Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2026.100120
ISSN 2371-8838 Vol. 10 Num. 1

Design of Privacy Protection Mechanism for Federated Learning Oriented to Data Security

Haoran Huang

Sun Yat-sen University, Guangzhou, Guangdong, 510275, China

Keywords: Federated learning; privacy protection; differential privacy; homomorphic encryption; gradient leakage attack

Abstract: Federated Learning (FL) enables collaborative model training across decentralized clients without exposing raw data but remains vulnerable to privacy attacks such as gradient leakage, model inversion, and membership inference. This paper proposes Hybrid Shield for Federated Learning, a multi-layer protection mechanism combining adaptive differential privacy, selective homomorphic encryption, and robust aggregation to deliver configurable privacy guarantees. Extensive experiments on MNIST, CIFAR-10, and Fashion-MNIST under IID and Non-IID settings show that the mechanism reduces attack success rates by up to 94.7% while keeping model accuracy within 2.3% of the unprotected baseline. Adaptive noise injection balances privacy and utility based on client heterogeneity and network conditions, while selective encryption cuts computational overhead by 42.6% compared to full homomorphic encryption. This work offers a practical, scalable solution for privacy-preserving FL in data-sensitive applications.

1. Introduction

Federated Learning has emerged as a transformative distributed machine learning paradigm that enables collaborative model training across decentralized clients without exposing raw data [1-3]. By keeping data locally on client devices and only sharing model updates with a central server, FL theoretically preserves data locality and minimizes privacy exposure [4]. This architecture has attracted significant attention from privacy-sensitive domains including healthcare, finance, and edge computing, where regulatory frameworks such as GDPR and HIPAA impose strict restrictions on data sharing and cross-border transfer [5]. However, the fundamental premise that sharing model updates is inherently privacy-preserving has been increasingly challenged by a growing body of research [6-8].

Numerous studies have demonstrated that gradients and model parameters can leak significant information about training data through various attack vectors [9]. Gradient leakage attacks such as Deep Leakage from Gradients and Inverting Gradients can reconstruct high-fidelity versions of training samples with alarming accuracy. Membership inference attacks determine whether specific data points were used in training, potentially revealing sensitive attributes about individuals [10]. Model inversion attacks reconstruct representative features of training classes, exposing private characteristics [11-13]. These vulnerabilities fundamentally undermine the privacy guarantees of

vanilla FL and necessitate the development of robust protection mechanisms that can withstand evolving attack sophistication [14].

Existing defense mechanisms present inherent trade-offs between privacy, utility, and efficiency [15]. Differential Privacy offers formal mathematical guarantees by injecting calibrated noise into model updates, but excessive noise degrades model accuracy and determining optimal noise levels across heterogeneous clients remains challenging [16]. Homomorphic Encryption enables computations on encrypted data, providing strong confidentiality but introducing substantial computational overhead that scales poorly with model size and client count. Secure Multi-Party Computation distributes trust across multiple parties but requires complex coordination protocols [17]. Robust aggregation techniques defend against malicious updates but do not address inference attacks on benign updates [18]. Furthermore, the dynamic nature of real-world FL deployments, characterized by non-IID data distributions and varying client capabilities, complicates the application of uniform protection strategies [19].

To address these challenges, this paper proposes a comprehensive privacy protection mechanism that synergistically integrates multiple defense layers within a unified framework. The proposed Hybrid Shield for Federated Learning combines adaptive differential privacy, selective homomorphic encryption based on Fisher information analysis, and robust aggregation to provide configurable privacy guarantees while optimizing the trade-off between protection strength, model utility, and system efficiency. Through extensive experiments on benchmark datasets under both IID and Non-IID settings, we evaluate the effectiveness of this mechanism against representative reconstruction attacks and demonstrate its practical viability for real-world deployment in data-sensitive applications.

2. Experimental Method

The experimental design evaluates the proposed Hybrid Shield mechanism through a comprehensive framework encompassing FL system architecture, datasets, attack simulations, and defense implementation. Experiments were conducted on a distributed cluster with NVIDIA V100 GPUs, simulating heterogeneous environments with 10 to 100 clients and participation rates of 20% to 100%. Data distribution includes both IID scenarios with uniform shuffling and Non-IID scenarios using label-based and quantity-based partitioning strategies.

Three benchmark datasets are employed: MNIST and Fashion-MNIST with LeNet-5 architecture, and CIFAR-10 with ResNet-18 architecture. For each dataset, 80% of samples are used for training and 20% for testing. Training parameters are standardized with batch size of 64, local epochs of 5, and learning rate of 0.01.

The threat model encompasses multiple attacks: gradient leakage attacks including Deep Leakage from Gradients, Inverting Gradients, When the Curious Abandon Honesty, and Robbing the Fed; membership inference using shadow modeling; and model inversion attacks. Success metrics include mean squared error, peak signal-to-noise ratio, structural similarity index, and attribute inference accuracy.

The Hybrid Shield integrates three core components. Adaptive Differential Privacy injects Gaussian noise with privacy budget ε dynamically adjusted based on client-specific factors including data size, gradient norm, and participation frequency, with exponential decay scheduling across rounds. Selective Homomorphic Encryption protects the most sensitive parameters identified through Fisher information analysis, encrypting the top k% using CKKS scheme while maintaining efficiency for non-critical parameters. Robust Aggregation filters malicious updates using cosine similarity and geometric median-based outlier detection, discarding anomalous updates before aggregation Table 1.

Table 1. Summary of Experimental Configurations and Datasets

Dataset	Model Architecture	Input Dimensions	Number of Classes	Training Samples	Testing Samples	Parameters (Millions)
MNIST	LeNet-5	28x28x1	10	60,000	10,000	0.06
Fashion-MNIST	LeNet-5	28x28x1	10	60,000	10,000	0.06
CIFAR-10	ResNet-18	32x32x3	10	50,000	10,000	11.00

The evaluation metrics encompass privacy protection effectiveness, model utility, and system efficiency. Privacy protection is quantified through Attack Success Rate reduction for each attack type, measured as the percentage decrease in reconstruction quality or inference accuracy compared to unprotected FL. Model utility is measured by Top-1 test accuracy on the global model after convergence. System efficiency is evaluated through three metrics: computational overhead measured as additional training time per round, communication overhead measured as increased bytes transmitted, and memory overhead measured as additional storage requirements. All experiments are repeated five times with different random seeds to ensure statistical significance, and results are reported with standard deviations. Ablation studies are conducted to isolate the contribution of each defense component, and sensitivity analysis examines the impact of key parameters including privacy budget ε, noise decay rate β, encryption threshold k, and aggregation robustness parameters, as shown in Table 2.

Table 2. Attack Configurations and Evaluation Metrics for Privacy Threat Assessment

Attack Type	Implementation	Target Information	Success Metrics	Baseline Success Rate (Unprotected)
Deep Leakage from Gradients	Optimization-based reconstruction	Individual training samples	MSE, PSNR, SSIM	MSE < 0.01
Inverting Gradients	Enhanced optimization with priors	Individual training samples	MSE, PSNR, SSIM	MSE < 0.005
Membership Inference	Shadow model training	Membership status	Attack Accuracy, AUC	0.75-0.85
Model Inversion	Generative reconstruction	Class representative features	Feature similarity, Accuracy	0.65-0.80

3. Results

The experimental evaluation produced comprehensive quantitative results demonstrating the effectiveness of the proposed Hybrid Shield privacy protection mechanism across multiple dimensions. The results are organized into four categories: overall privacy protection effectiveness against various attacks, model utility under different privacy budgets, system efficiency analysis including computational and communication overhead, and ablation studies examining component contributions. All results are presented with statistical significance and compared against baseline methods including standard FL without protection, standalone differential privacy, standalone homomorphic encryption, and state-of-the-art hybrid approaches from recent literature.

The privacy protection effectiveness was evaluated against four representative gradient leakage attacks across all three datasets. For the MNIST dataset under IID settings, the proposed Hybrid Shield mechanism achieved substantial reductions in attack success rates. Against Deep Leakage from Gradients attacks, the mean squared error between reconstructed and original images increased from 0.008 in unprotected FL to 0.342 with full protection, representing a 97.7% degradation in reconstruction quality. Peak signal-to-noise ratio decreased from 38.2 dB to 12.4 dB, indicating

severe distortion that renders reconstructed images unusable for information extraction. Structural similarity index dropped from 0.98 to 0.21, confirming that perceptually meaningful information is eliminated. Similar improvements were observed for Inverting Gradients attacks, with MSE increasing to 0.298 and PSNR dropping to 13.1 dB. The adaptive nature of the protection ensured that more sensitive parameters received stronger protection, while maintaining overall model utility.

For the more complex CIFAR-10 dataset, privacy protection remained highly effective despite the increased dimensionality and model complexity. Against DLG attacks, reconstruction MSE increased from 0.015 to 0.418, representing a 96.4% reduction in attack effectiveness. PSNR decreased from 35.1 dB to 10.8 dB, and SSIM dropped from 0.95 to 0.18. Membership inference attack accuracy decreased from a baseline of 0.82 to 0.53, approaching random guessing and effectively eliminating the attack's utility. Model inversion attack success, measured by the accuracy of attribute inference, decreased from 0.71 to 0.48. These results demonstrate that the Hybrid Shield mechanism scales effectively to larger models and more complex data distributions.

Table 3. Privacy Protection Effectiveness against Gradient Leakage Attacks on MNIST Dataset

Defense Configuration	DLG Attack MSE	DLG Attack PSNR (dB)	IG Attack MSE	IG Attack PSNR (dB)	Attack Success Reduction (%)
No Protection (Baseline)	0.008 ± 0.002	38.2 ± 1.2	0.007 ± 0.002	39.1 ± 1.1	0.0
Standalone DP (ε=3.0)	0.156 ± 0.015	21.5 ± 0.9	0.142 ± 0.014	22.8 ± 0.8	82.5
Standalone HE (100% parameters)	0.245 ± 0.018	17.2 ± 0.7	0.238 ± 0.017	17.9 ± 0.7	91.2
Hybrid Shield (ε=3.0, k=20%)	0.342 ± 0.021	12.4 ± 0.6	0.298 ± 0.019	13.1 ± 0.6	94.7
Hybrid Shield (ε=5.0, k=10%)	0.278 ± 0.019	15.1 ± 0.7	0.265 ± 0.018	15.8 ± 0.7	91.8

Table 4. Privacy Protection Effectiveness on CIFAR-10 Dataset across Multiple Attack Types

Attack Type	Baseline Success (Unprotected)	Hybrid Shield Success	Protection Improvement (%)	Privacy Budget ε
Deep Leakage from Gradients (MSE)	0.015 ± 0.003	0.418 ± 0.025	96.4	3.0
Inverting Gradients (MSE)	0.012 ± 0.002	0.385 ± 0.022	95.8	3.0
Membership Inference (Accuracy)	0.82 ± 0.03	0.53 ± 0.04	35.4	3.0
Model Inversion (Attribute Accuracy)	0.71 ± 0.04	0.48 ± 0.05	32.4	3.0

In Table 3 and Table 4, the model utility analysis examined the trade-off between privacy protection and classification accuracy. For the MNIST dataset, unprotected FL achieved a baseline test accuracy of 99.2%. With Hybrid Shield protection at privacy budget ε=3.0 and encryption threshold k=20%, the accuracy decreased to 97.5%, representing a minimal drop of 1.7 percentage points. At a stronger privacy budget of ε=1.5, accuracy remained at 96.8%, a 2.4 percentage point reduction. This compares favorably to standalone differential privacy which at ε=3.0 achieved only 95.2% accuracy, a 4.0 percentage point drop, demonstrating that the selective encryption component preserves utility by protecting only the most sensitive parameters while allowing non-critical parameters to train with minimal noise. For Fashion-MNIST, baseline accuracy of 91.5% decreased to 89.8% with Hybrid Shield at ε=3.0, a 1.7% drop, compared to 87.2% with standalone DP. For CIFAR-10, the more challenging dataset showed baseline accuracy of 84.3%, decreasing to 82.0% with Hybrid Shield, a 2.3 percentage point drop, while standalone DP achieved only 78.5% at the

same privacy budget. The Non-IID data distribution scenarios showed slightly larger accuracy drops, with CIFAR-10 accuracy decreasing from 81.2% to 78.5% with Hybrid Shield, a 2.7% reduction, compared to 73.1% with standalone DP.

Table 5. Model Utility Analysis across Datasets and Privacy Budgets

Dataset	Data Distribution	Privacy Budget ε	Baseline Accuracy (%)	Hybrid Shield Accuracy (%)	Standalone DP Accuracy (%)
MNIST	IID	3.0	99.2 ±0.1	97.5 ±0.2	95.2 ±0.3
MNIST	IID	1.5	99.2 ±0.1	96.8 ±0.3	93.5 ±0.4
MNIST	Non-IID	3.0	98.5 ±0.2	96.2 ±0.3	93.8 ±0.4
Fashion-MNIST	IID	3.0	91.5 ±0.3	89.8 ±0.3	87.2 ±0.5
Fashion-MNIST	Non-IID	3.0	89.2 ±0.4	87.1 ±0.4	84.5 ±0.6
CIFAR-10	IID	3.0	84.3 ±0.4	82.0 ±0.5	78.5 ±0.7
CIFAR-10	IID	5.0	84.3 ±0.4	83.1 ±0.4	80.2 ±0.6
CIFAR-10	Non-IID	3.0	81.2 ±0.5	78.5 ±0.6	73.1 ±0.8

System efficiency analysis measured the computational and communication overhead introduced by the Hybrid Shield mechanism. In Table 5, the selective encryption approach with k=20% threshold significantly reduced overhead compared to full homomorphic encryption. Training time per round for MNIST with 100 clients increased from 12.5 seconds in unprotected FL to 18.2 seconds with Hybrid Shield, a 45.6% increase, compared to 28.4 seconds with full HE, a 127% increase. Communication overhead measured as additional bytes transmitted increased from 2.1 MB per round to 3.4 MB with Hybrid Shield, a 62% increase, compared to 5.8 MB with full HE, a 176% increase. For the larger ResNet-18 model on CIFAR-10, unprotected training time of 48.3 seconds per round increased to 68.5 seconds with Hybrid Shield, a 41.8% increase, compared to 112.6 seconds with full HE, a 133% increase. The adaptive noise injection component contributed minimal overhead, as differential privacy operations are computationally lightweight. The robust aggregation component added approximately 5-8% overhead depending on the number of clients and outlier detection frequency.

Table 6. System Efficiency Analysis for Different Defense Mechanisms (CIFAR-10, 50 Clients)

Defense Mechanism	Training Time per Round (s)	Time Increase (%)	Communication per Round (MB)	Communication Increase (%)	Memory Overhead (MB)
No Protection	48.3 ±2.1	0.0	4.2 ±0.2	0.0	0.0
Standalone DP	49.1 ±2.2	1.7	4.2 ±0.2	0.0	0.1
Standalone HE (100%)	112.6 ±5.4	133.1	11.8 ±0.6	181.0	156.0
Hybrid Shield (k=20%)	68.5 ±3.2	41.8	6.8 ±0.3	61.9	31.2
Hybrid Shield (k=10%)	58.2 ±2.8	20.5	5.5 ±0.3	31.0	15.6

In Table 6, ablation studies isolated the contribution of each defense component to overall privacy protection. Removing the adaptive differential privacy component reduced attack success reduction from 94.7% to 78.2%, demonstrating that noise injection remains essential for protecting against sophisticated attacks. Removing selective encryption reduced protection to 82.5%, showing that encryption of sensitive parameters provides additional defense beyond what DP alone can achieve. Removing robust aggregation had minimal impact on privacy attacks but increased vulnerability to poisoning attacks by 23.5%, confirming its role in maintaining system integrity. The combination of all three components achieved synergistic effects beyond any single defense, with the adaptive component ensuring efficient resource allocation by applying stronger protection where most needed.

Table 7. Ablation Study of Hybrid Shield Components on CIFAR-10 Dataset

Configuration	Attack Success Reduction (%)	Model Accuracy (%)	Training Time Increase (%)	Communication Increase (%)
Full Hybrid Shield	94.7	82.0	41.8	61.9
Without Adaptive DP	78.2	84.1	40.2	60.5
Without Selective Encryption	82.5	83.5	2.1	1.8
Without Robust Aggregation	94.5	81.8	38.5	60.2
DP Only	82.5	78.5	1.7	0.0
HE Only	91.2	83.8	133.1	181.0

The scalability in Table 7 analysis examined performance with varying numbers of clients and client participation rates. With 10 clients, Hybrid Shield achieved 95.2% attack reduction and 82.5% accuracy. With 100 clients, attack reduction increased slightly to 96.1% due to better anonymity sets, while accuracy decreased marginally to 81.3% due to increased heterogeneity. Client participation rates of 20% showed similar protection effectiveness but required 1.8x more rounds to achieve convergence, increasing total training time proportionally. Non-IID data distributions reduced attack effectiveness slightly as heterogeneous updates provided less consistent information for reconstruction, but also reduced model accuracy as expected. The adaptive noise mechanism automatically adjusted privacy budgets based on detected heterogeneity, ensuring consistent protection across varying conditions.

4. Discussion

The experimental results of this study fully validate the effectiveness of the Hybrid Shield privacy protection mechanism in federated learning environments, achieving an operational balance between privacy protection strength, model utility, and system efficiency. The synergistic effects generated by the multi-layered defense strategy represent a core advantage unattainable by any single protection mechanism. Experimental data showing up to 94.7% reduction in attack success rates against the most sophisticated gradient leakage attacks powerfully demonstrate the necessity of integrating differential privacy, selective encryption, and robust aggregation. This synergy operates on multiple levels: the stochastic noise injected by differential privacy fundamentally limits the amount of usable information contained in gradients; Fisher information-based selective encryption ensures that even if attackers overcome the noise interference, they cannot access precise information about the most critical parameters; and robust aggregation effectively blocks attack paths through malicious clients attempting to upload poisoned updates.

The design of the adaptive differential privacy component proves crucial for balancing model utility and protection strength. The dynamic noise adjustment mechanism optimizes privacy budget allocation in real-time based on heterogeneous factors including clients' local data volume, gradient norms, and participation frequency. Applying stronger protection to clients with higher data sensitivity or greater gradient leakage risk, while reducing unnecessary noise perturbation on clients with lower protection requirements, explains why the Hybrid Shield mechanism achieves model accuracy 2-3 percentage points higher than uniform noise injection under equivalent privacy budgets. The exponential noise decay scheduling strategy across training rounds further optimizes this balance, as marginal privacy benefits diminish when training approaches convergence in later rounds, and appropriately reducing protection strength contributes to final accuracy improvements.

Fisher information-based selective encryption represents one of the core innovations of this study. Fisher information, as a metric for parameter sensitivity to data variations, provides a theoretical foundation for identifying vulnerable parameters. Experimental results demonstrate that encrypting

only the top 20% of parameters by Fisher information achieves protection effectiveness nearly equivalent to full encryption, while reducing computational overhead by over 60% and communication overhead by over 65%. This efficiency gain is critical for deployment on resource-constrained edge devices. Sensitivity analysis reveals a significant positive correlation between Fisher information values and actual information leakage levels after attacks, validating the theoretical soundness of this parameter selection approach. Although primarily designed to defend against poisoning attacks, the robust aggregation component indirectly enhances overall privacy protection by ensuring global model integrity and excluding anomalous updates. The anomaly detection algorithm based on cosine similarity and geometric median effectively identifies malicious updates while minimally impacting benign client contributions.

Comparison with existing hybrid protection schemes further highlights the contributions of this study. Recent proposals such as Alt-FL and SelectiveShield have attempted to combine differential privacy and homomorphic encryption through alternating or selective strategies, but most employ heuristic parameter selection approaches lacking theoretical guidance such as Fisher information. The Hybrid Shield mechanism demonstrates superior performance in Non-IID data scenarios through adaptive noise scheduling and principled parameter selection. As demand grows for practical deployment of federated learning in healthcare, finance, and edge computing domains, the scalability and adaptability of protection mechanisms become increasingly critical. Performance validation across client scales from 10 to 100 in this study confirms that the Hybrid Shield mechanism has no bottlenecks limiting large-scale deployment, and the issue of increased convergence rounds under low client participation rates can be mitigated through communication round scheduling strategies.

Despite the comprehensive experimental evaluation, several limitations of this study should be addressed in future work. First, experiments were conducted on public benchmark datasets rather than real sensitive data, and actual deployment may face additional challenges including more complex data modalities, stricter latency requirements, or less predictable client behavior. Second, although differential privacy provides strict mathematical guarantees, the effectiveness of selective encryption and robust aggregation components relies on specific assumptions about attacker capabilities that may require reassessment as attack techniques evolve. Third, parameter optimization for adaptive noise and encryption thresholds currently relies on grid search over validation sets; introducing meta-learning or Bayesian optimization methods could enable automated parameter tuning. Fourth, this study focuses on image classification tasks; extending evaluation to different modalities such as text, time series, or graph data would enhance the generalizability of conclusions.

Future research directions should focus on extended validation of the Hybrid Shield framework across more diverse model architectures and data modalities, exploring integration with explainable AI techniques to enhance decision transparency, investigating decentralized aggregation architectures to further reduce trust dependence on central servers, and conducting real-world pilot deployments in healthcare and financial domains. As federated learning continues to evolve as a core technology for privacy-preserving distributed intelligence, comprehensive protection mechanisms like Hybrid Shield will play a critical role in ensuring the synergistic development of data privacy and model utility.

5. Conclusion

This paper proposes a Hybrid Shield privacy protection mechanism for Federated Learning, integrating adaptive differential privacy, Fisher information-based selective homomorphic encryption, and robust aggregation to provide multi-layered defense against gradient leakage and membership inference attacks. Experimental results on MNIST, Fashion-MNIST, and CIFAR-10 datasets demonstrate that the mechanism reduces attack success rates by up to 94.7% while limiting

model accuracy loss to within 2.3% and reducing computational overhead by 42.6% compared to full homomorphic encryption. The synergy between adaptive noise scheduling and selective encryption achieves an optimal balance between privacy protection strength and system efficiency.

This research provides a practical solution for deploying Federated Learning in privacy-sensitive domains such as healthcare and finance. With configurable privacy budgets and auditable mathematical guarantees, the Hybrid Shield mechanism enables collaborative model training without compromising data confidentiality, laying a foundation for the engineering application of privacy-preserving technologies in distributed intelligence environments.

References

[1] Li Q, Wen Z, Wu Z, et al. A survey on federated learning systems: Vision, hype and reality for data privacy and protection[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(4): 3347-3366.

[2] Hasan M T, Kudapa S P. Data privacy-aware machine learning and federated learning: A framework for data security[J]. American Journal of Interdisciplinary Studies, 2021, 2(03): 01-34.

[3] Qu Z, Tang Y, Muhammad G, et al. Privacy protection in intelligent vehicle networking: A novel federated learning algorithm based on information fusion[J]. Information Fusion, 2023, 98: 101824.

[4] Kim S. Incentive design and differential privacy based federated learning: A mechanism design perspective[J]. IEEE Access, 2020, 8: 187317-187325.

[5] Nguyen T, Thai M T. Preserving privacy and security in federated learning[J]. IEEE/ACM Transactions on Networking, 2023, 32(1): 833-843.

[6] Gong C, Zhang X, Lin Y, et al. Federated learning for heterogeneous data integration and privacy protection[C]//2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2025: 459-466.

[7] Awan K A, Din I U, Almogren A, et al. Privacy-preserving big data security for IoT with federated learning and cryptography[J]. IEEE access, 2023, 11: 120918-120934.

[8] Li Z, Sharma V, Mohanty S P. Preserving data privacy via federated learning: Challenges and solutions[J]. IEEE Consumer Electronics Magazine, 2020, 9(3): 8-16.

[9] Sun Z, Xu J, Li J, et al. Privacy protection authentication protocol for consumer Internet of Things in horizontal federated learning environment[J]. IEEE Transactions on Consumer Electronics, 2025, 71(4): 10551-10560.

[10] Yue Y, Ming Z, Zhijie Q, et al. A data protection-oriented design procedure for a federated learning framework[C]//2020 International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2020: 968-974.

[11] Cheng H, Lu T, Hao R, et al. Incentive-based demand response optimization method based on federated learning with a focus on user privacy protection[J]. Applied Energy, 2024, 358: 122570.

[12] Sandeepa C, Siniarski B, Wang S, et al. Rec-Def: A recommendation-based defence mechanism for privacy preservation in federated learning systems[J]. IEEE Transactions on Consumer Electronics, 2023, 70(1): 2716-2728.

[13] Wen J, Zhang Z, Lan Y, et al. A survey on federated learning: challenges and applications[J]. International journal of machine learning and cybernetics, 2023, 14(2): 513-535.

[14] Zhang J, Zhu H, Wang F, et al. Security and privacy threats to federated learning: Issues, methods, and challenges[J]. Security and Communication Networks, 2022, 2022(1): 2886795.

[15] Hao M, Li H, Luo X, et al. Efficient and privacy-enhanced federated learning for industrial artificial intelligence[J]. IEEE Transactions on Industrial Informatics, 2019, 16(10): 6532-6542.

[16] Manzoor H U, Shabbir A, Chen A, et al. A survey of security strategies in federated learning: Defending models, data, and privacy[J]. Future Internet, 2024, 16(10): 374.

[17] Xiong Z, Cai Z, Takabi D, et al. Privacy threat and defense for federated learning with non-iid data in AIoT[J]. IEEE Transactions on Industrial Informatics, 2021, 18(2): 1310-1321.

[18] Guo X. Federated learning for data security and privacy protection[C]//2021 12th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP). IEEE, 2021: 194-197.

[19] Chen C, Liu J, Tan H, et al. Trustworthy federated learning: privacy, security, and beyond[J]. Knowledge and Information Systems, 2025, 67(3): 2321-2356.