International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 7, Issue 1, 29-33, 2026
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V7I1P107

Original Article

Privacy-Preserving Technologies in Edge Analytics: Architectures, Mechanisms, and Performance Frontiers

Naresh Kalimuthu
Independent Researcher, USA.

Received On: 29/11/2025 ***Revised On: 30/12/2025*** ***Accepted On: 08/01/2026*** ***Published On: 20/01/2026***

Abstract - *The shift of data analytics from centralized clouds to the network edge enhances latency and bandwidth efficiency, but it also raises significant privacy concerns. As sensitive biometric, behavioral, and operational data are gathered on resource-limited IoT devices, traditional security measures become outdated. This paper explores the main challenges in securing edge analytics, emphasizing the "Privacy-Utility Paradox" and the risks of inference attacks. It provides a detailed analysis of distributed learning architectures—Federated Learning (FL), Split Learning (SL), and Swarm Learning—and assesses cryptographic techniques such as Differential Privacy (DP), Homomorphic Encryption (HE), and Secure Multi-Party Computation (SMPC). Based on recent quantitative research, we propose mitigation strategies that balance accuracy, latency, and energy use, guiding the development of compliant and trustworthy edge intelligence systems.*

Keywords - *Edge Computing, Privacy-Preserving Analytics, Federated Learning, Differential Privacy, Split Learning, Swarm Learning, Homomorphic Encryption, Non-IID Data.*

1. Introduction

The digital ecosystem is shifting from centralized cloud data silos to decentralized edge processing. This change is driven by the surge in Internet of Things (IoT) devices, including autonomous vehicles, industrial sensors, and wearable health tech, which generate vast amounts of data. These large data volumes render traditional "store-and-send" approaches obsolete due to bandwidth limits, latency issues, and high transmission costs. As analytics approaches data sources, security risks grow, particularly when sensitive user information is stored on devices that are physically accessible and resource-constrained, posing significant privacy and security challenges.

In this context, Privacy-Preserving Technologies (PPTs) in edge analytics have become fundamental architectural elements rather than optional security features. The collection of biometric data, location histories, and behavioral insights at the network edge poses significant risks, as data leaks could lead to severe personal consequences and breach strict global regulations like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). These laws emphasize principles such as data minimization and purpose limitation, prompting system designers to adopt a "Privacy by Design" strategy that integrates protection mechanisms into learning algorithms and communication protocols from the outset.

Securing edge analytics involves addressing the "Privacy-Utility Paradox," where enhancing data protection, such as noise injection or encryption, often diminishes data utility for machine learning or causes delays during real-time tasks. Techniques such as Federated Learning (FL) mitigate this by exchanging model updates rather than raw data. Nonetheless, recent research indicates that gradients can still expose sensitive information through reconstruction attacks, necessitating additional safeguards like Differential Privacy (DP), Homomorphic Encryption (HE), and Secure Multi-Party Computation (SMPC).

2. Key Research Challenges

To enable secure and effective edge analytics, three key challenges must be addressed. These arise from the inherent traits of edge environments and the characteristics of distributed learning.

2.1. Challenge I: Statistical Data Heterogeneity (Non-IID Data)

Unlike centralized data centers, where data can be shuffled to be Independent and Identically Distributed (IID), edge data is inherently Non-IID. Data distributions differ considerably across devices because of diverse user behaviors, geographic locations, and device usage patterns.

- Impact: This statistical heterogeneity leads to differences between clients' local optimization goals and the overall global objective. Standard aggregation methods such as FedAvg often struggle to converge or exhibit "client drift," hindering the global model's ability to generalize well for individual users.
- Hierarchical Complexity: In complex networks, "hierarchical non-IIDness" emerges, with devices connected to the same edge server sharing contextual similarities distinct from those in other clusters, making model aggregation more difficult.

2.2. Challenge II: Inference and Reconstruction Vulnerabilities

While distributed architectures like FL prevent sharing raw data, transferring model updates (gradients) can serve as a secondary channel for potential information leaks.

- Gradient Leakage: Adversaries with access to the central server can analyze shared gradients to reconstruct the original training data. Methods such as Deep Leakage from Gradients (DLG) can recover detailed images or readable text from private inputs by repeatedly optimizing dummy data to match the received gradients.
- Membership Inference: Attackers can determine whether a specific person's data was used to train a model, potentially revealing sensitive information such as medical diagnoses.
- Property Inference: Beyond targeting individual records, adversaries can perform property inference attacks to uncover overall traits of the training dataset, such as the percentage of a specific demographic or the presence of a certain device class. This type of "meta-privacy" breach is especially problematic in regulated industries such as finance or healthcare, where disclosing population-level information can breach confidentiality agreements or regulatory requirements, even if individual data points stay private and secure.

2.3. Challenge III: Resource Constraints at the Edge

Edge nodes and IoT devices face strict limitations in computation, storage, energy, and network bandwidth.

- Computational Bottlenecks: Advanced privacy mechanisms such as Homomorphic Encryption (HE) and Secure Multi-Party Computation (SMPC) require significant computational power. Using them on battery-powered devices can deplete resources and cause unacceptable delays in real-time processing applications.
- Communication Overhead: High-dimensional models require transmitting millions of parameters. In cellular or Long Range Wide Area Network (LoRaWAN) networks with limited bandwidth, the cost of frequent model updates, especially with encrypted payloads, can overload the network and delay convergence.

3. Distributed Learning Architectures

The core principle of privacy-preserving edge analytics is shifting computation closer to where the data resides.

3.1. Federated Learning (FL)

Federated Learning (FL) is the prevailing standard for privacy-preserving edge intelligence. In FL, a central server sends out a global model to edge clients, who train it using their local private data and send back model updates (gradients). The server then combines these updates, usually through FedAvg, to improve the global model.

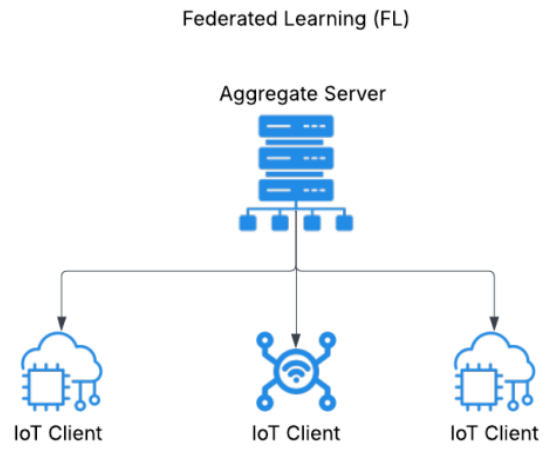

Fig 1: Federated Learning Architecture for Distributed IoT Clients

3.2. Split Learning (SL) and Federated Split Learning (SFL)

Split Learning (SL) divides the neural network between the client and server. The client runs the initial layers and sends the intermediate activations, known as "smashed data," to the server, which then completes the forward pass.

- **Trade-offs:** SL decreases the computational burden on clients, making it suitable for low-power IoT devices. Nonetheless, it causes significant communication overhead due to frequent activation sharing across data batches, potentially overwhelming bandwidth in standard cellular networks. Federated Split Learning (SFL) merges the parallel training approach of FL with the model splitting technique of SL to better balance these challenge factors.

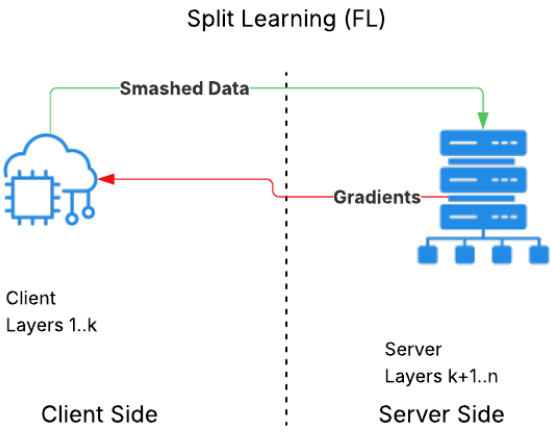

Fig 2: Split Learning Architecture with Client–Server Model Partitioning

3.3. Swarm Learning (SL)

Swarm Learning removes the need for a central aggregator by employing blockchain smart contracts to coordinate peer-to-peer model sharing.

Advantages: It eliminates the single point of failure and the risk posed by "honest-but-curious" servers. The blockchain

ensures an immutable audit trail, safeguarding data sovereignty and preventing model poisoning. This is especially useful in consortia such as banking or genomics, where participants are competitors and do not have a trusted third party.

4. Cryptographic and Statistical Privacy Mechanisms

4.1. Differential Privacy (DP)

DP introduces controlled noise into data or model updates to mask individual information contributions.

- Local DP (LDP): Noise is added at the edge device, providing maximum privacy but significantly reducing model accuracy because of the high noise variance required.
- Central DP (CDP): The server adds noise after aggregation, improving utility but requiring trust in the server.

4.2. Homomorphic Encryption (HE)

HE enables calculations on encrypted data. In federated learning (FL), Paillier encryption allows a server to sum encrypted weight updates without decrypting them. Although theoretically secure, HE incurs significant computational overhead and data expansion, often greatly increasing training time and hindering real-time edge applications.

4.3. Secure Multi-Party Computation (SMPC)

SMPC protocols, such as Secure Aggregation, enable users to collaborate on computing updates while ensuring that the server accesses only the final aggregate. Common techniques include pairwise masking, where random masks cancel each other out when summed. Recent protocols, such as e-SeaFL, have improved this process for edge computing, enabling secure aggregation in a single communication round and reducing overhead compared to traditional methods and schemes.

4.3. Lightweight Cryptography (LWC)

LWC bridges the gap between strict security needs and the limited resources of edge sensors. NIST-standardized Lightweight Cryptography (LWC) is crucial here. Algorithms like ASCON offer authenticated encryption with low computational demands, tailored for environments where primitives like AES or SHA-3 are too energy-consuming. Incorporating LWC ensures sensitive data remains secure on the sensor before entering a distributed learning process, without depleting the battery or causing delays or bottlenecks.

5. Handling Data Heterogeneity and Personalization

To tackle the Non-IID challenge, Personalized Federated Learning (pFL) customizes models for individual users or edge devices nodes.

- PHE-FL (Personalized Hierarchical Edge-enabled FL) specifically tackles "hierarchical non-IIDness" — variations across edge servers. Experiments demonstrate that PHE-FL can reach up to 83% higher accuracy than standard FL in complex hierarchical networks by tailoring aggregation to each edge's unique data clusters.
- PPFL: In medical settings, PPFL has shown higher accuracy than local training (0.941 versus 0.875), enabling hospitals to collaborate on mortality prediction models while maintaining site-specific data features.

6. Quantitative Analysis: Performance and Trade-offs

Table 1: Comparative Analysis of Distributed Privacy-Preserving Learning Paradigms (Federated, Split, and Swarm Learning)

Metric	Federated Learning (FL)	Split Learning (SL)	Swarm Learning
Privacy	High (with DP/SMPC)	Moderate (activations can leak)	Very High (Blockchain + no central server)
Comm. Overhead	Medium (Updates per round)	High (Activations per batch)	Medium (Peer-to-peer sync)
Comp. Load (Client)	High (Full backward pass)	Low (Partial model)	High (Training + Blockchain consensus)
Accuracy (Non-IID)	Degrades (needs personalization)	Stable	High (Robust to poisoning)

Key Findings:

- Latency: Using HE for secure aggregation can raise training latency by 23x to 93x compared to plaintext training.
- Accuracy: Personalized algorithms, such as Top-k Shuffled Differential Privacy Federated Learning (TopkSDP-FL), achieve approximately 3.6% higher accuracy on CIFAR-10 (Canadian Institute For Advanced Research) than standard Federated Averaging (FedAvg) in heterogeneous settings.

7. Strategic Mitigations and Validation Case Studies

Addressing the key challenges necessitates targeted mitigation strategies. Below, we link each strategy to a specific challenge and validate it with recent case studies.

7.1. Mitigation for Data Heterogeneity: Hierarchical Personalization

- The Strategy: Implement Personalized Hierarchical Edge-enabled Federated Learning (PHE-FL). Rather than applying a single global model to all users, this method develops intermediate "edge models" tailored to the data distribution of each local cluster (e.g., a specific hospital or city district).
- How It Works: The architecture recognizes "hierarchical non-IIDness' and conducts partial aggregation at the edge server to capture local

details before transmitting a more generalized update to the cloud. This approach helps prevent "client drift," where local models deviate significantly from the global model average.

Case Study Validation:

- PHE-FL Experiments: In scenarios with complex hierarchical non-IID data, PHE-FL reached accuracy levels up to 83% higher than traditional FedAvg. It also significantly reduced fluctuations in accuracy, demonstrating greater stability across conditions.
- Medical Mortality Prediction (PPFL): In a study forecasting in-hospital mortality across various hospitals, the PPFL framework attained an accuracy of 0.941, exceeding both local training results (0.875) and conventional Federated Learning. This demonstrates that personalization is crucial for high-stakes decision-making in heterogeneous environments.

7.2. Mitigation for Inference Attacks: Verifiable Secure Aggregation & Decentralization

- The Strategy: Implement Verifiable Secure Aggregation protocols and use Swarm in environments lacking trust.
- How It Works: Secure aggregation employs cryptographic masking by adding random noise that cancels out when summed, preventing the server from viewing individual updates. Swarm Learning obliterates the central server and relies on blockchain smart contracts to manage the model, removing the single point of failure and the "honest-but-curious' threat aggregator.

7.2.1. Case Study Validation:

- e-SeaFL Protocol: Traditional secure aggregation can be slow. The e-SeaFL protocol was designed to perform secure aggregation in just one communication round. Experiments demonstrated that it delivers a tenfold efficiency boost compared to leading protocols when handling large gradient vectors, making high-security federated learning (FL) practical even on regular cellular networks.
- HPE Swarm Learning: In fields like financial fraud detection and genomics, HPE's Swarm Learning framework enables organizations to collaborate without exchanging raw data. By eliminating the central aggregator, it prevents any one entity from dominating the model, encouraging cooperation even among competitors, such as banks, who typically refrain from sharing data due to regulatory and competitive concerns.

7.3. Mitigation for Resource Constraints: Federated Split Learning (SFL)

- The Strategy: Implement Federated Split Learning (SFL) for devices with constrained battery life or computational capacity, such as drone wearables.
- How It Works: The significant computational load of training a deep neural network is divided. The edge device handles only the initial layers (feature extraction), while a more powerful edge server manages the remaining layers. SFL integrates this approach with parallel training to accelerate the process convergence.

7.3.1. Case Study Validation:

- Human Activity Recognition (HAR): A study on Human Activity Recognition with wearables found that the FSL-DP (Federated Split Learning with Differential Privacy) framework outperformed traditional federated learning across both accuracy and loss metrics. Importantly, it also provided faster communication times per training round, demonstrating that dividing the model is a practical approach for resource-limited IoT devices requiring real-time processing responsiveness.
- Edge Video Analytics (FedEVA): To manage high-bandwidth video data on limited devices, FedEVA employs a simple perturbation method (noise addition) rather than complex encryption. This approach enables privacy protection while avoiding the significant latency typically caused by cryptographic processes on video streams.

8. Conclusion

Privacy-preserving edge analytics has evolved from a theoretical idea to an essential part of infrastructure. Although Federated Learning remains the primary approach, the "one-size-fits-all" method struggles with Non-IID data and limited resources. The future points to adaptive, hierarchical systems that combine lightweight cryptography (SMPC) with customized learning models. By leveraging architectures such as Swarm Learning for trust and Split Learning for efficiency, organizations can harness the potential of edge data while ensuring strict adherence to global privacy standards.

References

[1] M. Abadi et al., "Deep Learning with Differential Privacy," in Proc. of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308–318. doi: 10.1145/2976749.2978318

[2] P. Voigt and A. von dem Bussche, The EU General Data Protection Regulation (GDPR): A Practical Guide, Springer International Publishing, 2017. doi: 10.1007/978-3-319-57959-7

[3] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," in IEEE Symposium on Security and Privacy (SP), 2017, pp. 3–18. doi: 10.1109/SP.2017.41

[4] K. Bonawitz et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in Proc. of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191. doi: 10.1145/3133956.3133982

[5] Richard Kim, Max Kleiman-Weiner, Andrés Abeliuk, Edmond Awad, Sohan Dsouza, Joshua B. Tenenbaum, and Iyad Rahwan. 2018. A Computational Model of Commonsense Moral Decision Making. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18). Association for Computing

Machinery, New York, NY, USA, 197–203. https://doi.org/10.1145/3278721.3278770
[6] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated Learning with Non-IID Data. *ArXiv*. https://doi.org/10.48550/arXiv.1806.00582
[7] Vepakomma, P., Gupta, O., Swedish, T., & Raskar, R. (2018). Split learning for health: Distributed deep learning without sharing raw patient data. *ArXiv*. https://arxiv.org/abs/1812.00564
[8] Zhu, L., Liu, Z., & Han, S. (2019). Deep Leakage from Gradients. *ArXiv*. https://arxiv.org/abs/1906.08935
[9] A. Salem et al., "ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models," in Network and Distributed System Security Symposium (NDSS), 2019. doi: 10.14722/ndss.2019.23119
[10] Singh, A., Vepakomma, P., Gupta, O., & Raskar, R. (2019). Detailed comparison of communication efficiency of split learning and federated learning. *ArXiv*. https://arxiv.org/abs/1909.09145
[11] T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," in IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, May 2020, doi: 10.1109/MSP.2020.2975749.
[12] W. Y. B. Lim et al., "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 2031–2063, 2020. doi: 10.1109/COMST.2020.2986024
[13] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of Edge Computing and Deep Learning: A Comprehensive Survey," IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 869–904, 2020. doi: 10.1109/COMST.2020.2970550
[14] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2018). Federated Optimization in Heterogeneous Networks. *ArXiv*. https://arxiv.org/abs/1812.06127
[15] C. Thapa, M. A. P. Chamikara, and S. Camtepe, "SplitFed: When Federated Learning Meets Split Learning," arXiv preprint arXiv:2004.12088, 2020. doi: 10.48550/arXiv.2004.12088
[16] P. Kairouz et al., "Advances and Open Problems in Federated Learning," Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021. doi: 10.1561/2200000083
[17] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, *216*, 106775. https://doi.org/10.1016/j.knosys.2021.106775
[18] H. Yin, A. Mallya, A. Vahdat, J. M. Alvarez, J. Kautz and P. Molchanov, "See through Gradients: Image Batch Recovery via GradInversion," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 16332-16341, doi: 10.1109/CVPR46437.2021.01607.
[19] V. Turina, Z. Zhang, F. Esposito, and I. Matta, "Federated or Split? A Performance and Privacy Analysis of Hybrid Split and Federated Learning Architectures," in IEEE International Conference on Cloud Computing (CLOUD), 2021. doi: 10.1109/CLOUD49388.2021.00069
[20] S. Warnat-Herresthal et al., "Swarm Learning for Decentralized and Confidential Clinical Machine Learning," Nature, vol. 594, pp. 265–270, 2021. doi: 10.1038/s41586-021-03583-3
[21] Wu, Yulei & Choo, Kim-Kwang Raymond & Yang, Laurence. (2019). IEEE Access Special Section Editorial: Internet-of-Things Big Data Trust Management. IEEE Access. 7. 65223-65227. 10.1109/ACCESS.2019.2915489.
[22] J. Zhang et al., "Secure Aggregation Techniques in Federated Learning," IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1–15, 2023. doi: 10.1109/TIFS.2023.3235652
[23] Nguyen, T. T., Huynh, T. T., Ren, Z., Nguyen, T. T., Nguyen, P. L., Yin, H., & Nguyen, Q. V. (2024). A Survey of Privacy-Preserving Model Explanations: Privacy Risks, Attacks, and Countermeasures. *ArXiv*. https://arxiv.org/abs/2404.00673
[24] Behnia, R., Riasi, A., Ebrahimi, R., Chow, S. S., Padmanabhan, B., & Hoang, T. (2023). Efficient Secure Aggregation for Privacy-Preserving Federated Machine Learning. *ArXiv*. https://arxiv.org/abs/2304.03841
[25] Tae Hyun Kim, Jae Yong Yu, Won Seok Jang, Sun Cheol Heo, MinDong Sung, JaeSeong Hong, KyungSoo Chung, Yu Rang Park, PPFL: A personalized progressive federated learning method for leveraging different healthcare institution-specific features, iScience, Volume 27, Issue 10, 2024. https://doi.org/10.1016/j.isci.2024.110943.
[26] Xiao, D., Fan, X., & Chen, L. (2025). Top-k Shuffled Differential Privacy Federated Learning for Heterogeneous Data. Sensors, 25(5), 1441. https://doi.org/10.3390/s25051441
[27] Giehl, Alexander & Schneider, Peter & Busch, Maximilian & Schnoes, Florian & Kleinwort, Robin & Zaeh, Michael. (2019). Edge-computing enhanced privacy protection for industrial ecosystems in the context of SMEs. 1-6. 10.1109/CMI48017.2019.8962138.
[28] J. Liu et al., "FedEVA: Federated Learning for Edge Video Analytics," arXiv preprint arXiv:2401.08872, 2024. doi: 10.48550/arXiv.2401.08872.