algorithm, Wireless Networks, vol. 31, no. 2, pp. 1435-1448, 2025, doi: 10.1007/s11276-024-03828-9.
18. O. Younis and S. Fahmy, HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks, IEEE Transactions on Mobile Computing, vol. 3, no. 4, pp. 366-379, Oct.-Dec. 2004.
19. S. Lindsey and C. S. Raghavendra, PEGASIS: Power-efficient gathering in sensor information systems, in Proc. IEEE Aerospace Conference, Big Sky, MT, USA, 2002, pp. 1125-1130.
20. A. Manjeshwar and D. P. Agrawal, TEEN: A routing protocol for enhanced efficiency in wireless sensor networks, in Proc. 15th International Parallel and Distributed Processing Symposium Workshops (IPDPSW), San Francisco, CA, USA, 2001, pp. 2009-2015.
21. A. Manjeshwar and D. P. Agrawal, APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks, in Proc. 16th International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Fort Lauderdale, FL, USA, 2002, pp. 195-202.
22. A. A. Abbasi and M. Younis, A survey on clustering algorithms for wireless sensor networks, Computer Communications, vol. 30, no. 14-15, pp. 2826-2841, 2007.
23. X. Liu, A survey on clustering routing protocols in wireless sensor networks, Sensors, vol. 12, no. 8, pp. 11113-11153, 2012.
24. Y. Liu, W. Guo, and H. Fan, A game theory based energy efficient clustering protocol for wireless sensor networks, Wireless Communications and Mobile Computing, vol. 2019, Art. no. 9048607, 2019, doi: 10.1155/2019/9048607.
25. M. Zhu, Q. Liu, and X. Wang, Multi-agent reinforcement learning-based energy-efficient routing for wireless sensor networks, IEEE Access, vol. 9, pp. 47996–48007, 2021, doi: 10.1109/ACCESS.2021.3068963.

26. L. Yang, Y. Lu, S. X. Yang, T. Guo, and Z. Liang, A secure clustering protocol with fuzzy trust evaluation and outlier detection for industrial wireless sensor networks, IEEE Transactions on Industrial Informatics, vol. 17, no. 7, pp. 4837–4847, Jul. 2021, doi: 10.1109/TII.2020.3019286.
27. Y.-D. Yao, X. Li, Y.-P. Cui, J.-J. Wang, and C. Wang, Energy-efficient routing protocol based on multi-threshold segmentation in wireless sensor networks for precision agriculture, IEEE Sensors Journal, vol. 22, no. 7, pp. 6216–6231, Apr. 2022, doi: 10.1109/JSEN.2021.3131260.
28. B. Han, Y. Zhu, Z. Jiang, Y. Hu, and H. D. Schotten, Optimal blocklength allocation towards reduced Age of Information in wireless sensor networks, in Proc. IEEE GLOBECOM Workshops (GC Wkshps), Waikoloa, HI, USA, Dec. 2019, pp. 1–6, doi: 10.1109/GCWkshps45667.2019.9024516.
29. S. Gong, L. Cui, B. Gu, B. Lyu, D. T. Hoang, and D. Niyato, Hierarchical deep reinforcement learning for Age-of-Information minimization in IRS-aided and wireless-powered wireless networks, IEEE Transactions on Wireless Communications, vol. 22, no. 11, pp. 7437–7451, Nov. 2023, doi: 10.1109/TWC.2023.3259721.
30. R. D. Yates, Age of information in a network of preemptive servers, in Proc. IEEE INFOCOM Workshops, Honolulu, HI, USA, Apr. 2018, pp. 697–702, doi: 10.1109/INFCOMW.2018.8406983.
31. M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, Machine learning in wireless sensor networks: Algorithms, strategies, and applications, IEEE Communications Surveys&Tutorials, vol. 16, no. 4, pp. 1996-2018, Fourth Quarter 2014.