IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 3, July 2025

Impact Factor: 7.67

Privacy-Preserving Data Aggregation in IoT Environments: A Lightweight Edge-Based Hybrid Framework

Kakumanu Haritha[1] and K. Sundari[2]

Lecturer in Computer Science[1]
III B.Sc. Computer Science[2]
Sir. C. R. Reddy College of Engineering, Eluru

Abstract: *As Internet of Things (IoT) ecosystems scale across sectors like smart healthcare, urban infrastructure, and industrial monitoring, concerns about user privacy and data security have become paramount. Traditional aggregation methods—where raw data is collected and processed centrally—expose sensitive information to breaches, inference attacks, and misuse. This paper presents a hybrid privacy-preserving data aggregation model that leverages edge computing, lightweight encryption, and differential privacy techniques to ensure confidentiality without compromising data utility or system performance.*

The proposed framework distributes aggregation tasks to edge nodes, where sensor data is pre-processed, obfuscated, and locally encrypted before transmission. Experimental evaluation using both simulated and physical IoT devices demonstrates that the hybrid model achieves over 90% reduction in data exposure risk, maintains aggregation accuracy above 95%, and reduces bandwidth usage by more than 50% compared to centralized systems. Statistical analysis confirms the model's resilience to inference and reconstruction attacks while maintaining low latency and computational overhead.

This research contributes a scalable and practical solution for secure data processing in IoT deployments, offering critical insights for architects seeking to balance privacy with real-time analytics in sensitive domains. Future directions include adaptive privacy budgets, decentralized trust models, and integration with federated learning and zero-trust infrastructures..

Keywords*:* Privacy-Preserving Aggregation, Edge Computing in IoT, Differential Privacy**,** Lightweight Encryption**,** IoT Data Security

I. INTRODUCTION

The rapid growth of the Internet of Things (IoT) has revolutionized how we interact with the physical world, enabling real-time monitoring, control, and automation across diverse domains such as smart homes, transportation, agriculture, healthcare, and urban infrastructure. IoT systems collect massive volumes of data from connected sensors and devices, often including sensitive information such as a person's health statistics, location history, behavioral patterns, or environmental conditions within private spaces. While this data is instrumental in driving innovation and efficiency, it also raises significant concerns regarding privacy and data security.In traditional IoT frameworks, individual devices transmit raw data directly to a centralized server or cloud infrastructure for storage, aggregation, and analysis. Although this model offers scalability and processing power, it introduces a major vulnerability: the central collection point becomes a high-risk target for breaches, surveillance, and misuse. A single compromise of the central system can expose the entire dataset—impacting not only individual users but potentially millions of devices and applications. Additionally, centralized models often violate principles of data minimization, informed consent, especially in sensitive domains like healthcare and smart surveillance. To address these concerns, researchers, and developers are increasingly focusing on privacy-preserving data aggregation (PPDA) techniques in IoT environments. PPDA aims to perform computations such as averages, counts, or trends without revealing individual-level data. In other words, it allows

DOI: 10.48175/IJARSCT-28537

useful inferences from collective data without compromising the privacy of any one device or user. This is particularly important in sectors governed by strict privacy regulations, including GDPR, HIPAA, and similar laws in Asia and the Middle East.However, ensuring privacy in IoT aggregation is inherently complex. First, IoT devices are often resource-constrained—limited in terms of processing power, memory, and energy. Therefore, any privacy solution must be lightweight enough to run on minimal hardware. Second, IoT networks are highly dynamic, with devices frequently joining or leaving, suffering from intermittent connectivity, and operating over heterogeneous protocols. Third, preserving privacy must not come at the cost of data utility—aggregated results must remain accurate and timely for effective use in applications such as smart grids, anomaly detection, or personalized services.

Several techniques have emerged to support privacy-preserving aggregation, including homomorphic encryption, differential privacy, secure multiparty computation (SMC), edge computing, and blockchain-based trust models. Yet, most of these either impose significant computation and communication overhead, rely on unrealistic assumptions (such as synchronized clocks or fully trusted nodes), or fail to scale in real-world, high-volume deployments. As a result, there is no "one-size-fits-all" solution to privacy-preserving data aggregation in IoT, and a hybrid approach—balancing computation, privacy, and system performance—is needed.

This research aims to address this gap by proposing a practical, scalable, and secure framework for privacy-preserving data aggregation in IoT environments. The framework leverages lightweight edge-based aggregation, obfuscation methods, and distributed trust mechanisms to ensure individual privacy without significantly compromising system responsiveness or accuracy. Through a combination of literature synthesis, architecture design, simulation, and performance benchmarking, this study seeks to identify best practices and novel techniques that can be adopted in real-world IoT applications at scale.

II. LITERATURE REVIEW

Research into privacy-preserving aggregation for IoT environments has intensified in recent years—prompted by stricter data protection regulations and increased consumer awareness. Early solutions relied on homomorphic encryption, which enables mathematical operations on encrypted data. Li et al. (2019) demonstrated homomorphic schemes applied to smart-meter data, but the computational overhead rendered them impractical for resource-constrained IoT sensors. More recent approaches incorporate secure multiparty computation (SMC), where multiple devices jointly compute aggregated results without shared raw values. While effective in theory, SMC protocols often require heavy communication rounds and assume synchronous, reliable connectivity—conditions often absent in real-world IoT networks. To balance practicality with security, differential privacy has emerged as a promising technique. Abdelzaher et al. (2020) proposed injecting calibrated noise into sensor readings before aggregation, ensuring that individual contributions remain indistinguishable while preserving statistical trends. Although lightweight and effective under offline scenarios, traditional differential privacy methods can degrade precision or require centralized calibration—potentially introducing trust vulnerabilities.An alternative paradigm focuses on edge-assisted aggregation architectures. In this model, local data from clusters of devices are aggregated at nearby edge nodes using secure partial aggregation methods such as randomized blinding or secret sharing, before transmission to the central server. Wang et al. (2021) showed that short-range edge aggregation reduces communication overhead and enhances privacy, but scalability remained limited when device cluster sizes increased beyond 500 nodes.Another area of research explores hybrid encryption-obfuscation techniques, where devices perturb data using lightweight masked transformations before sending to a semi-trusted aggregator. In a study by El-Sayed et al. (2022), these methods achieved privacy protection with minimal computational load but required strong assumptions about adversarial capabilities and trust in edge nodes.Finally, the integration of blockchain and decentralized ledgers has recently been proposed to enforce transparent audit trails and secure aggregation without centralized trust. Zhang et al. (2023) introduced blockchain-based IoT aggregation where edge nodes or gateways act as validators, but the consensus overhead and complexity limit real-time applicability in many IoT environments. In summary, while a range of privacy-preserving aggregation approaches exist—homomorphic encryption, SMC, differential privacy, edge clustering, blockchain—the critical challenge remains to harmonize privacy, computational efficiency, network resource constraints, and scalability in real-world IoT deployments.

DOI: 10.48175/IJARSCT-28537

III. RESEARCH OBJECTIVES

This research aims to address the growing need for secure and privacy-conscious data handling in large-scale IoT networks. Given the sensitive nature of data captured by IoT devices—ranging from environmental parameters to personal health metrics—it becomes essential to ensure that aggregation operations do not compromise individual privacy. The overarching goal of this study is to design and evaluate a **scalable, efficient, and privacy-preserving data aggregation framework** that can be applied across diverse IoT applications with minimal overhead.

The specific objectives of the study are:

- To analyze the key privacy threats in centralized and semi-centralized IoT aggregation systems, focusing on data exposure, inference attacks, and unauthorized access.
- To investigate and compare existing privacy-preserving aggregation techniques—including differential privacy, secure multiparty computation (SMC), homomorphic encryption, and edge-based pre-aggregation.
- To propose a hybrid privacy-preserving data aggregation model that leverages lightweight obfuscation, edge node coordination, and minimal trust assumptions to protect user data.
- To implement the proposed model in a simulated IoT environment using diverse device types and communication protocols, validating its feasibility and adaptability.
- To evaluate the system performance in terms of latency, computational overhead, aggregation accuracy, bandwidth efficiency, and resistance to privacy breaches.
- To validate the effectiveness of the privacy model through statistical analysis, demonstrating its balance between utility and confidentiality.

IV. RESEARCH HYPOTHESES

Based on the challenges and research goals, the following hypotheses are proposed to guide the investigation and provide a basis for empirical validation:

H_1: A hybrid privacy-preserving data aggregation framework using edge-based obfuscation significantly reduces individual data exposure risk compared to centralized aggregation methods. This hypothesis addresses the primary privacy concern and seeks to measure actual data leakage probabilities under both architectures.

H_2: The proposed model achieves comparable aggregation accuracy ($\geq 95\%$) to raw data aggregation, with less than 10% overhead in latency and computation. This evaluates the trade-off between privacy and performance, ensuring that privacy enforcement does not degrade system utility beyond acceptable limits.

H_3: Edge-based pre-aggregation reduces network bandwidth usage by at least 50% compared to centralized models transmitting raw data.By quantifying bandwidth savings, this hypothesis explores the model's scalability benefits in constrained environments.

H_4: The hybrid model demonstrates higher robustness to common privacy attacks (e.g., reconstruction or membership inference) than baseline models without privacy enhancements. This measures privacy efficacy in adversarial conditions, validating the system's security posture beyond theoretical guarantees.

V. RESEARCH DESIGN

To examine the practicality and effectiveness of privacy-preserving data aggregation in IoT, this study adopts an experimental-comparative research design using a combination of simulated IoT environments, custom-built prototypes, and statistical validation. The research integrates both quantitative metrics (e.g., data leakage probability, latency, accuracy, bandwidth usage) and scenario-based threat testing to evaluate real-world performance of various aggregation strategies. The study follows a three-phase process.

Phase 1: System Architecture and Method Selection

In the initial phase, a modular hybrid aggregation architecture is designed. The architecture consists of:

- IoT Sensor Nodes: Resource-constrained devices that collect data such as temperature, humidity, or activity status.

- Edge Aggregators: Intermediate nodes (e.g., Raspberry Pi or ESP32) responsible for partial aggregation, data obfuscation, and local encryption.
- Central Collector: A lightweight server that collects already obfuscated or aggregated data for final computation.

Various aggregation techniques are implemented and tested, including:

- Baseline (centralized aggregation with raw data)
- Edge-only aggregation (no privacy mechanisms)
- Obfuscation with differential privacy noise injection
- Lightweight encryption with partial homomorphic properties
- Edge-clustered aggregation using randomized masking and modular transformation

Phase 2: Simulation and Prototype Deployment

To simulate realistic IoT environments, two platforms are used:

- Docker-based sensor network simulation with up to 1,000 virtual nodes.
- Physical prototype with 50 devices using ESP8266/ESP32 microcontrollers connected over MQTT and HTTP protocols.

Scenarios include:

- Regular data reporting (e.g., every 30 seconds)
- Burst data during abnormal events (e.g., fire alert)
- Fault injection (e.g., dropped packets, compromised node simulation)

Each configuration is tested for:

- Latency (ms)
- Bandwidth usage (KB/s)
- Computational load (% CPU/RAM)
- Aggregation error (%)
- Data exposure probability (%)

Phase 3: Statistical Analysis and Threat Testing

The performance results are analysed using:

- t-tests to compare privacy-enabled and baseline models
- ANOVA for latency and load variation across methods
- Chi-square tests for comparing leakage under different privacy threats
- Regression analysis to model overhead scalability

Threat scenarios are crafted to test:

- Membership inference attacks
- Data reconstruction from aggregates
- Compromised node injection

This experimental setup ensures that the proposed model is evaluated not only for efficiency but also for resilience under privacy risks in diverse and high-scale IoT deployments.

VI. SAMPLE AND SAMPLING TECHNIQUES

The “sample” in this context refers to the IoT device configurations, aggregation strategies, privacy models, and simulated traffic profiles used to test the system. Given the exploratory nature of this research, a purposive sampling approach is employed to ensure the inclusion of representative variations in IoT devices, communication protocols, and attack scenarios.

6.1 Device and Protocol Sample

Sensor Nodes:

- 50 physical microcontroller units (ESP32, ESP8266, Arduino Uno + WiFi)
- 1,000 Docker-based virtual devices simulating temperature, motion, air quality, and noise sensors

Protocols Used:

- MQTT for real-time telemetry
- HTTP/REST for device onboarding and control
- CoAP for lightweight communication from low-power devices

Edge Aggregators:

- Raspberry Pi 4 (4GB) with Python-based data processors
- ESP32 with MicroPython and lightweight encryption libraries

Central Server:

- Ubuntu-based virtual machine with MongoDB, Flask, and visualization via Grafana

6.2 Sampling Rationale

- Devices selected represent low-cost, resource-constrained environments, mimicking realistic consumer or municipal deployments.
- Communication protocols reflect diverse interoperability challenges in current IoT ecosystems.
- Privacy mechanisms sampled range from lightweight to strong cryptographic approaches, allowing comparison in performance vs. privacy tradeoffs.
- Data profiles are designed to include normal activity, rare events, and privacy-sensitive entries (e.g., home presence detection, medical data).

6.3 Limitations and Scope of Sample

- Some privacy threats (e.g., physical side-channel attacks) are not tested in this simulation.
- The system assumes honest-but-curious aggregators (semi-trusted edge nodes) but not fully malicious actors unless explicitly tested.
- Battery lifetime and energy consumption, while relevant, are not deeply evaluated in this round of experiments.
- Despite these limitations, the sample design ensures a balanced, realistic, and technically diverse setup for testing both privacy efficacy and system scalability in modern IoT environments.

VII. DATA ANALYSIS

This section presents the statistical analysis of the data collected from the various aggregation models tested. The objective is to compare the performance, privacy, and efficiency of different aggregation techniques—namely centralized raw aggregation, edge-based aggregation without privacy, and the proposed hybrid model integrating differential privacy and lightweight encryption at the edge.

7.1 Key Performance Metrics

The following metrics were evaluated across all simulations and prototype runs:

- Latency (ms): Time between data generation and aggregate result at the central server.
- Bandwidth Usage (KB/s): Total upstream network usage per node cluster.
- Aggregation Error (%): Deviation of aggregated result from ground truth (used to evaluate utility).
- CPU Load (%): Processor utilization on edge devices.
- Data Exposure Probability (%): Measured as the rate at which individual values could be reconstructed in threat models.

DOI: 10.48175/IJARSCT-28537

- Resilience to Attack: Number of successful inference attacks vs. attempts.

Each configuration was run with increasing device counts (from 100 to 1,000 virtual sensors and 50 physical devices) and repeated three times per load setting.

7.2 Descriptive Statistics Summary

Metric	Centralized Raw	Edge-Only	Proposed Hybrid Model
Avg. Latency (ms)	825	412	498
Bandwidth Usage (KB/s)	38.5	23.6	16.2
Aggregation Error (%)	0%	2.3%	3.6%
CPU Load on Edge (%)	11.2	13.5	21.8
Data Exposure Probability (%)	100%	87%	9.3%
Successful Attacks (%)	92%	48%	6%

These results show that while the hybrid model introduces moderate computation overhead and minimal aggregation error, it significantly enhances privacy while conserving bandwidth.

7.3 Inferential Statistics

A. Data Exposure Probability (Chi-Square Test)

H_0: No significant difference in data exposure across models.

$\chi^2(2, N=3000) = 213.67, p < 0.001$

Interpretation: There is a highly significant difference in privacy leakage across models; the hybrid model is most secure.

B. Latency Comparison (t-test)

Comparing hybrid vs. centralized:

$t(58) = -7.22, p < 0.001$

Interpretation: Although slightly slower than raw edge aggregation, the hybrid model is significantly faster than centralized raw aggregation.

C. Bandwidth Usage (ANOVA)

$F(2, 87) = 28.76, p < 0.001$

Interpretation: Bandwidth usage differs significantly between models, with the hybrid system showing the least usage.

D. Aggregation Error vs. Utility (Regression Analysis)

$R^2 = 0.88$, suggesting strong predictability of how aggregation noise impacts accuracy.

Conclusion: The aggregation accuracy remains above 95% in the hybrid model, confirming Hypothesis H_2.

7.4 Threat Simulation Results

Three primary adversarial scenarios were tested:

- Membership Inference Attack – Trying to infer if a specific user's data is part of an aggregate.
- Data Reconstruction Attack – Attempting to reverse engineer individual values from the total.
- Compromised Node Injection – Deploying a rogue device that feeds manipulated data.

DOI: 10.48175/IJARSCT-28537

VIII. FINDINGS

The centralized raw model failed under all three.

The edge-only model resisted node injection but leaked data under inference attacks.

The hybrid model successfully resisted all attacks, with <10% success rate under extreme conditions.

VIII. RESULTS AND DISCUSSION

This study validates the proposed hybrid privacy-preserving aggregation framework as a viable solution for balancing utility, performance, and data confidentiality in IoT environments. The most critical achievement was a 91% reduction in data exposure risk compared to centralized models, while still delivering accurate aggregate insights (error under 4%) and maintaining latency below 500 ms.

The system proved highly bandwidth-efficient, consuming less than half the network resources of non-obfuscated raw-data systems. This makes the architecture especially suitable for rural or bandwidth-limited deployments, where privacy and communication overhead must be tightly managed.

Additionally, the hybrid architecture scaled well with increasing device loads. The overhead on edge processors increased slightly due to encryption and noise calibration, but CPU usage remained under 25% even at peak loads—indicating feasibility on modest edge hardware.

From a security standpoint, the proposed system effectively countered inference and reconstruction threats. The combination of differential privacy and randomized masking introduced just enough ambiguity to frustrate adversaries without degrading data utility—striking a strong privacy-performance balance.

However, the system does have trade-offs:

- Requires semi-trusted edge nodes (honest-but-curious assumption)
- Needs careful noise tuning to avoid excessive aggregation error
- Introduces slight delay (~80 ms) due to local obfuscation and edge-level computations

Despite these limitations, the system demonstrates practical promise for sectors such as:

- Smart healthcare (protecting patient metrics)
- Smart grid analytics (preserving residential consumption privacy)
- Urban mobility monitoring (aggregating vehicle data without tracking individuals)

In conclusion, this research shows that privacy-preserving aggregation is not only possible but also practical—especially when edge computing is leveraged to offload sensitive computations from centralized infrastructure. The hybrid model outperforms legacy architectures in both privacy protection and resource management, providing a strong foundation for future smart systems where data trust is paramount.

IX. CONCLUSION AND FUTURE SCOPE

Conclusion

As the Internet of Things (IoT) becomes increasingly integrated into critical aspects of modern life—spanning homes, cities, industries, and healthcare—the imperative to secure and protect the privacy of sensitive data grows exponentially. This research addressed one of the most pressing challenges in IoT systems: how to aggregate data from distributed, often personal, sources without compromising privacy, system efficiency, or scalability.

Through the design, implementation, and analysis of a hybrid privacy-preserving aggregation architecture, this study demonstrated that a balance between data utility and confidentiality is not only achievable but also efficient when leveraging edge computing and lightweight privacy mechanisms. The proposed model combined edge-level differential privacy, randomized data masking, and bandwidth-efficient hierarchical aggregation to outperform traditional centralized approaches across all tested metrics.

Key findings include:

- Over 90% reduction in data exposure risk compared to centralized models
- 50%+ reduction in bandwidth usage
- Less than 4% aggregation error, maintaining over 95% data accuracy

DOI: 10.48175/IJARSCT-28537

- Strong resistance to privacy attacks such as membership inference and data reconstruction
- Scalability to thousands of devices with minimal performance degradation

While centralized systems remain simpler to manage, they pose unacceptable privacy risks in many use cases. This research provides concrete evidence that privacy can be enforced without sacrificing performance, especially when computation is intelligently pushed to the network edge.

Future Scope

Despite its promising results, the research also highlights several directions for enhancement and broader applicability:

Dynamic Privacy Budget Management

Integrating adaptive differential privacy algorithms that adjust noise levels based on contextual sensitivity and network activity could optimize the trade-off between privacy and accuracy over time.

Zero-Trust and Fully Decentralized Architectures

Future work could explore architectures that eliminate semi-trusted edge nodes, using blockchain or secure multiparty computation to ensure end-to-end zero-trust models.

Privacy-Aware Federated Learning

Combining the proposed aggregation model with federated learning frameworks could allow distributed AI training on obfuscated data—enhancing both intelligence and confidentiality in IoT networks.

Energy-Aware Privacy Protocols

Designing energy-efficient obfuscation algorithms tailored for battery-powered or energy-harvesting IoT devices would extend system life while preserving data privacy.

Cross-Domain Interoperability Standards

Development of standardized APIs and modular privacy-preserving aggregation layers can help unify privacy strategies across different IoT verticals like healthcare, agriculture, and industrial automation.

Real-World Longitudinal Pilots

Testing the hybrid model in live, long-term deployments—e.g., in smart buildings or public transport systems—would validate scalability, reliability, and maintainability under evolving conditions.

User-Controlled Privacy Preferences

Enabling end users to set their own privacy levels dynamically through edge-based privacy interfaces could promote transparency, trust, and compliance with emerging data protection laws. In conclusion, this study contributes a practical, tested, and highly adaptable solution to the critical issue of data privacy in IoT aggregation. As smart environments continue to scale, architectures like the one proposed here will form the backbone of ethical, secure, and sustainable data ecosystems.

REFERENCES

[1]. Abdelzaher, T., Lee, C., & Liu, J. (2020). Enabling privacy-preserving data aggregation in IoT networks through differential privacy. IEEE Transactions on Dependable and Secure Computing, 17(3), 590–603. https://doi.org/10.1109/TDSC.2019.2899010

[2]. El-Sayed, A., Mohamed, N., & Al-Jaroodi, J. (2022). Lightweight encryption and obfuscation for edge-based privacy in IoT. Future Generation Computer Systems, 125, 215–228. https://doi.org/10.1016/j.future.2021.06.009

[3]. Li, F., Luo, B., & Liu, P. (2019). Secure information aggregation for smart grids using homomorphic encryption. Journal of Network and Computer Applications, 134, 66–78. https://doi.org/10.1016/j.jnca.2019.02.001

[4]. Wang, S., Zhang, Y., & Chen, J. (2021). Edge-based secure aggregation in large-scale IoT systems. IEEE Internet of Things Journal, 8(6), 4910–4922. https://doi.org/10.1109/JIOT.2020.3037400

[5]. Zhang, K., Liang, X., & Wang, L. (2023). Blockchain-based decentralized trust for secure IoT data aggregation. Computer Networks, 226, 109621. https://doi.org/10.1016/j.comnet.2022.109621

DOI: 10.48175/IJARSCT-28537

[6]. Ghinita, G. (2020). Privacy for mobile and IoT devices using adaptive noise generation. ACM Transactions on Privacy and Security, 23(1), 5. https://doi.org/10.1145/3386343

[7]. Tan, R., Yang, Z., & Sohraby, K. (2021). Preserving privacy in smart environments: A survey and taxonomy. IEEE Communications Surveys & Tutorials, 23(2), 1357–1388. https://doi.org/10.1109/COMST.2021.3059298

[8]. Lin, H., & Zhang, Y. (2020). Efficient secure multiparty computation for privacy-preserving data aggregation. IEEE Transactions on Information Forensics and Security, 15, 2381–2396. https://doi.org/10.1109/TIFS.2019.2952120

[9]. Raza, S., Seitz, L., & Voigt, T. (2021). CoAP and DTLS-based lightweight security for IoT applications. Sensors, 21(4), 1259. https://doi.org/10.3390/s21041259

[10]. Kumar, R., & Singh, M. (2022). Obfuscation techniques for data security in constrained IoT devices. Journal of Information Security and Applications, 65, 103122. https://doi.org/10

DOI: 10.48175/IJARSCT-28537