Journal of Adaptive Learning Technologies
ISSN **2997-3902** (Online)
Vol. 2, No. 11, November 2025,
Available online at: https://scientificbulletin.com/index.php/JALT

PRIVACY IN IoT: FEDERATED LEARNING AND TELEMETRY ANONYMIZATION

Kim Andrey Odylovich
Samarkand Branch of the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,
Samarkand, Uzbekistan
newwayondri@gmail.com

Article information:
Manuscript received: 12 Oct 2025; **Accepted:** 13 Novl 2025; **Published:** 30 Nov 2025

Abstract: The article examines privacy issues in IoT systems and approaches for addressing them through federated learning and telemetry anonymization methods. It provides a detailed overview of the federated learning (FL) approach, its architecture, and the mathematical optimization framework that enables models to be trained on distributed devices without transferring raw data [1][2]. Classical anonymization techniques (pseudonymization, k-anonymity, L-diversity) and differential privacy with the corresponding formulas are described [3][4]. Practical application cases are considered: in healthcare, FL enables joint model training on medical data without violating patient privacy [5]; in smart cities, IoT devices (e.g., smartphones and sensors) can train models locally (speech recognition, traffic prediction) and transmit only updates [6]; in the energy sector, regional networks collaboratively build consumption prediction models [6]. Schematic illustrations and tables are presented (e.g., distributed learning architecture in Fig. 1 and a comparison of privacy methods in Table 1).

Keywords: IoT, privacy, federated learning, data anonymization, differential privacy.

Introduction

Modern IoT systems consist of billions of sensor-based and smart devices collecting and transmitting vast amounts of data about the environment, user actions, and the users themselves. It is projected that by 2030 there will be 50–100 billion such devices worldwide. This creates significant privacy risks.

First, continuous data collection (location, video, audio, biometrics) allows detailed user profiles to be built, enabling tracking of movements and behavior. Second, many IoT devices lack robust protection: they use simple passwords, receive updates infrequently, and are therefore susceptible to data breaches. For example, one study found that smart speakers begin recording conversations accidentally without user commands (on average ~100 times over 5 days) [7]. Such recordings (20–40 seconds long) can be uploaded to the cloud without the owner's knowledge [7].

Third, IoT device data is often highly sensitive: fitness trackers and medical sensors transmit health-related information, and leakage of such data is unacceptable. There have been cases of health-monitoring applications being hacked, exposing patients' personal information to third parties [8]. Finally, manufacturers and service providers may use data for marketing purposes: creating advertising profiles and sharing them with partners [9]. Many countries enforce strict data protection regulations (GDPR, CCPA, etc.) governing the collection and transfer of personal data; non-compliance can result in severe fines [10].

Thus, the main privacy challenges in IoT are the risk of unauthorized collection of personal information, weak protection of communication channels and data collection points, and the possibility of identifying individuals through unique device characteristics and behavioral

patterns. The integration of numerous devices and cloud services expands the attack surface, increasing the risk of data leakage and unauthorized analysis.

Federated learning (FL) offers a distributed training architecture in which models are trained locally on devices rather than centrally [1][2]. This keeps data at the collection point and transmits only parameter updates to the server. In the classical FL scheme, devices (clients) receive the current global model from the server, train it on their local data, and send back gradients or updated weights. The server aggregates (averages) these updates to produce a new global model. Formally, the global objective function can be expressed as the mean of local loss functions:

$$f(x) = \frac{1}{K}\sum_{i=1}^{K} f_i\,(x_i),$$

Where f_i – is the loss function on the i-th device [2]. Each device computes the local gradient ∇f_i on its data and updates its local weights wiw_iwi. The server then averages them to obtain a new global weight vector: w=1/K ∑_i^▒w_i [2][11]. This algorithm (FedAvg) is equivalent to parameter averaging under equal initial weights and significantly reduces traffic compared to transferring full datasets [11].

To address data heterogeneity, FedProx was introduced—an extension of FedAvg that adds a proximal regularizer: during optimization, a term
μ/2‖w-w^((t)) ‖^2, is added, limiting the divergence of local models and preventing severe drift under non-IID data distributions [12].

FL is classified into several types: horizontal FL (HFL), where devices share the same feature space but have different samples; vertical FL (VFL), where devices have data about the same entities but with different features; and federated transfer learning (Transfer FL) [1]. Each approach applies to specific data partitioning scenarios.

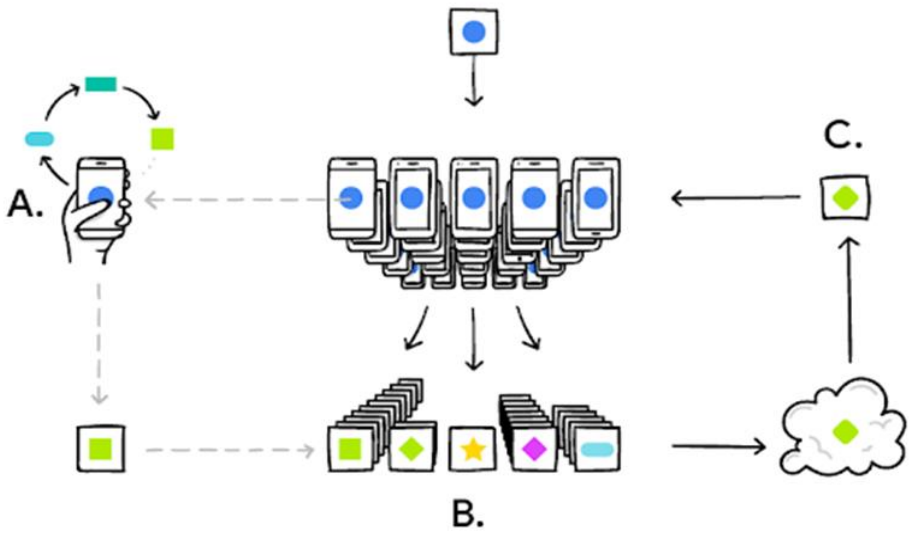

Figure 1. Architecture of centralized federated learning. Several client devices train local models on their own data and send updates (gradients or weights) to a central server for aggregation..

The architecture of FL is shown in Fig. 1. Clients (IoT devices) train the model locally and send either updated weights or computed gradients to the central server. The server collects these updates and forms a new global model. At the same time, the most private data (raw sensor records, request logs, etc.) remains on the device, reducing the risk of compromise. Publications note that this approach "minimizes the need to transmit sensitive data to central servers" [1] and "enables training personalized models without violating user privacy" [13]. However, FL also has its limitations: a large number of participants creates communication and synchronization overhead, and collaborative training under non-representative (non-IID) data may suffer from convergence issues. To optimize communication, compression techniques and adaptive client-selection schemes are introduced; and to ensure robustness, aggregation algorithms resilient to outliers (Byzantine-robust) are used.

3. Anonymization Methods and Differential Privacy

To enhance the privacy protection of IoT data, anonymization methods are employed. These include:

• pseudonymization (replacing unique identifiers with surrogate keys);
• removal of direct identifiers (names, addresses, etc.);
• partitioning and aggregation (group-level summarization of indicators);
• randomization, i.e., the addition of random noise.

For example, in k-anonymity, data records are grouped so that each anonymized record corresponds to at least k original records [3]. This prevents direct identification of a specific user. The drawback of k-anonymity is that with background knowledge of a particular attribute, one may still isolate a group. To address this, the concept of L-diversity was introduced, requiring that each group contain at least L distinct values of a sensitive attribute [3]. However, L-diversity also becomes vulnerable when the data distribution is highly skewed.

A simple example of randomization is adding independent noise to numerical data: if X is the original value, the published value is X+Y, where Y has a known distribution (e.g., normal or Laplace) [14]. The choice of noise parameters aligns with privacy requirements (see below).

Today, the gold standard of privacy protection is differential privacy (DP) [4]. An algorithm A is called ε -differentially private if for any two neighboring datasets D1 and D2 (differing by one record) and any output S, the following inequality holds:

$$\Pr[\mathcal{A}(D_1) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(D_2) \in S].$$

This means that the presence or absence of a single record has virtually no effect on the distribution of outputs [4]. The parameter ε (the privacy budget) reflects the level of protection: the smaller the ε, the greater the obfuscation of the results. One of the main implementations of DP is the Laplace mechanism: for a numerical function f(D), the published value is f(D + Y), where the noise Y is drawn from the Laplace distribution with a certain density. It has been shown that with an appropriate choice of scale, the mechanism satisfies differential privacy. Thus, adding Laplace noise ensures that the output of the

algorithm changes only minimally when a single record is removed from the database.

Method	Idea	Advantages	Disadvantages
Pseudonymization	Replacing explicit identifiers (IDs) with pseudonyms	Preserves data structure	Easily reversible if the key is known
k-anonymity	Grouping records so that each group contains ≥ k similar entries	Simple and efficient	Vulnerable to background knowledge
L-diversity	Requiring L different values of sensitive attributes within each group	Improved protection of sensitive data	May lack sufficient unique values
Differential privacy	Adding controlled noise (e.g., Laplace noise)	Strong mathematical privacy guarantee	Reduces accuracy (the smaller ε is, the stronger the noise)
Local DP	Noise is added on the client side before sending data to servers	Stronger protection before transmission	Significantly reduces data utility

Table 1. Comparison of the main data privacy protection methods.

4. Applications and Use Cases Across Different Industries

Healthcare. In medical IoT (wearable sensors, lab tests, electronic health records), collaborative learning is crucial, but strict patient privacy regulations create constraints. Federated learning allows several hospitals or devices to train a shared diagnostic model (for example, for disease detection based on images or ECG signals) without exchanging raw medical data [5]. For instance, one study demonstrated that the proposed Federated Agentic Learning scheme based on PhysioNet achieved an F1-score of 92.5% in detecting cardiac anomalies, outperforming traditional FL (~88.1%) while also providing “fairness” in training distribution [16]. Publications also describe FL applications in personalized medicine, MRI image analysis, and epidemic spread prediction.

Smart cities. In urban infrastructure, IoT devices generate massive volumes of data: surveillance video, traffic sensor readings, public transportation data. FL is used for jointly training computer vision and forecasting models. For example, many smartphones or street cameras can locally train object recognition models and send updated weights to a central cluster without revealing video data [6]. Horizontal federated learning (HFL) can be applied to improve speech or image recognition algorithms, as devices synchronously enhance the model through local training [6]. In addition, organizations possessing different fragments of city datasets (e.g., separate transportation companies) can collaborate to build global models for predicting transportation demand or road traffic density without exchanging raw data.

Energy sector. In smart grids, load and consumption forecasting is critical for system stability. Regional energy providers can use FL: each operator trains a model on its local consumption data and then aggregates results via a central server. This enables the creation of a unified forecasting model without transmitting private customer data. Such cross-

segment FL is described in [81] as an example of cooperation between regional networks for joint energy consumption forecasting [6]. Similarly, FL is applied in industrial IoT and smart home systems for equipment failure prediction, resource allocation optimization, and other scenarios where data privacy and real-time analytics are important.

Differential privacy and anonymization in practice. In all of the aforementioned cases, FL methods are often combined with additional anonymization. For example, before sending updates, clients may apply local differential privacy by adding noise to gradients (Local DP), so that the server receives only noised parameters. This further complicates reconstruction of original data from updates. Thus, the combination of FL and DP provides multi-layer protection: data never leaves the device in raw form, and what is transmitted is additionally perturbed with statistically controlled noise.

Conclusion

The presented article demonstrates that ensuring privacy in IoT requires a comprehensive approach. Federated learning significantly reduces risks associated with centralized data collection by enabling decentralized model training [1][2]. However, FL is not a complete solution: due to data heterogeneity and network constraints, its limitations must be considered alongside additional privacy safeguards. Anonymization techniques (pseudonymization, k-anonymity) and especially differential privacy offer formal guarantees and practical implementations (e.g., Laplace noise) [4][15]. The application of these approaches in real-world IoT scenarios (healthcare, smart cities, energy, and others) is already producing concrete results: models become “privacy-friendly” without compromising quality. Future research will involve further optimization of aggregation, communication, and privacy mechanisms—for example, developing new heterogeneous FL methods (FedDyn, HyFDCA, etc.) and DP implementations tailored to the specifics of IoT data, enabling even more effective trade-offs between model accuracy and privacy preservation.

References

1. Al-Tamimi M.M., Hassan M.B., Abbas S.A. Federated Learning (FL) – A Review // Izvestiya SPbGETU “LETI”. 2024. Vol. 17, No. 5. pp. 74–82. DOI: 10.32603/2071-8985-2024-17-5-74-82.

2. Borisov A.V., Bosov A.V., Ivanov A.V. Methods and Algorithms for Personal Data Anonymization // Programming. 2023. No. 4. pp. 58–74.

3. IoT Security: How a Smart Home Can Harm Privacy // Data Privacy Office, October 10, 2024. URL: https://data-privacy-office.com/iot-security-how-smart-homes-can-harm-privacy/ (accessed: 05.11.2025).

4. Differential Privacy. Wikipedia. URL: https://ru.wikipedia.org/wiki/Дифференциальная_приватность (last accessed: 05.11.2025).

5. Federated Learning. Wikipedia. URL: https://en.wikipedia.org/wiki/Federated_learning (last accessed: 05.11.2025).

6. Dritsas E., Trigka M. Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications // Journal of Sensor and Actuator Networks. 2025. Vol. 14, No. 1, 9. DOI: 10.3390/jsan14010009.

7. izv.etu.ru

8. https://izv.etu.ru/assets/files/izvestiya-5-2024-74-82.pdf

9. Federated Learning – Wikipedia https://en.wikipedia.org/wiki/Federated_learning

10. Programming. Issue 4, 2023 https://sciencejournals.ru/view-article/?j=program&y=2023&v=0&n=4&a=Program2304004Borisov

11. Differential Privacy — Wikipedia

https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D1%84%D1%84%D0%B5%D1%80%D0%B5%D0%BD%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D0%BF%D1%80%D0%B8%D0%B2%D0%B0%D1%82%D0%BD%D0%BE%D1%81%D1%82%D1%8C

12. Federated Learning for IoT and Edge Computing: Horizontal and Vertical Models

https://guardora.ru/blog/federativnoye-obucheniye-v-iot/

13. IoT Security: How a Smart Home Can Harm Privacy – Data Privacy Office

https://data-privacy-office.com/iot-security-how-smart-homes-can-harm-privacy/

14. Federated Learning for IoT Applications | Request PDF

https://www.researchgate.net/publication/354435192_Federated_Learning_for_IoT_Applications

15. Types of FL Based on Data Partitioning, System Architecture, and Operational Strategies | Scientific Diagram

https://www.researchgate.net/figure/Types-of-FL-based-on-data-partitioning-system-architecture-and-operational-strategies_fig2_388311969

16. Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications

https://www.mdpi.com/2224-2708/14/1/9