

Secure and Efficient Data Aggregation in Wireless Sensor Networks: A Review, Research Gaps, and a Proposed Lightweight Framework

Dr. Amit Singhal¹, Dr. Sai Kiran Oruganti²

¹Professor, Department of CSE, Raj Kumar Goel Institute of Technology, Ghaziabad, India. amit1408@gmail.com

²Faculty of Engineering and Built Science, Lincoln University College, Malaysia. saisharma@lincoln.edu.my

Abstract— Wireless Sensor Networks (WSN) consist of a large number of tiny and constrained sensor nodes to sense the physical or environmental phenomena and relay their observations towards a base station (BS). Most deployments amplify and correlate readings in the network before sending them all upstream to consume node energy and congest the network. However, aggregation focuses trust into a few nodes, where one malicious or compromised node can contribute false partial sums, replay old packets, or steal information carried in the sensitive readings via offline corruptions, without being detected by the base station and thus without any impact on the value that this base station is taking action on. In this paper, we examine the status of secure data aggregation protocols for WSNs, as well as discuss their reported advantages and disadvantages, and list the open challenges: scalability, energy efficiency, security coverage, deployability in the real world, additional emerging technologies (such as blockchain, machine learning) and resource-constrained operations. These gaps provide the basis for a Secure Lightweight Hybrid Data Aggregation (SLHDA) framework in which the base station securely learns about the secured clustering, performs aggregated in-network data processing and adds integrity guarantees for the data via keyed-hash components, filters for corruption in the network using a trust-based approach that leaves the auditing record to the backbone of cluster-head-to-base station, and implements an audit ledger using a lightweight blockchain in the cluster. We also describe a first-order radio energy model-based simulation approach and metrics for assessing the framework in comparison to current baselines – energy consumption, network lifetime, aggregation accuracy, detection rate and end-to-end latency. We argue that a hybrid layered design can provide an acceptable compromise for the security level desired with respect to the energy and computation resources of the WSN nodes.

Index Terms— *Wireless sensor networks, secure data aggregation, energy efficiency, privacy preservation, trust management, blockchain, Internet of Things.*

I. INTRODUCTION

A Wireless Sensor Network (WSN) is an ad hoc network composed of numerous tiny battery-powered sensor nodes that are designed to collaborate to monitor, record and transmit a physical phenomenon, such as humidity, temperature, and seismic events, or pollution concentration [1]. Today, such networks are embedded in a variety of Internet of Things (IoT) applications such as smart agriculture, structural health monitoring, industrial automation, environmental monitoring and remote healthcare, as cited in [1, 2]. All of them have limited memory, computation, and communication bandwidth at the nodes, and batteries that cannot be changed once the network is deployed on a dangerous terrain or a distant area. Long network lifetime is thus not an exigency or an add-on to the design; it is a first order design concern.

The EUS4All project draws on existing data; that is, SIW-W readings that are closely located in space and/or time tend to be correlated, and therefore it would be wasteful to send each individual SIW-W reading to the base station. Direct redundancy attack on data aggregation: intermediate nodes, typically cluster heads (CHs) coalesce the readings received and just send summarized data to the sink [1, 7]. Summing, averaging, and taking min/max or counts reduce the amount of bits sent by an order of magnitude, and the resulting energy savings are a side benefit, as is a lower probability of congestion.

However, there are catch effects to these gains. Seeding data to CHs does also lead to centralization of risk: if a CH is compromised, or if a malicious node is introduced in the aggregation path, false partial aggregates, selectively

dropping packets, replaying of old messages or eavesdropping on plaintext readings can take place to lead to any compromise of the confidentiality, integrity, and availability of the aggregated result [5], [6]. However, cryptographic protection come with a cost too: It needs to maintain the same energy budget as aggregation does, but also demands the computations and additional messages that come with it. It is an open and ongoing problem to design a secure, energy-efficient and practicable aggregation scheme for large-scale deployments with limited resources, simultaneously.

The rest of the paper is presented as follows. Second, representative secure data aggregators techniques reviewed and compared from recent literatures by analysing in Section II. Given the problem statement is formalized in Section III and the motivation and research objectives in Section IV. In Section V, the proposed methodology – a system model, the SLHDA framework, the algorithmic steps and the simulation-based evaluation plan – are described. The remaining knowledge gaps and open research challenges are discussed in Section VI. Section VII gives the outcomes and Section VIII summarizes the paper with the suggestions for future work.

II. RELATED WORK

Whilst Table I gives a brief overview of representative recent research for secure and energy-efficient data aggregation in WSN and IoT applications. Each contribution specifies technique, its principal advantage and its main limitation as reported in the literature.

There are two overall design themes that permeate these contributions. The first follows the strategy of reduction via topology-aware clustering and rendering scheduling with energy and scalability trade-offs, typically at the expense of the absence or lack of an adversarial model [2, 7]. The second minimizes computational/communication costs but still provides cryptographic or blockchain-based security guarantees [3, 9, 10]. It is this tension that drives the hybrid design developed in Section V: move the heavier security mechanisms to the higher tiers in the cluster where cluster-head and gateway are more well provisioned, and where they are relatively abundant; move the “small” portions of the cluster load closer to the leaf-node tier.

TABLE I
COMPARATIVE SUMMARY OF SECURE DATA AGGREGATION TECHNIQUES

Year	Author(s)	Technique	Advantage(s)	Limitation(s)
2021	Yousefpoor et al. [5]	Comprehensive review of secure aggregation countermeasures	Broad taxonomy of attacks and defenses across cryptographic and non-cryptographic schemes	Primarily analytical; does not propose a new lightweight scheme
2021	Yousefpoor et al. [6]	Hierarchical secure aggregation with dragonfly algorithm	Fuzzy-scheduled intra-cluster aggregation and optimized inter-cluster tree	Optimization overhead grows with cluster count
2022	Begum & Nandury [2]	Energy-efficient aggregation protocol survey	Improves network lifetime through topology-aware scheduling	Limited native security integration
2022	Tao et al.	EPPSA privacy-preserving statistical aggregation	Supports multiple aggregation functions with formal privacy guarantees	Higher cryptographic (homomorphic) computation cost
2023	Begum & Nandury [1]	Comprehensive WSN–IoT aggregation survey/taxonomy	Extensive classification of aggregation protocols and optimization goals	Survey in nature; no protocol

Year	Author(s)	Technique	Advantage(s)	Limitation(s)
				implementation or evaluation
2024	Aga et al. [4]	Secure and private IoT data aggregation techniques	Preserves privacy of aggregated readings	Increased communication overhead from added protocol messages
2026	MCAGCN-BBO	Blockchain with graph convolutional network optimization	High security and adaptive resource optimization	Computational complexity unsuitable for the most resource-constrained nodes

III. PROBLEM STATEMENT

The individual sensor nodes have limited resources to spare (i.e., energy and computing resources), and adopt data aggregation to reduce communication overhead and extend the lifetime of the network. But this process is jeopardized by malicious nodes or compromised nodes which threaten the integrity and reliability of the process and may cause transposed or bogus aggregations to be sent to the base station [5].

Clearly, such an aggregation mechanism must be both lightweight and secure, while maintaining a low energy consumption as well, to ensure data confidentiality and integrity, resistance to such attacks as false-data injection, selective forwarding, and replay, and be scalable and practical for large-scale real-world deployments of WSNs and IoTs.

IV. MOTIVATION AND OBJECTIVES

The motivation in this work is simple: a lightweight, secure, scalable and low-energy consuming data aggregation framework that shields aggregated data from malicious attacks and that prolongs the operation time of WSNs. Practical problems that motivate this research along with operational impact are listed in Table II.

TABLE II
PROBLEMS AND THEIR OPERATIONAL IMPACT

Problem	Operational Impact
High energy consumption	Short network lifetime
Massive data transmission	Network congestion
Security attacks on aggregators	Data manipulation
Node compromise	False aggregated results
Large-scale WSN deployment	Scalability issues
IoT integration	Need for lightweight security

A. Research Objectives

- 1) To design a clustering and aggregation scheme that keeps per-node energy consumption low without sacrificing aggregation accuracy.
- 2) To add lightweight cryptographic confidentiality and keyed-hash integrity verification within the computational budget of typical sensor hardware.
- 3) To incorporate a trust-based mechanism that can detect and isolate compromised or malicious nodes during the aggregation process.
- 4) To quantify the trade-off between security strength and network lifetime through simulation against established baseline protocols.

V. PROPOSED METHODOLOGY

This section provides the system model, the proposed Secure Lightweight Hybrid Data Aggregation (SLHDA) framework, its algorithmic workflow, and the methodology to be used for simulation-based evaluation of the framework.

A. Network and Adversary Model

The sensing field is a two-dimensional space and we assume that there are N stationary sensor nodes deployed in it in a uniform and random manner and clustered based on a modified low energy adaptive cluster hierarchy where the sensor's residual energy is used to determine its cluster. Each cluster chooses one CH who gathers data from all of its nodes and forwards it towards the base station, either directly or by passing it through a series of CHs. The base station is rich in resources and reliable. The adversary can inflict a bounded degree of damage on sensor nodes or CHs, securing enough resources to enable it to maliciously modify the data, select purposefully forward particular packets, and be able to eavesdrop on unprotected payload data, but has no ability to compromise the base station itself.

B. Proposed SLHDA Framework

SLHDA overlay these four cooperative mechanisms on top of typical cluster-oriented aggregation: (i) Residual-energy-aware cluster-head (CH) election to balance energy usage across clusters and promote longer lifetimes: (ii) An additive-homomorphic, lightweight encryption scheme is added to the member-to-CH and CH-to-CH hops to detect any sum in the middle, without the need for end-to-end decryption; and (iii) A keyed-hash message authentication code (MAC) is placed on each aggregated packet allowing consistency of aggregates to be checked as an aggregated packet moves through its cluster toward the CH, which allows the detection and removal of nodes that are misreporting or modifying aggregates in a manner that is inconsistent with their history of reported aggregates. A light-weight blockchain ledger that is designed to complete the design. It is computed within the CH-to-base-station backbone, not all sensor nodes and it is storing hashes of the aggregation rounds, providing energy efficient tamper-evident auditing without consensus overhead for energy constrained leaf nodes.

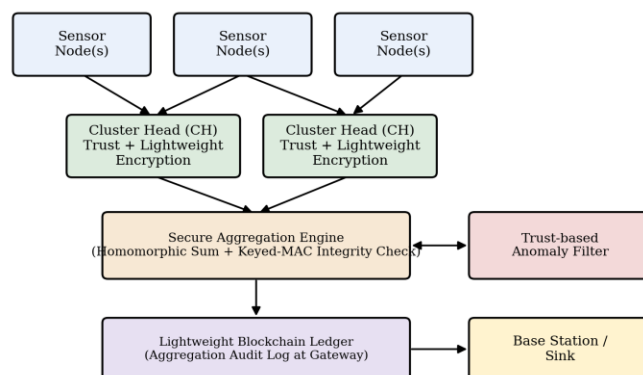


Fig. 1. System architecture of the proposed SLHDA framework.

C. Algorithmic Workflow

- 1) Cluster formation: Nodes are grouped into clusters, and CH election weights residual energy and node centrality to balance load.
- 2) Secure sensing and encryption: Each member node encrypts its reading with the shared homomorphic key, appends a keyed-hash tag, and transmits the result to its CH.
- 3) In-network aggregation: The CH verifies each incoming MAC, discards packets that fail verification, and homomorphically combines the remaining valid ciphertexts into a partial aggregate.
- 4) Trust update: Each CH's aggregate history is compared against its reputation profile at the base station or an upstream CH, and large deviations lower its trust score.
- 5) Anomaly isolation: A CH whose trust score falls below the threshold is excluded from the active aggregation path for a cooldown interval, then re-evaluated.
- 6) Ledger commit and forwarding: The verified aggregate and its hash are appended to the lightweight ledger at the gateway tier, and the result is forwarded toward the base station.

D. Simulation-Based Evaluation Methodology

We intend to evaluate the framework using a discrete-event network simulator (e.g., NS-3 or a MATLAB-based custom simulator) under the first-order radio energy dissipation model, in which transmitting or receiving a k -bit message over distance d consumes energy proportional to k and to d^2 (free-space) or d^4 (multipath), consistent with widely used WSN energy models [11]. Table III lists the representative simulation parameters that will configure this evaluation.

TABLE III
REPRESENTATIVE SIMULATION PARAMETERS

Parameter	Representative Value / Range
Number of sensor nodes (N)	100 – 500
Deployment field	100 m × 100 m to 500 m × 500 m
Initial energy per node	0.5 J
Base station location	Center of field / fixed outside field
Compromised-node fraction	5% – 20% of deployed nodes
Data packet size	500 bytes (aggregated payload)
MAC / control packet size	25 – 50 bytes
Radio energy model	First-order model (free-space / multipath)
Simulation rounds	1000 – 5000 aggregation rounds
Baseline protocols compared	Energy-efficient (non-secure), privacy-preserving statistical, blockchain-based aggregation

The framework will be benchmarked against baseline schemes representative of the three categories in Table I — an energy-efficient but security-agnostic protocol, a privacy-preserving statistical aggregation scheme, and a blockchain-based secure aggregation scheme — on the following metrics:

- Network lifetime: rounds elapsed until a defined percentage of nodes exhaust their energy (first-node-death and half-node-death metrics).
- Energy consumption: average and total energy dissipated per aggregation round.
- Aggregation accuracy: deviation of the recovered aggregate from the ground-truth value with no attack present.
- Malicious-node detection rate and the false-positive rate of the trust-based filter.
- End-to-end latency and the communication overhead introduced by the cryptographic and ledger operations.

E. Complexity Considerations

For each member node, there is one homomorphic encryption operation and one keyed-hash operation at the sensor tier, for each reporting round. Both are $O(1)$, are not dependent on network sizes, and the computational and energy per node does not significantly deviate from that of an unsecured scheme. The main cost per CH at each round is $O(m)\#$ MAC checks where m is the number of m report(s) received, and $O(m)\#$ homomorphic addition, in addition to the $O(1)\#$ comparison of a reputation value stored in the CH. This ledger append operation on the gateway-tier applies solely to the communication between the CH and the base-station, and is not proportional to the number of leaf sensor nodes. This is because $O(1)$ cost is incurred at the leaf tier and $O(m)$ cost is incurred at the CH tier, where m denotes the cluster size, and not the network size, N , consequently place a significantly lower bound on the scalability of SLHDA through N growth.

VI. KNOWLEDGE GAPS AND RESEARCH CHALLENGES

Drawing on the literature reviewed in Section II, we identify six interrelated categories of open research gaps that stand in the way of secure and efficient data aggregation at scale.

- Scalability: Most cryptographic aggregation schemes are validated on small to moderate topologies, and their behavior under thousands of nodes and multi-hop CH chains remains poorly characterized.
- Energy efficiency: Security mechanisms (encryption, MAC computation, consensus) add energy cost that can eat into the savings aggregation itself provides, particularly on the most resource-constrained nodes.
- Security challenges: Existing schemes typically address a subset of attacks, such as false-data injection, while leaving others — collusion among multiple compromised CHs, side-channel leakage — unaddressed.
- Real-world deployment: Many proposals are validated only in simulation; field trials that capture realistic channel conditions, node failures, and mobility are still comparatively rare.
- Emerging technologies: Integrating blockchain or deep-learning-based anomaly detection improves security but brings computational and storage demands that must be partitioned carefully across the network hierarchy.
- Resource limits: Hard constraints on memory, processing power, and battery capacity continue to bound the complexity of the cryptographic and consensus protocols that are feasible on the sensor tier.

VII. EXPECTED OUTCOMES AND DISCUSSION

The layered design should prove the feasibility of adding confidentiality and integrity protection to the standard cluster-based aggregation with minimal energy cost: first, homomorphic combination eliminates the need for per-hop decryption at the network tier; second, the blockchain ledger is limited to the less energy-intensive 'gateway' tier. In a small number of aggregation rounds, we expect that the trust-based anomaly filter will correctly isolate a large portion of the malicious CHs at the cost of a bounded false-positive rate, which is a compromise between detection rate and false-positive rate. In sum, the design must be resilient to false data injection and give the network greater

lifespan than security-agnostic frameworks that don't do in-network aggregation and that are based on end-to-end encryption.

A summary-qualitative evaluation of the position of SLHDA based on the three baseline categories identified in Table 1 is shown in Table IV. The comparison is only intended to assist in the quantitative evaluation suggested in Section V-D, and should not be used rather than it.

TABLE IV
QUALITATIVE COMPARISON OF SLHDA WITH BASELINE CATEGORIES

Property	Energy-Efficient (Non-Secure)	Privacy-Preserving Statistical	Blockchain-Based Secure	Proposed SLHDA
Data confidentiality	No	Yes (cryptographic)	Partial (ledger integrity, not confidentiality by default)	Yes (homomorphic encryption)
Per-hop integrity check	No	Limited	Yes (via ledger/consensus)	Yes (keyed-hash MAC)
Compromised-node detection	No	No	Limited	Yes (trust-based filter)
Relative energy overhead	Lowest	High (cryptographic cost)	Highest (consensus cost)	Low–moderate (confined to CH/gateway tier)
Auditability	No	No	Yes	Yes (lightweight, gateway-tier ledger)

VIII. CONCLUSION AND FUTURE WORK

This paper summarizes the current status of secure data aggregation for wireless sensor networks; summarizes the frequent successes and limitations of previously reported techniques, and identifies six key categories of open research issues that cover scalability, energy efficiency, security, deployability, emerging technologies and integrated technologies, and resources. We then presented SLHDA, a hybrid approach combining residual-energy-aware clustering, keyed-hash integrity verification, homomorphic in-network aggregation, trust-based anomaly filtering, and a gateway-tier blockchain audit ledger, and we proposed a simulation-based experiment technique to compare performance of our approach with representative baselines.

Future work will incorporate the SLHDA framework into a network simulator for future simulation, perform a comparative evaluation of the baseline techniques summarized in Table I and explore lightweight post-quantum resistant primitives and federated or on-device machine learning based candidate methods for enhancing the security of the cryptographic mechanisms that do not violate the strict energy and computation constraint of sensor nodes.

ACKNOWLEDGMENT

The author acknowledges the Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, for institutional support extended toward this research.

REFERENCES

- [1] B. A. Begum and S. V. Nandury, "Data aggregation protocols for WSN and IoT applications—A comprehensive survey," *J. King Saud Univ.—Comput. Inf. Sci.*, vol. 35, no. 2, pp. 651–681, Feb. 2023.
- [2] B. A. Begum and S. V. Nandury, "A survey of data aggregation protocols for energy conservation in WSN and IoT," *Wireless Commun. Mobile Comput.*, vol. 2022, Art. no. 8765335, 2022.
- [3] X. Tao, S. Peng, Y. Zhu, and W. Liu, "EPPSA: Efficient privacy-preserving statistical aggregation scheme," *Secur. Commun. Netw.*, vol. 2022, 2022.
- [4] T. Aga et al., "Exploring secure and private data aggregation techniques for IoT," *Discover Internet of Things*, vol. 4, 2024.
- [5] M. S. Yousefpoor, E. Yousefpoor, H. Barati, A. Barati, A. Movaghar, and M. Hosseinzadeh, "Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: A comprehensive review," *J. Netw. Comput. Appl.*, vol. 190, p. 103118, 2021.
- [6] E. Yousefpoor, H. Barati, and A. Barati, "A hierarchical secure data aggregation method using the dragonfly algorithm in wireless sensor networks," *Peer-to-Peer Netw. Appl.*, vol. 14, pp. 1917–1942, 2021.
- [7] S. Abbasian Dehkordi, K. Farajzadeh, J. Rezazadeh, R. Farahbakhsh, K. Sandrasegaran, and M. Abbasian Dehkordi, "A survey on data aggregation techniques in IoT sensor networks," *Wireless Netw.*, vol. 26, no. 2, pp. 1243–1263, 2020.
- [8] B. Vikas, B. Sagar, and M. Munjul, "Security issues in wireless sensor network—A survey," *J. Discrete Math. Sci. Cryptogr.*, vol. 24, no. 5, pp. 1415–1427, 2021.
- [9] M. Dener and A. Orman, "BBAP-WSN: A new blockchain-based authentication protocol for wireless sensor networks," *Appl. Sci.*, vol. 13, no. 3, p. 1526, 2023.
- [10] "MCAGCN-BBO: Blockchain and graph convolutional network-based secure data aggregation for wireless sensor networks," *Optik*, 2026, in press.
- [11] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, Maui, HI, USA, 2000, pp. 1–10.