International Journal of Artificial Intelligence and EvolveTech
ISSN: 0000-0000
Volume 1, Issue 1 | PP 18-25| www.ijaiet.com |

Federated Learning at the Edge: A Survey of Privacy Preserving Architectures for IoT Networks

Dr. Sandeep Kumar

Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India City-Nagpur, Country-India, Pincode-440008

*Corresponding author
Dr. Sandeep Kumar
Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India City-Nagpur, Country-India, Pincode-440008
E-Mail: er.sandeepsahratia@gmail.com Orcid : https://orcid.org/0000-0002-4752-7884

Received : 24-06-2026 | Revised : 17-08-2026 | Accepted : 24-08-2026 | Published : 31-08-2026

Abstract

As the Internet of Things (IoT) continues to grow, the majority of data generated is being collected from the network edge, which is too burdensome to be sent to a central server for model training due to bandwidth, latency, energy, and most importantly privacy and regulatory considerations. One alternative is federated learning (FL), where devices use raw data to train a shared model without the data ever being sent away from the device. But when deployed at the edge, FL has to face the limits of resource-constrained hardware, unreliable networks, and especially, not being private by default: the model updates that devices exchange can expose a lot of information about the underlying data. This survey provides an overview of the architectures and mechanisms that enable FL to be privacy preserving in IoT environments. It outlines the field along four axes: taxonomy of aggregation topologies and privacy strategies; accuracy, communication, and energy trade-offs of the privacy enhancing mechanisms (differential privacy, secure aggregation, homomorphic encryption, secure multiparty computation, and trusted execution); threat models and attacks that these mechanisms need to withstand (gradient inversion, membership inference, poisoning attacks); and a layered reference architecture spanning the device, edge, and cloud tier, with dimensions along which this architecture is evaluated. The survey shows that there is no one mechanism that is both private, accurate and cheap and the key remaining open problem is composing mechanisms under the small budgets of IoT hardware. Future work priorities involve adaptive privacy budgeting, communication and energy efficient secure aggregation, robustness to poisoning and standardised and reproducible evaluation.

Keywords: federated learning; edge computing; privacy preserving machine learning; Internet of Things; differential privacy; secure aggregation; distributed learning

1. Introduction

Today, the Internet of Things (IoT) extends to tens of billions of sensors, wearables, vehicles and industrial controllers generating constant streams of data at the network edge. Traditionally, to derive value from this data using machine learning, it has been necessary to move it to a central server or cloud to be used for model training. This centralised paradigm is putting pressure on four fronts in the IoT scale: bandwidth cost of data transfer, latency, energy consumption on resource-limited devices, and most importantly privacy and regulatory implications of collecting

sensitive personal information in one location [1]. Wholesale data centralisation is now a burden, rather than a benefit, due to data protection regimes and increasing user expectations.

Federated learning (FL) is a rephrasing of the problem. The introduction of FL [2] introduced a practical training scheme in which multiple clients learn the same model while keeping their raw data on device, and only sending the updates to the server which aggregates them to create a better model. This makes the natural fit for IoT where computation is pushed to where the data is created and sensitive records never leave the device [3], [4]. FL can be deployed on the edge to leverage the intermediate edge or gateway nodes to aggregate updates towards the device, offloading the load from the wide area network [5], [6].

There are two problems here. Edge FL faces many challenges in the real world, such as: devices are diverse and not always on hand, the compute, memory and energy cost of the devices is limited, the network among the devices can be slow and unreliable [7]. Second, subtlely, FL is not private by default. Despite the fact that the raw data remains local, the data sent to the model updates are functions of the data and can leak a lot of information about them: attacks can recover training samples from shared gradients [8]; attacks can reveal whether a record was used or exhibit sensitive attributes of the training set [9], [10]. To realize true privacy, however, it is necessary to take deliberate architectural and cryptographic steps, added to the basic FL protocol.

This survey reviews those measures for the IoT edge. Four questions for the organisation of the design space are posed: (RQ1) which architectural patterns and means of controlling privacy form the design space; (RQ2) which mechanisms can enhance privacy, and what are their accuracy, communication and energy costs; (RQ3) what are the threat models and attacks they are required to defend against; (RQ4) how the resulting systems can be organised into a reference architecture and evaluated. It provides a taxonomy of privacy preserving FL approaches (Figure 1), a comparative analysis of mechanisms (Table 3), a consolidated account of threats, and a layered device – edge – cloud reference architecture (Figure 2), along with a prioritised research agenda (Table 4).

2. Review Approach

2.1 Method. It is a structured narrative survey that was carried out on a pre-established protocol that set the research questions, the resources of literature, the inclusion criteria and the procedure for extracting and synthesizing the data prior to the screening. A narrative rather than a meta analytic synthesis is appropriate, due to the methodologically diverse primary literature dealing with cryptography, systems and machine learning, and the incommensurable results.

2.2 Scope and eligibility. The survey looks at both primary and secondary literature in which Federated or Collaborative Learning on device is combined with an explicit privacy or security mechanism in the Edge or IoT context. Purely centralised machine learning work and FL work that does not involve the privacy dimension fall out of scope. Complete inclusion/exclusion criteria are detailed in Table 1.

2.3 Sources of information and search strategy. Backward and forward citation chasing from selected surveys was combined with searching five databases (IEEE Xplore, ACM Digital Library, Scopus, Web of Science and arXiv for pre publication). Three concept blocks (federated learning, privacy and security, and the edge/IoT setting) were used in combination with the Boolean operators used in the specific databases; the representative terms used are shown in Table 2.

2.4 Selection and extraction. De-duplicated and screened against eligibility criteria retrieved records should be against eligibility criteria, in two stages: title and abstract, then full text and the number retained at each stage should be recorded in the final manuscript. A common template was used for each work retained to provide a snapshot of the aggregation topology, privacy mechanism(s), threat model(s) addressed, and that considered edge/IoT constraints, and evaluations approach.

2.5 Synthesis.

The findings were thematised on the basis of the four research questions. Architectural patterns have been grouped into the taxonomy in Figure 1, privacy mechanisms compared in Table 3, system architecture in Figure 2 is used to position the privacy mechanisms at different system layers, and the dimensions for evaluation of such systems are collected in Table 4.

Table 1. *Scope and eligibility criteria for the survey.*

Dimension	Inclusion criteria	Exclusion criteria
Focus	Federated / collaborative on device learning with an explicit privacy or security mechanism	Centralised ML; FL with no privacy dimension
Setting	Edge, gateway, mobile, or IoT deployment context	Purely cloud/datacentre training with no edge relevance
Contribution	Architecture, mechanism, attack, defence, or survey	Marketing or non technical opinion pieces
Type	Peer reviewed papers and well cited preprints	Non archival abstracts and posters
Language	Published in English	Full text unavailable in English
Window	Modern FL era (2016–present)	Pre dates the federated learning literature

Table 2. *Search concept blocks and representative terms; syntax was adapted to each database.*

Concept block	Representative search terms (OR combined within block; blocks AND combined)
A · Federated learning	"federated learning" OR "collaborative learning" OR "on device training" OR "decentralized learning" OR FedAvg
B · Privacy & security	privacy OR "differential privacy" OR "secure aggregation" OR "homomorphic encryption" OR "secure multiparty" OR "trusted execution" OR poisoning OR inference
C · Edge / IoT	edge OR "edge computing" OR IoT OR "Internet of Things" OR gateway OR mobile OR "resource constrained"
D · Model (optional)	"deep learning" OR "neural network" OR "machine learning"

3. Findings

3.1 The convergence of federated learning, edge computing, and IoT. This literature is the intersection of three trends. Edge computing moves storage and computation closer to the data source to reduce latency and backhaul traffic [1]; FL moves model training closer to the data owners, who have constrained hardware resources with which this must operate; and IoT provides both the data and the constrained hardware on which this must be done. Their convergence will allow for local, low latency intelligence that will be aware of data locality, but also it will collapse all the hard constraints in one place: models will be trained in many weak, unreliable, and non identically-distributed clients, with protection against adversaries who will learn the updates being passed. The following section addresses the four research questions, in turn.

3.2 RQ1 A taxonomy of privacy preserving FL for IoT. The design space separates into two orthogonal aspects: aggregation of updates, and enforcement of privacy. On the first one, aggregation topology can be the centralised star, where a single server gathers all updates from the clients, or can be hierarchical, where the aggregated clients form a hierarchy, with edge aggregators sitting between devices and the cloud to reduce the amount of WAN traffic and to take advantage of locality [5], or full decentralisation, with clients communicating peer to peer directly with each other and/or the cloud, thus eliminating the central server as a single point of trust and failure. On the second, privacy enforcement methods include perturbation-based methods where noise is added to the data to modify its content; cryptographic methods for masking individual updates while preserving the sum of the updates; and trust and robustness methods, which depend on protected hardware or filtering out malicious contributions of updates. These are arranged in a working taxonomy in figure 1, and the mechanisms in the middle branch are discussed in detail in section 3.3.

International Journal of Artificial Intelligence and EvolveTech
ISSN: 0000-0000
Volume 1, Issue 1 | PP 18-25| www.ijaiet.com |

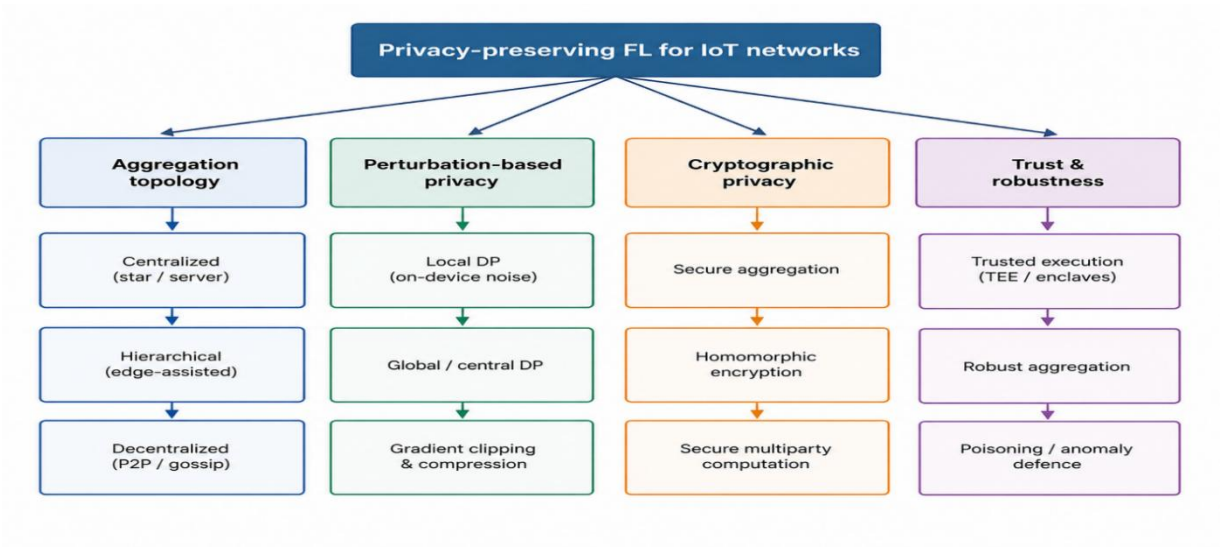

Figure 1. *A taxonomy of privacy preserving federated learning approaches for IoT networks.*

3.3 RQ2 Privacy enhancing mechanisms and their trade offs. There are 5 major mechanism families. Differential Privacy (DP): A calibrated noise is added to updates so that adding or removing any record is statistically hidden and is quantified by a privacy budget ε [11], [12]. Utilizing the central variant requires more noise to achieve the same guarantee than the local variant, which widens the utility gap, while it is cheap and composable, suitable for IoT, but noise deteriorates the model accuracy [13], [14]. Secure aggregation is a lightweight cryptographic masking approach that allows the server to learn only the sum of the client update (not any specific one) with moderate overhead that is well suited for mobile or IoT applications [15]. The homomorphic encryption (HE) method enables the aggregation of ciphers directly on the ciphertext, which is good for confidentiality, but requires high computational and ciphertext size costs for constrained devices [16]. Secure multiparty computation (SMPC) distributes trust between the parties so that each party doesn't see the data of the other parties and the communication cost is proportional to the number of parties. The trusted execution environments (TEEs) approach is to move the problem from cryptography to hardware trust and availability in the form of enclaves. These mechanisms are complementary to each other in providing statistical guarantees and update confidentiality (DP for statistical guarantees or secure aggregation/DP for integrity and robust aggregation), and are often used together. Their properties are summarised in Table 3.

Table 3. *Comparative taxonomy of privacy and integrity enhancing mechanisms for federated learning at the edge.*

Mechanism	Guarantee	Main overhead	Threat addressed	Edge / IoT fit	Key limitation
Local DP	Statistical (ε per client)	Accuracy loss	Inference from updates	High	Large utility loss at small ε
Central DP	Statistical (ε global)	Accuracy loss	Inference from model	High	Requires trusted aggregator
Secure aggregation	Update confidentiality	Moderate crypto / comms	Server sees single updates	High	Dropout handling; sum only
Homomorphic encryption	Update confidentiality	High compute / ciphertext size	Server / eavesdropper	Low–moderate	Costly on constrained devices
Secure MPC	Update confidentiality	Comms grows with parties	Curious participants	Moderate	Scales poorly with clients

International Journal of Artificial Intelligence and EvolveTech
ISSN: 0000-0000
Volume 1, Issue 1 | PP 18-25| www.ijaiet.com |

Mechanism	Guarantee	Main overhead	Threat addressed	Edge / IoT fit	Key limitation
Trusted execution (TEE)	Hardware isolation	Enclave constraints	Server side leakage	Moderate	Hardware trust; side channels
Robust aggregation	Integrity	Aggregation overhead	Poisoning / backdoors	High	Trades off with heterogeneity

3.4 RQ3 Threat models and attack surfaces. There are several distinct privacy mechanisms in the FL literature, and they have meaning only when there is a declared adversary. An honest but curious server follows the protocol, but attempts to deduce the client's data from the information made available to it via the updates.Secure aggregation and HE target specifically this adversary, an honest but curious server. Powerful aggregation and anomaly detection can help prevent or limit the impact of poisoning and backdoor attacks by malicious clients, which involve the submission of malicious updates that affect the global model or introduce hidden behaviours [17]. On the inference side, gradient inversion attacks infer training inputs from common gradients, and sometimes do so in surprisingly accurate ways [8]; membership inference and property inference attacks can tell if a specific record was part of the training set, or can infer population level properties [9], [10]. The threats to the system that are mentioned repeatedly in various surveys of the area are: intermittent participation, non IID data and limited defence [18], [19]. Good designs start by explicitly naming the adversary and then writing mechanisms to match, not believing that "keep the raw data local" is enough.

3.5 RQ4 A reference architecture and its evaluation. Putting it all together: FL in an edge setting with privacy protection can be seen as a tri-layer model (Figure 2). At the device tier, constrained IoT nodes learn from their own local data and use on device protections like gradient clipping and local DP noise prior to sending out data. At the edge/aggregator tier, nodes such as gateway or base station nodes can do secure or partial aggregation, select which clients should be involved in each round, and can aggregate in a TEE absorbing the data that would otherwise travel through the wide area network. The global model is aggregated, versioned and redistributed at the cloud/coordinator tier, and training rounds are orchestrated. The model updates are encrypted or noised from bottom to top while the global model is passed from top to bottom. The concerns that cut across the three tiers are the ones for feasibility: communication efficiency, energy budget, device heterogeneity, the privacy budget ε, robustness to poisoning, and scalability. Testing the accuracy of such a system is therefore not enough; the dimensions that are required to evaluate such a system are listed in Table 4. There is a general lack of literature that optimises more than one of these dimensions, and that reports on non standard benchmarks, making cross study comparisons difficult.

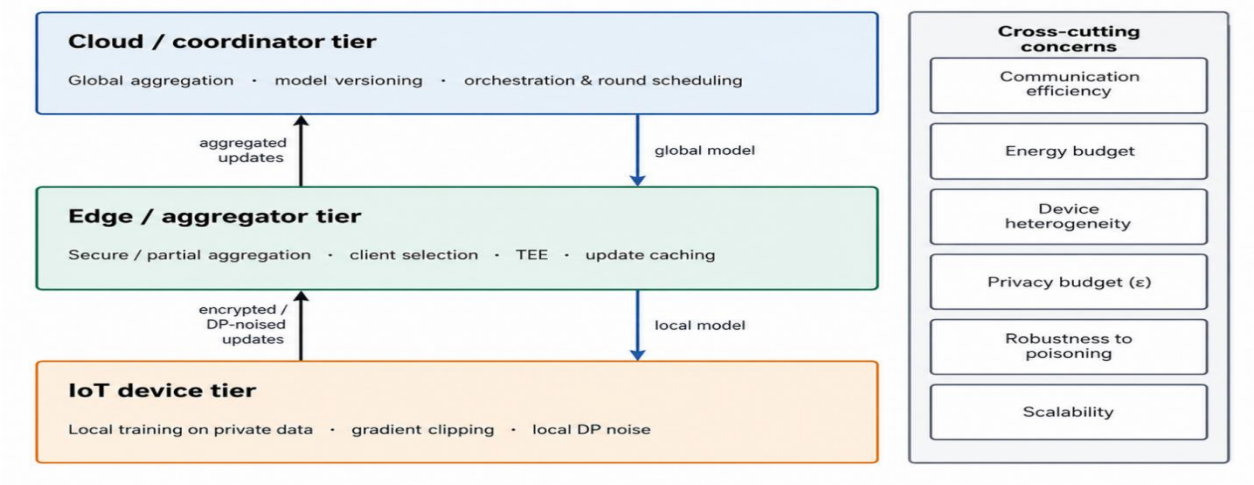

Figure 2. *A layered device–edge–cloud reference architecture for privacy preserving federated learning.*

Table 4. *Evaluation dimensions for privacy preserving federated learning at the edge.*

Evaluation dimension	What it assesses	Example metric / approach
Privacy leakage	Residual information recoverable from updates or model	Attack success rate; privacy budget ε
Model utility	Accuracy retained under privacy protection	Task accuracy vs. non private baseline
Communication cost	Data exchanged per round and to convergence	Bytes / round; rounds to target accuracy
Energy & compute	On device cost of training and protection	Energy per round; wall clock on target hardware
Robustness	Resilience to poisoning and malicious clients	Accuracy under attack; backdoor success rate
Scalability	Behaviour as clients and data grow	Throughput vs. client count
Heterogeneity handling	Performance under non IID data and stragglers	Accuracy across client distribution; fairness
Convergence	Stability and speed of training	Rounds to convergence; variance across runs

4. Discussion

4.1 Principal findings. Three conclusions are repeated in the work sampled. First, there is a well known and well understood menu of privacy mechanisms, and each of these comes with a different form of payment: Differential privacy pays for its guarantee in the expense of accuracy, homomorphic encryption and secure multiparty computation pay for their guarantee in the price of computation and communication, and trusted execution pays for its guarantee in the cost of a trust assumption of the hardware. There are no private, accurate and cheap mechanisms that are all at once. Second, because IoT devices have strict energy, compute and bandwidth constraints, the choice of binding mechanism is not the constraint itself, but a collection of them used precisely at the point where the constraints of the edge bite the hardest. Third, the most obvious and intuitive aspect of FL – that raw data stays on device – is required but it is not enough for privacy; the update channel itself is an attack surface that needs to be explicitly closed.

4.2 The privacy–utility–cost trilemma at the edge. The middle theme is a three-way compromise. Increasing the privacy budget ε increases the noise and decreases the accuracy; increasing the cryptographic confidentiality increases computation and communication; increasing the robustness against poisoning may interact badly with the non IID data which is typical in IoT. On the hardware that has plenty of resources these tensions can be held in check, but on the battery powered sensor they may make the difference. Much of the most useful recent work, then, is related to lower cost, efficient communication security mechanisms for aggregation that tolerate client dropout, or DP mechanisms that spend the privacy budget adaptively rather than uniformly to make acceptable points in the trilemma reachable on otherwise constrained devices [15], [14].

4.3 Systems and heterogeneity. Without systems setting, privacy cannot be dealt with. Devices come and go, have data from entirely disparate distributions, and have capabilities orders of magnitude different. These realities intersect with privacy in real ways: client selection has an impact on convergence, as well as on the privacy accounting; the straggler problem makes it more difficult to develop secure aggregation protocols; and non IID data slows training and makes simple robust aggregation defences less effective. Hierarchical and edge assisted topologies are helpful in that they help to localise aggregation/communication, but introduce an additional trust boundary that needs to be secured [5], [6].

4.4 Open challenges and future directions. Four priorities follow. (1) Adaptive privacy budgeting: spread the privacy budget across rounds, clients and model components, spending ε where it is needed, instead of uniformly, to compensate for accuracy at a fixed guarantee. (2) Dropout-tolerant secure aggregation: efficient protocols for communicating and computing with large, unreliable IoT populations. (3) Integrated privacy and robustness: Defences that are both inference resistant and poisoning resistant, as they both occur in real deployments. (4) Standardised and

reproducible evaluation: common benchmarks and hardware in the loop measurement of privacy, utility, communication and energy in combination, for the comparison of results across studies. Work on these would shift the focus from "is it possible to implement privacy preserving edge FL" to "which designs are feasible in the real IoT budget".

4.5 Limitations of this survey. This review is a narrative synthesis, and thus involves editorial choices in the selection and organization of a sizeable and rapidly evolving literature; it also limited the search to English language studies from the modern FL period. The taxonomy and reference architecture are not inventory lists, since they are intended to organize the field rather than to list everything that exists, and any given point in time is soon outdated. The limitations should be clearly stated in the final paper.

5. Conclusion

While federated learning is a logical solution to the bandwidth, latency, energy, and privacy concerns that arise when using machine learning with IoT data, putting it at the edge increases each constraint and reveals a channel that carries information about the model—unless it is purposefully protected. This survey has categorized the resulting design space into a taxonomy of topologies and privacy strategies, a comparative analysis of the different types of differential privacy, secure aggregation, homomorphic encryption, secure multiparty computation, and trusted execution, an overview of the inference and poisoning attack they have to resist, and a layered device-Edge-cloud reference architecture that can be evaluated along various dimensions. The overall conclusion is a privacy-utility-cost triple dilemma, which can be solved by no single mechanism: practical systems have to be composed of mechanisms in the constrained budgets of IoT hardware. The most promising paths to future demonstrations becoming viable and trusted edge intelligence are adaptive privacy budgeting, efficient and drop-tolerant secure aggregation, integrated privacy and robustness defence, and standardised evaluation.

Declarations

Conflict of Interest

The authors report that there are no conflicts of interest to declare.

Funding Statement

This work was not funded with any specific lines of public, commercial or not for profit funding.

Ethical Approval

No human or animal subjects were involved in this study and hence no ethical approval was required.

References

[1] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: Vision and challenges,” *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016, doi: 10.1109/JIOT.2016.2579198.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication efficient learning of deep networks from decentralized data,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, PMLR, vol. 54, pp. 1273–1282, 2017.

[3] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, Art. no. 12, 2019, doi: 10.1145/3298981.

[4] P. Kairouz et al., “Advances and open problems in federated learning,” *Foundations and Trends in Machine Learning*, vol. 14, nos. 1–2, pp. 1–210, 2021, doi: 10.1561/2200000083.

[5] 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.

[6] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, “Federated learning for Internet of Things: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021, doi: 10.1109/COMST.2021.3075439.

[7] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020, doi: 10.1109/MSP.2020.2975749.

[8] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Advances in Neural Information Processing Systems 32*, Curran Associates, pp. 14774–14784, 2019.

[9] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *2019 IEEE Symposium on Security and Privacy*, pp. 691–706, 2019, doi: 10.1109/SP.2019.00029.

[10] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white box inference attacks against centralized and federated learning," in *2019 IEEE Symposium on Security and Privacy*, pp. 739–753, 2019, doi: 10.1109/SP.2019.00065.

[11] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, nos. 3–4, pp. 211–407, 2014, doi: 10.1561/0400000042.

[12] M. Abadi et al., "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318, 2016, doi: 10.1145/2976749.2978318.

[13] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," arXiv:1712.07557, 2017.

[14] K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020, doi: 10.1109/TIFS.2020.2988575.

[15] K. Bonawitz et al., "Practical secure aggregation for privacy preserving machine learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, 2017, doi: 10.1145/3133956.3133982.

[16] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Computing Surveys*, vol. 51, no. 4, Art. no. 79, 2018, doi: 10.1145/3214303.

[17] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," in *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, PMLR, vol. 108, pp. 2938–2948, 2020.

[18] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Generation Computer Systems*, vol. 115, pp. 619–640, 2021, doi: 10.1016/j.future.2020.10.007.

[19] L. Lyu, H. Yu, and Q. Yang, "Threats to federated learning: A survey," arXiv:2003.02133, 2020.