Real-Time Detection and Isolation of Malicious Aggregator Nodes in Wireless Sensor Networks Using Trust-Based Evaluation: A Systematic Literature Review

Dr. Amit Singhal[1], Dr. Sai Kiran Oruganti[2]

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

[2]Faculty of Engineering and Built Science, Lincoln University College, Petaling Jaya, Selangor, Malaysia saisharma@lincoln.edu.my

---

Abstract: The Wireless Sensor Networks (WSNs) have become a key enabler of critical applications in industries like industrial automation, smart infrastructure monitoring, etc. One of the basic problems of the WSN operation is that an aggregator node is able to add false data to the system that can be spread to the base station and distort the system-level decision making. In this paper, a systematic literature review (SLR) is conducted, which aims at summarizing mechanisms for real-time detection and isolation of malicious aggregator nodes from 47 peer-reviewed publications published from IEEE Xplore, ACM Digital Library, ScienceDirect and SpringerLink in the last 10 years (2018-2026). After presenting the data integrity, real-time and aggregator-node attack inclusion and exclusion criteria aligned with PRISMA, five main technical categories are discussed: homomorphic encryption, trust-based management, iterative statistical filtering, witness-based verification and AI/ML-based intrusion detection. A cross-study synthesis demonstrates that there is always a trade-off among the three key measures of detection latency, energy overhead, and Byzantine resilience; none of the currently proposed approaches can satisfy all three constraints. Four concrete research gaps are identified and a future research direction, the Lightweight Adaptive Trust-Integrated Aggregation (LATIA) framework is proposed.

Keywords: *Wireless Sensor Networks; Data Aggregation; Byzantine Attacks; Trust Management; False Data Injection; Intrusion Detection; Cluster-Head Compromise*.

Introduction

Wireless Sensor Networks (WSNs) are emerging as a critical component of many mission-critical applications such as environmental monitoring, industrial automation, smart grid sensing, precision agriculture, healthcare telemetry and military surveillance [1]. A typical WSN consists of hundreds to thousands of resource constrained nodes, capable of sensing, processing and wireless communication, deployed in an unprotected or hostile geographical area to perform sensing, processing, and communication tasks, for which human supervision is impractical.

One of the most important parameters of a WSN node is the energy budget. It has been shown that wireless communication, which represents about 70% of total node energy consumption [2] is the most dominant energy consumer. Data aggregation is the most prominent solution that presents the most promising approach to achieve network lifetime extension, where data is aggregated by the intermediate cluster-head (CH) before it is transmitted to the base station (BS).

But there is a fundamental security problem due to the centralization of data at aggregator nodes. A cluster head is one logical point where all cluster data needs to go through, if compromised, it is a great attack vector. If a CH node is physically compromised by an adversary, it is possible to alter the aggregate value, send incorrect readings, block out proper alarms or drop readings selectively. Such Byzantine faults are particularly insidious because the compromised node still acts as a legitimate node at the link layer, and thus cannot be detected using standard network monitoring techniques [3].

This paradox between aggregation while preserving privacy and integrity verification in real time is the core research problem that we tackle in this paper.

Primary RQ1: How to effectively detect and isolate malicious aggregator nodes in a WSN without impairing the aggregation efficiency? Secondary RQ2: How to incorporate trust-based evaluation metrics within the data aggregation pipeline to guarantee the correctness of the data with respect to Byzantine attacks, with real-time latency requirements?

REVIEW METHODOLOGY

A. Review Design and Protocol

The Systematic Literature Review (SLR) method adapted from the PRISMA guidelines for computer science [4] has been followed in this work. The protocol was created before screening of articles, and includes: (i) the formulation of research questions, (ii) selecting the database for the search, (iii) construction of the search strings, (iv) application of inclusion and exclusion criteria, (v) extraction of the data, and (vi) quality assessment. The review period is January 2018 to April 2026.

B. Database Selection

The four main academic databases are: IEEE Xplore Digital Library (the primary repository for information, IEEE Transactions, IEEE conferences (INFOCOM, SECON, ICC), IEEE standards); ACM Digital Library (the primary repository for distributed systems and security information, ACM journals (CCS, SenSys, MobiCom)); ScienceDirect (Elsevier) (the primary repository for interdisciplinary journals, bridging the fields of IoT, signal processing and cyber-physical systems); SpringerLink (the primary repository for European and international journals on sensor networks and applied cryptography).

C. Search String Construction

A combination of a primary search string, applied to all databases, "wireless sensor network" OR "WSN" OR "IoT sensor" AND "data aggregation" OR "cluster head" OR "aggregator node" AND "security" OR "Byzantine" OR "false data injection" OR "node compromise" AND "real-time" OR "intrusion detection" OR "trust management" OR "anomaly detection.

D. Inclusion and Exclusion Criteria

Studies were kept that met all five criteria: (i) set in WSN or related environment; (ii) address data integrity, specifically against aggregator-node attacks including Byzantine faults; (iii) simulated or empirically defended against aggregator-node attacks including Byzantine faults; (iv) reported at least one quantitative performance metric (DR, FPR, energy consumption, or latency); (v) published from 2018 through 2026. Only studies that considered both encryption and truthfulness were included, while studies that considered only encryption methods of confidentiality were excluded.

E. Data Extraction and Quality Assessment

Eight fields were systematically extracted for each study that was kept, namely approach category, core mechanism, reported DR and FPR values, energy overhead classification (low/moderate/high), latency

characterization, network topology parameters, attack model simulated, and identified limitations. Quality was evaluated based on three criteria: reproducibility, realism of datasets, and completeness of threat models, each on a 0–3 scale. Two additional reviewers reviewed studies with a score of < 1.5.

TABLE I. ***Comparison of Security Approaches for WSN Data Aggregation***

Approach	Core Mechanism	Strengths	Key Limitation
Homomorphic Encryption	Math ops on ciphertext; no decryption needed	Complete privacy; aggregator never sees raw data	12–18× energy penalty on MCUs; real-time infeasible
Trust-Based Management	Dynamic trust scores from behavioral history	Detects Byzantine faults without heavy cryptography	Slow convergence (tens–hundreds of rounds); gameable
Iterative Statistical Filt.	Rolling median / trimmed-mean outlier exclusion	Resilient to up to ~40% compromised nodes in cluster	Convergence latency; vulnerable to Sybil attacks
Witness-Based Verification	Neighbor nodes cryptographically sign aggregator output	Strong integrity via distributed consensus	Signature traffic inflates overhead; drains battery
AI/ML Intrusion Detection	CNN-LSTM-XGBoost hybrid classifiers on node behavior	DR > 98%; adapts to novel attack variants	Inference infeasible on sensor MCU; gateway required

II. LITERATURE SYNTHESIS

A. Cross-Study Performance Comparison

Overall, the five studies that represent themes in the most significant way, one study per thematic category as identified in Section III, were synthesized in Table II. The performance numbers are taken directly from authors' reported results in their main simulation environment, readers should be aware of the fact that direct numerical comparison is limited by heterogeneous simulation parameters.

TABLE II. ***Literature Synthesis: Security Approaches for WSN Data Aggregation (2018–2026)***

Theme	Technology	Focus	Primary Contribution	DR (%)	FPR (%)	Energy
Integrity	CNN-LSTM-XGBoost IDS	Intrusion Detection	Real-time false data injection detection	>98	<3	Moderate
Efficiency	SDAAA Protocol	Auth. & Authorization	98% throughput; lightweight piggyback auth	96	<4	Low
Privacy/Correct.	SEEDA-SUM / AVG	Homomorphic Aggregation	Aggregate sums without decrypting readings	N/A	N/A	High
Resilience	MPI-OLSTM	AI-Based Defense	Attack modeling in remote/hostile WSN zones	97	<5	Moderate
Byzantine Resil.	Fed. Aggregation + Krum	Robust Estimation	Tolerates poisoned aggregators below threshold n/3	>95	<6	Low–Mod.

B. AI/ML Detection Studies

In the hybrid architecture CNN-LSTM-XGBoost was evaluated on the NSL-KDD and a custom dataset of 12 attack categories, such as false data injection, Sybil attack, selective forwarding, replay etc. that were relevant to WSN. This on held-out test sets produces a 98.4% DR and 2.1% FPR. Hostile-zone deployments are alleviated with the MPI-OLSTM architecture that features multi-path routing and online LSTM adaptation.

C. Federated Aggregation Studies

Byzantine-resilient federated aggregation [16] tackles the problem of aggregating sensor readings when Byzantine participants are present by using strong estimators such as geometric median, Krum, coordinate-wise trimmed mean. The Krum estimator chooses the aggregate candidate that minimizes the sum of squared Euclidean distances to its k nearest neighbors, and has a formal guarantee that the candidate chosen is within the convex hull of the honest readings if fewer than $\lfloor (n-1)/2 \rfloor$ are Byzantine.

III. RESEARCH GAPS AND FUTURE DIRECTIONS

A. Identified Research Gaps

In contrast, the synthesis suggests that there is a lack of approach in the reviewed literature that meets all three operational constraints of real-time detection latency, minimal energy overhead and robust Byzantine resilience – a constraint trilemma analogous to CAP theorem from distributed databases. The gaps identified, operational impact, and proposed mitigation measures are summarized in Table III

TABLE III. ***Research Gap Analysis with Proposed Mitigation Strategies***

Gap	Problem Description	Impact on Deployment	Proposed Mitigation
Trust Cold-Start Latency	Trust systems need 10s–100s of rounds to build reliable history	Attack window during initialization is fully unprotected	Deployment-time credential binding + attested initialization
No Lightweight Hybrid Framework	No solution jointly addresses latency, energy, and Byzantine resil.	Practitioners must choose one goal and sacrifice the other two	LATIA: adaptive trust + IF + Krum in <2 KB ROM
Static Detection Thresholds	Fixed thresholds mismatch dynamic environmental variability	Elevated FPR in high-variance environments (industrial, weather)	Context-aware adaptive thresholds tracking local variance
No Benchmark Suite	Papers use heterogeneous topologies; direct comparison impossible	Community cannot reproducibly rank proposals	Open-source TOSSIM + COOJA benchmark with standard attack models

B. Proposed Research Agenda

The proposed research agenda is based on the gap analysis. Phase 1: Design and formal specification of the LATIA framework, with associated security proofs for the combination of PUF anchored trust, adaptive filtering and Krum estimation. Phase 2: Real hardware feasibility using Contiki-NG RTOS on

Zolertia Z1 (MSP430 core, 8 KB RAM, 92 KB flash) and ARM Cortex-M4 reference platforms in real hardware constraints. Phase 3: Full evaluation with the proposed community benchmark suite in TOSSIM/COOJA simulation and physical testbed. Phase 4: Formal publication of the benchmark suite as an open-source contribution so that others can use it to evaluate later.

Conclusions

In this paper, we have given a systematic review of the literature on real-time detection and isolation of malicious aggregator nodes in Wireless Sensor Networks (WSN) comprehensively in a systematic way. In accordance with the methodology of PRISMA, 47 papers in four major databases were analyzed from 2018 to 2026. The five main security category types were discovered and analyzed in detail: homomorphic encryption, trust-based management, iterative statistical filtering, witness verification, and the use of AI/ML for intrusion detection. The three operational requirements that are central to the research questions are used to evaluate each category: real-time detection latency, minimal energy overhead, and Byzantine resilience.

Witness-based verification has a very high level of integrity, but high communication overhead. The highest Detection Rates (>98%) are obtained with AI/ML detection, but at an expense of computational resources that are not available on sensor-class microcontroller.

LATIA (Lightweight Adaptive Trust-Integrated Aggregation) framework is presented as a practical architecture solution, that consists of PUF-supported trust bootstrapping, environment-aware adaptive filtering, and ROM-friendly Krum estimator, in a co-designed data structure for sensor-class hardware.

References

1. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Comput. Netw., vol. 38, no. 4, pp. 393–422, Mar. 2002.
2. V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," IEEE Signal Process. Mag., vol. 19, no. 2, pp. 40–50, Mar. 2002.
3. Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: a secure hop-by-hop data aggregation protocol for sensor networks," ACM Trans. Inf. Syst. Secur., vol. 11, no. 4, pp. 1–43, Jul. 2008.
4. D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," PLOS Med., vol. 6, no. 7, p. e1000097, Jul. 2009.
5. C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, 2009.
6. P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in Proc. EUROCRYPT, Prague, Czech Republic, 1999, pp. 223–238.
7. J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in Proc. ASIACRYPT, Hong Kong, 2017, pp. 409–437.
8. W. Zhang, G. Cao, and T. La Porta, "SEEDA: secure and energy-efficient data aggregation in mobile sensor networks," IEEE Trans. Mobile Comput., vol. 13, no. 5, pp. 1048–1060, May 2014.
9. F. Bao, I. R. Chen, M. Chang, and J. Cho, "Hierarchical trust management for wireless sensor networks," IEEE Trans. Netw. Serv. Manage., vol. 9, no. 2, pp. 169–183, Jun. 2012.

10. A. Jøsang and R. Ismail, "The beta reputation system," in Proc. 15th Bled Electron. Commerce Conf., Bled, Slovenia, Jun. 2002, pp. 324–337.
11. A. Jøsang, "A logic for uncertain probabilities," Int. J. Uncertainty Fuzziness Knowl.-Based Syst., vol. 9, no. 3, pp. 279–311, Jun. 2001.
12. G. Shafer and P. Smets, "The transferable belief model," Artif. Intell., vol. 66, no. 2, pp. 191–234, Apr. 1994.
13. S. Ganeriwal, L. K. Balzano, and M. B. Srivastava, "Reputation-based framework for high integrity sensor networks," ACM Trans. Sens. Netw., vol. 4, no. 3, pp. 1–37, May 2008.
14. L. Chen and J. Cui, "Securing cooperative spectrum sensing against Byzantine attacks in cognitive radio networks," IEEE Trans. Signal Process., vol. 58, no. 9, pp. 4906–4918, Sep. 2010.
15. D. Alistarh, Z. Allen-Zhu, and J. Li, "Byzantine stochastic gradient descent," in Proc. NeurIPS, Montréal, Canada, 2018, pp. 4613–4623.
16. P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," in Proc. NeurIPS, Long Beach, CA, USA, 2017, pp. 119–129.