Federated Learning Privacy Protection Collaboration Model Based on Improved Cyber Threat Collaborative Analysis Capability

Rui Zhuang[1] and Xiling Yang[2,*]

[1]*Department of Fundamental Network Technology, China Mobile Research Institute, Beijing 100053, China*
[2]*College of Business Administrations, Henan Finance University, Zhengzhou 451464, China*
E-mail: yangxl18@outlook.com
[*]*Corresponding Author*

Received 08 January 2026; Accepted 18 May 2026

Abstract

Traditional federated learning has challenges such as slow model convergence, lagging situational awareness, and risk of gradient privacy leakage in cyber threat collaborative analysis capability. This study proposes a federated learning privacy-preserving collaboration model based on threat intelligence drive and hierarchical aggregation. The model optimizes participating nodes through a dynamic client selection mechanism, uses a hierarchical aggregation strategy to balance the learning of basic features and advanced threat patterns, and introduces an adaptive differential privacy mechanism to strengthen gradient protection. The experiment is based on the CIC-IDS-2018 public dataset, which covers various types of attacks in real network environments, with a data volume of approximately 5 million pieces. It is divided into 50 clients in a non-independent and identically distributed manner to simulate cross organizational collaboration scenarios. In the threat

Journal of Cyber Security and Mobility, Vol. 15_4, 915–938.
doi: 10.13052/jcsm2245-1439.1545

detection task, the model designed by the research institute achieved an accuracy of 91.8% and an F1 Score of 89.8% compared to baseline models such as FedAvg, FedProx, and DP FedAvg. All indicators were superior to the comparison model. In addition, in the advanced persistent threat attack scenario, the attack chain detection rate increased to 95.6%, and the average detection time was shortened to 2.8 hours. In terms of privacy protection, the Rényi privacy loss was only 2.89 with a budget of $\varepsilon = 3.0$. The proposed model effectively improves the efficiency and timeliness of collaborative detection of cross-organizational threats while ensuring data privacy and provides a feasible solution for building a safe and reliable collaborative defense system. It should be pointed out that while improving detection performance, the model introduces additional communication overhead caused by hierarchical aggregation and dynamic selection mechanisms. The average communication traffic in the experiment was about 13.8 GB. As the number of clients expands to a larger scale, the computational load and scheduling complexity of the coordination layer will further increase. In addition, although the non-independent and identically distributed data partitioning used in the experiment can simulate real heterogeneous scenarios, the convergence efficiency of the model under extreme distributions still needs further verification. The above limitations will be optimized in future work by introducing asynchronous aggregation and lightweight communication protocols.

Keywords: Federated learning, cyber threats, privacy leakage, privacy protection, advanced persistent threats.

1 Introduction

As cyber-attack methods become increasingly advanced, organized, and concealed, complex cyber threats such as Advanced Persistent Threat (APT) pose severe challenges to the data defense capabilities of a single organization [1–3]. Traditional threat detection methods based on isolated data are limited by local vision and limited samples, making it difficult to comprehensively perceive and accurately judge cross-domain coordinated attacks. In existing research, federated learning, as a distributed learning paradigm, provides new ideas for cross-organizational security collaborative. Shen et al. proposed a multiplicative double privacy mask algorithm combined with a gradient screening mechanism to solve the privacy leakage problem caused by uploading plaintext gradients in federated learning. The results showed that this

solution reduced the computational overhead by 51.78% and the communication traffic by 64.87% while maintaining almost no loss in model accuracy, allowing the privacy-preserving federated learning framework to be deployed on lighter and smaller cross-institution or cross-device terminals [4]. Yang et al. proposed a federated multi-task machine learning framework, which combined fairness modeling and privacy protection in multi-task learning scenarios. Simulation experiments showed that this framework had strong privacy protection capabilities while ensuring model accuracy and fairness between tasks. It could provide a feasible solution for the privacy and security of educational data and accelerate the digital transformation of education [5]. Xu et al. embedded blockchain smart contracts and homomorphic encryption into the existing IoT architecture to achieve privacy-preserving aggregation without sharing plaintext or trusted centers in IoT federated learning scenarios. The results showed that this scheme could support cross-terminal joint modeling within an acceptable range of communication and computing overhead and provided a feasible path for privacy federated learning on large-scale lightweight IoT devices [6]. Yang et al. proposed a federated deep neural network ownership verification scheme (Federated Intellectual Property Right, FedIPR). By embedding verifiable digital watermarks in federated models, it solved the intellectual property ownership that was prone to plagiarism, misappropriation, and abuse during the distribution and sharing process. The results showed that FedIPR watermarking was robust against collusion, fine-tuning, and pruning while maintaining model accuracy, and could provide an integrated framework for building trusted AI [7].

However, the above methods still have significant limitations in the context of network threat intelligence collaboration. From the perspective of convergence speed, although the multiplication dual privacy mask proposed by Shen et al. [4] reduces computational and communication overhead, its fixed client participation mode cannot prioritize the activation of clients with the latest threat samples when facing dynamic evolving threat situations, resulting in a lag in the model's response to new attack modes. From the perspective of non-independent and identically distributed data processing capabilities, the federated multitasking framework proposed by Yang et al. [5] and the blockchain scheme proposed by Xu et al. [6] both adopt a homogeneous aggregation strategy, where all client updates are equally weighted. In the reality of highly heterogeneous distribution of network security data, this can easily bias the model towards clients with large data volumes, thereby weakening its learning ability for small sample but high-value threat patterns (such as rare attack types). From the perspective of privacy utility trade-off,

the above methods often adopt fixed privacy budget allocation strategies and do not consider the differences in noise sensitivity at different stages of training, resulting in a significant decrease in model utility under strong privacy protection requirements (such as Yang et al.'s FedIPR scheme [7], which protects model intellectual property, but the balance between privacy protection and detection performance still needs to be optimized). In contrast, threat intelligence driven federated learning needs to simultaneously meet the requirements of fast convergence, adaptation to data heterogeneity, and dynamic adjustment of privacy protection strength. However, existing methods have shortcomings in these dimensions, and there is an urgent need to design privacy protection collaboration models specifically for network threat collaborative analysis scenarios. Therefore, the study proposes a federated learning privacy protection collaboration model driven by threat intelligence, aiming to build a collaborative threat analysis framework that takes into account detection efficiency and privacy protection, thereby achieving safe sharing of cross-organizational security capabilities.

The research scope and system assumptions of this article are defined as follows. The proposed model is aimed at collaborative threat assessment scenarios across network security organizations. It is assumed that all participants are operating under a semi honest threat model, meaning that both the client and coordination server will strictly follow the protocol process, but may attempt to infer sensitive information of other participants from the received model updates or aggregation results. The attacker's ability is limited to being unable to break through the provable privacy protection provided by the underlying differential privacy mechanism, nor can they directly access the raw local data of any client. In terms of trust boundaries, participating organizations do not trust each other, but they all trust the global training protocol and privacy budget parameters published by the coordination layer. The model is not suitable for Byzantine attack scenarios where malicious clients actively poison or damage the model, and additional robust aggregation mechanisms need to be introduced to cope with such scenarios.

In addition, from the perspective of practical deployment, the application of the proposed model in cross organizational network defense environments needs to pay attention to the following practical constraints. In terms of communication overhead, although hierarchical aggregation and dynamic client selection mechanisms improve detection performance, they introduce additional model update transmission burden, which needs to be optimized in practical deployment by combining gradient compression or sparsity techniques. In terms of scalability, when the scale of participating clients expands

to hundreds or more, the scheduling complexity of the coordination layer and the computational load of model aggregation will significantly increase, which may become a system bottleneck. In the future, a layered federated learning architecture can be explored to support larger scale collaborative defense. In terms of cross organizational deployment constraints, there may be differences in network latency, bandwidth limitations, and compliance requirements among the participating parties. The proposed model naturally fits these constraints through local training and gradient sharing, but the actual implementation still requires coordination among all parties to establish a unified privacy budget negotiation mechanism and training protocol. The above practical considerations will be further analyzed in subsequent experiments and discussions.

The innovation of the research lies in: (1) constructing a dynamic client selection mechanism, adaptively optimizing participating nodes based on threat intelligence relevance, expected to accelerate the convergence speed of the model to new threat patterns, and reduce training epochs; (2) designing a hierarchical model aggregation strategy to handle both basic feature updates and advanced threat pattern updates. Experiments have shown that this strategy can increase the APT attack chain detection rate to 95.6% and significantly enhance the learning ability for rare attack types; (3) designing a dual protection mechanism based on adaptive differential privacy to achieve dynamic allocation of privacy budget and precise control of gradient disturbance. Under a medium privacy budget of $\varepsilon = 3.0$, only 2.89 Rényi privacy loss is generated while maintaining a detection accuracy of 91.8%, achieving a good balance between privacy protection and model utility.

2 Methods and Materials

2.1 Federated Learning Architecture Design for Cyber Threat Collaborative Analysis Capability

As cyber-attacks become increasingly complex, organized, and concealed, it is often difficult for a single organization to comprehensively and accurately perceive and judge complex cyber threats such as APTs relying only on its own limited local data. Cross-organizational data collaborative is seen as a key path to improving threat detection and analysis capabilities. However, cybersecurity data often contains sensitive operational information, user privacy and even the operating status of critical infrastructure. Directly sharing raw data faces legal and regulatory risks as well as security risks. Federated

learning is an emerging distributed machine learning paradigm. "The data does not move but the model moves" provides a potential technical solution for achieving cross-subject collaborative threat analysis under the premise of privacy protection [8, 9]. However, due to its static client selection and homogenized aggregation strategy, the standard federated learning framework is prone to problems such as slow model convergence and lagging situational awareness when dealing with dynamically evolving cyber threats. The original gradient updates it transmits are also susceptible to privacy attacks such as member inference and attribute inference. To deal with these problems, the study designs a federated learning architecture for collaborative threat assessment. Based on standard federated learning, this architecture introduces a threat intelligence-driven client selection mechanism and a hierarchical model aggregation strategy, aiming to achieve efficient and safe collaborative security capability improvement, as shown in Figure 1.

Figure 1 shows the federated learning architecture for collaborative threat analysis. The architecture consists of three core logic layers, including multiple cybersecurity entities participating in collaborative as the client layer, a central server responsible for coordinating the federated learning process as the coordination layer, and a threat intelligence management layer that runs throughout. At the client layer, each participating organization (including different enterprise security operations centers, cloud service providers, and national computer emergency response teams) maintains their own cybersecurity data locally. This data may include network traffic logs, system audit events, and feature vectors of malware samples [10]. Each client uses its local data set to train a local threat analysis model. The initial weight of the model is uniformly distributed by the coordination layer to ensure a consistent model structure. After training is completed, the client sends the calculated model updates to the coordination layer. As the organizer of federated learning, the coordination layer is mainly responsible for model aggregation, process scheduling, resource allocation and other functions. The specific operation process of the coordination layer is illustrated in Figure 2.

Figure 2 shows the specific process of the coordination layer. This layer first implements a dynamic client selection mechanism driven by threat intelligence. Before each round of federated training begins, the coordination layer will access a real-time updated global threat intelligence source and calculate the semantic correlation between each client's local data and the current threat situation. To accurately quantify the client's relevance to the current threat landscape, the study introduces text embedding technology based on pre-trained language models. Specifically, the Sentence

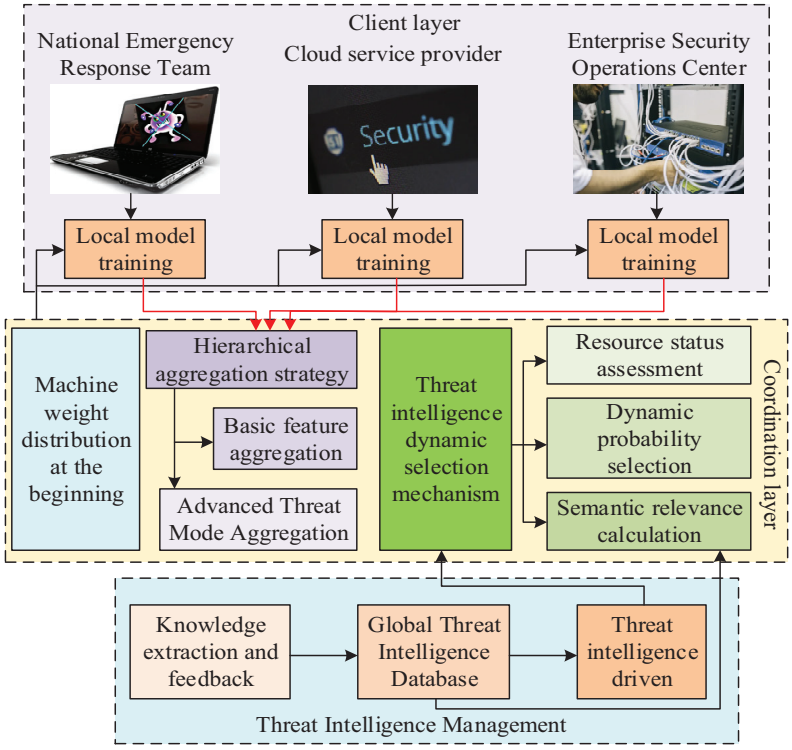

Figure 1 Federated learning architecture for cyber threat collaborative analysis capability (Image source: https://colorhub.me/photos/voelE; https://colorhub.me/photos/JPG7q; https://colorhub.me/photos/3r8zx).

Bidirectional Encoder Representations from Transformers (Sentence-BERT) is used to convert unstructured threat intelligence text and metadata such as industry attributes and system environment submitted by each client into semantic embedding vectors in high-dimensional vector space [11, 12]. Then, the semantic relevance is objectively measured by calculating the cosine similarity between the threat intelligence vector and each client metadata vector. On this basis, the selection probability is calculated based on a comprehensive function that takes into account the aforementioned correlation measures, the client's cumulative historical participation frequency, and the current available computing resource status. By introducing the negative exponential decay term of historical participation frequency, this mechanism can effectively avoid the loss of model contribution caused by some nodes not being selected for a long time while giving priority to high-correlation

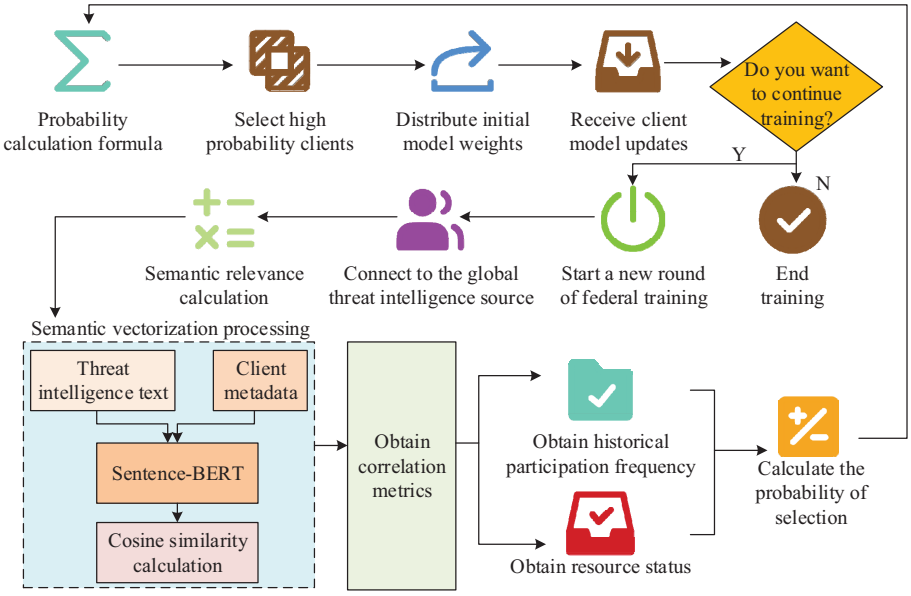

Figure 2 Step process of coordination layer (Icon source: https://iconpark.oceanengine.com/official).

clients, thereby ensuring the fairness of federated training and the stability of model convergence. This dynamic selection process can be expressed as:

$$P_i^{(t)} = \frac{\mathrm{Sim}(D_i, TI^{(t)}) \cdot \exp(-\alpha \cdot C_i^{(t-1)}) \cdot R_i^{(t)}}{\sum_{j=1}^{N} \mathrm{Sim}(D_j, TI^{(t)}) \cdot \exp(-\alpha \cdot C_j^{(t-1)}) \cdot R_j^{(t)}} \tag{1}$$

where $P_i^{(t)}$ represents the probability that the i-th client is selected by the coordination layer in the t-th round of federated learning training, $\mathrm{Sim}(D_i, TI^{(t)})$ represents the semantic similarity function, i stands for client, D_i is the local data corresponding to the client, $\mathbf{T}I^{(t)}$ represents the global threat intelligence in the t-th round, $C_i^{(t-1)}$ represents the total number of times that the i-th client has been successfully selected to participate in federated training as of the $t-1$-th round in history, α is the attenuation coefficient, $R_i^{(t)}$ represents the available resource weight of the i-th client in the t-th round, N is the total number of clients participating in federated learning [13]. After completing client selection and receiving model updates from the selected clients, the coordination layer will continue to execute the hierarchical model aggregation strategy, as shown in Figure 3.

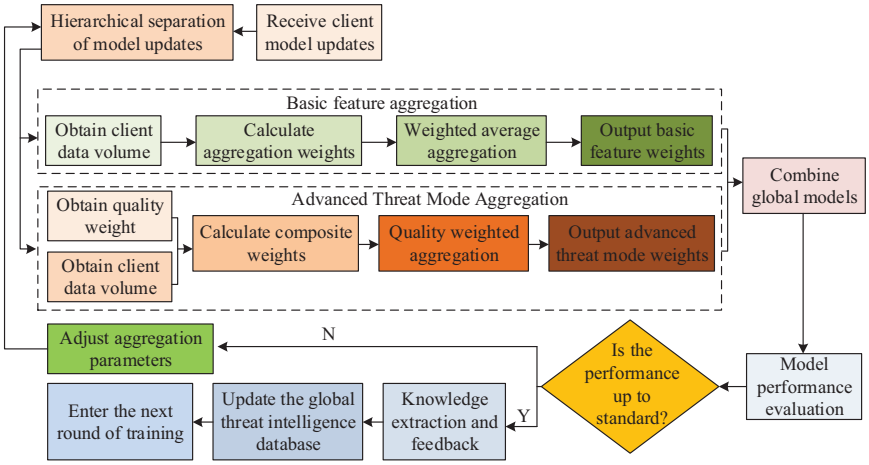

Figure 3 Schematic diagram of hierarchical-based model aggregation strategy.

Figure 3 is a schematic diagram of a hierarchical-based model aggregation strategy. Since the standard federated averaging algorithm weights updates from all clients equally, in cybersecurity scenarios with heterogeneous data, this may cause the model to be biased towards clients with large amounts of data, while ignoring clients with high-value, small-sample threat data. Therefore, the research adopts a hierarchical aggregation strategy, which divides model update into two levels: "basic feature update" and "advanced threat pattern update". The basic feature update corresponds to the shallow and general network behavior feature extractor of the model, and its aggregation adopts the traditional weighted average method, and the weight is proportional to the local data volume of each client:

$$w_{global}^{base,(t+1)} = \sum_{k \in S_t} \frac{|D_k|}{\sum_{j \in S_t} |D_j|} w_k^{base,(t)} \tag{2}$$

where S_t is the set of clients selected in round t, $|\mathrm{D}_k|$ and $|\mathrm{D}_j|$ represent the sizes of the local data sets of clients k and j, respectively [14], k and j are summation indexes, used to traverse all clients in the set S_t, $w_k^{base,(t)}$ is the weight of the basic feature layer in the local model of client k after the t round of local training, $w_{global}^{base,(t+1)}$ is the weight of the global model in the basic feature layer [15]. To update advanced threat patterns corresponding to the deeper layers of the model, the aggregation strategy introduces a quality weight based on the performance of the client's local model on the latest

threat intelligence verification set:

$$w_{global}^{adv,(t+1)} = \sum_{k \in S_t} \frac{\beta_k \cdot |D_k|}{\sum_{j \in S_t} \beta_j \cdot |D_j|} w_k^{adv,(t)} \tag{3}$$

where $w_{global}^{adv,(t+1)}$ is the weight of the global model in the advanced threat mode layer, β_k and β_j are both quality weights, $w_k^{adv,(t)}$ is the weight of the advanced threat pattern layer in the local model [16]. Based on this hierarchical aggregation mechanism, it is ensured that the model not only learns general network behavior rules but also absorbs the latest insights of new advanced threat models. In addition, the threat intelligence management layer serves as the "intelligent hub" of the architecture and interacts closely with the coordination layer. This layer is mainly responsible for providing decision-making basis for client selection and quality assessment benchmark for hierarchical aggregation. It is also responsible for refining the collective knowledge contained in the aggregated model generated during the federated learning process and reversely enriching and updating the global threat intelligence library. Based on the above, the federated learning architecture for collaborative threat analysis is designed.

2.2 Privacy Protection Mechanism Embedding Dynamic Weights and Gradient Perturbations

Although the federated learning architecture based on the above design can protect data privacy at the basic level, in the actual deployment of collaborative threat analysis, the model gradient updates exchanged between the coordination node and participating clients still have the risk of leaking sensitive information. Attackers can reconstruct the characteristics of the original training data by analyzing the transmitted gradient information or implement privacy attacks such as member inference and attribute inference. This poses a serious threat to cybersecurity data that carries sensitive information such as the operating status of critical infrastructure. To strengthen the privacy protection boundary, based on the existing architecture, the research further designs a dual privacy protection mechanism that integrates dynamic weight adjustment and gradient perturbation, and then constructs a privacy protection collaboration model, as shown in Figure 4.

Figure 4 is a schematic diagram of the dual privacy protection mechanism. This mechanism first implements a dynamic weight clipping strategy in the local training phase of the client. Traditional fixed clipping thresholds often

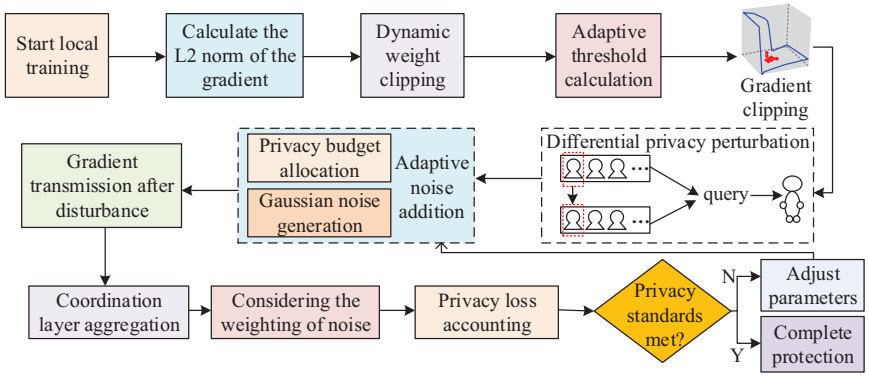

Figure 4 Dual privacy protection mechanism.

make model convergence unstable or have poor privacy protection when faced with non-independent and identically distributed cybersecurity data. The research uses an adaptive clipping threshold determination scheme based on local data distribution characteristics. Specifically, at the beginning of each round of training, each client first calculates the L2 norm of all parameter gradients in the current local model, recorded as $g_k^{(t)}$. Then, based on the sliding average and standard deviation of the client's historical gradient norm, the clipping threshold $C_k^{(t)}$ of this round of training is dynamically adjusted. The specific calculation is

$$C_k^{(t)} = \mu \cdot \mathrm{MA}(g_k)^{(t)} + \sigma \cdot \mathrm{Std}(g_k)^{(t)} \tag{4}$$

where $\mathrm{MA}(g_k)^{(t)}$ represents the sliding average of the historical gradient norm of client k up to the t-th round, $\mathrm{Std}(g_k)^{(t)}$ is the corresponding standard deviation, μ and σ are hyperparameters that control the clipping intensity [17]. After completing gradient clipping, considering that a single clipping operation can only constrain the gradient norm but cannot completely block the leakage of private information contained in the gradient, this study further introduces gradient perturbation based on differential privacy. Different from the traditional method using a fixed noise addition scheme, the study designs a privacy budget adaptive allocation strategy, as shown in Figure 5.

Figure 5 is a schematic diagram of the privacy budget adaptive allocation strategy. Considering the difference in sensitivity to noise at different training stages in the federated learning process, this strategy allocates a larger privacy budget in the early stages of training, allowing relatively small noise to be

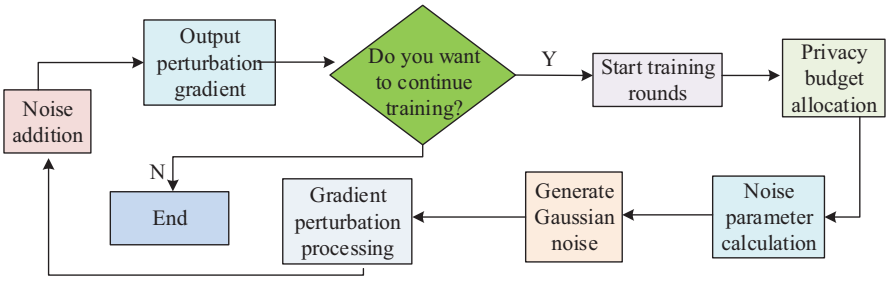

Figure 5 Privacy budget adaptive allocation strategy.

added to promote rapid model convergence. In the later stages of training, as the model gradually stabilizes, the privacy budget is reduced and the intensity of privacy protection is enhanced. The specific noise addition is:

$$\begin{cases} \tilde{g}_k^{(t)} = \text{clip}(g_k^{(t)}, C_k^{(t)}) + \mathcal{N}(0, \sigma^{(t)} \cdot (C_k^{(t)})^2 \mathbf{I}) \\ \sigma^{(t)} = \sqrt{2\ln(1.25/\delta)}/\epsilon^{(t)} \end{cases} \tag{5}$$

where $\sigma^{(t)}$ is the noise coefficient of the t-th round of training, $\tilde{g}_k^{(t)}$ is the gradient generated by client k in round t after clipping and noise perturbation, $\text{clip}(g_k^{(t)}, C_k^{(t)})$ is the gradient clipping function, $g_k^{(t)}$ is the original gradient, $C_k^{(t)}$ represents the radius, $\mathcal{N}(0, \sigma^{(t)} \cdot (C_k^{(t)})^2 \mathbf{I})$ represents a Gaussian distribution with mean 0 and covariance matrix $\sigma^{(t)} \cdot (C_k^{(t)})^2 \mathbf{I}$, δ is the relaxation parameter [18], $\epsilon^{(t)}$ is the privacy budget, which is allocated according to the exponential decay law:

$$\epsilon^{(t)} = \epsilon_{\text{total}} \cdot \gamma^t \tag{6}$$

where ϵ_{total} is the total privacy budget, γ is the decay rate. Based on this design, it is possible to ensure the optimal privacy-utility trade-off under a limited overall privacy budget. In the model aggregation stage, the coordination layer receives the post-perturbation gradient update $\tilde{g}_1^{(t)}, \tilde{g}_2^{(t)}, \ldots, \tilde{g}_K^{(t)}$ from each client and performs a weighted average. This mechanism takes into account both the data quality of each client and the added noise level in the design of the aggregation weight:

$$\Delta w^{(t)} = \sum_{k=1}^{K} \frac{|D_k|/\sigma_k^{(t)}}{\sum_{j=1}^{K} |D_j|/\sigma_j^{(t)}} \cdot \tilde{g}_k^{(t)} \tag{7}$$

where $\Delta w^{(t)}$ is the global model update amount obtained by aggregating all client gradients in the t-th round of coordination layer. Based on this weight

distribution method, clients that add less noise can obtain higher weights in model updates, thereby offsetting the negative impact of noise addition on model performance to a certain extent. To evaluate the overall privacy protection level, the research uses Rényi differential privacy for privacy loss accounting [19]. For the complete federated learning process including T rounds of training, the upper bound of the privacy loss can be expressed as:

$$\epsilon_{\text{total}} \leq \sum_{t=1}^{T} \frac{\alpha}{2(\sigma^{(t)})^2} + \frac{\ln(1/\delta)}{\alpha - 1} \tag{8}$$

where $\alpha > 1$ is the order of Rényi divergence. By carefully designing the noise parameter $\sigma^{(t)}$ of each round, it can be ensured that the overall privacy loss does not exceed the preset privacy budget ϵ_{total} [20]. Based on the above dynamic weight clipping and gradient perturbation mechanism, together with the aforementioned federated learning architecture, a complete privacy protection collaboration model can be formed, as shown in Figure 6.

Figure 6 shows the privacy protection collaboration model. The model maintains efficient gradient updates through dynamic clipping, while achieving provable privacy guarantees with adaptive noise injection, and optimizes final model performance with an aggregation strategy that takes noise levels into account. This collaborative design can ensure that the data privacy of all parties is strictly protected, effectively improve the collective collaborative

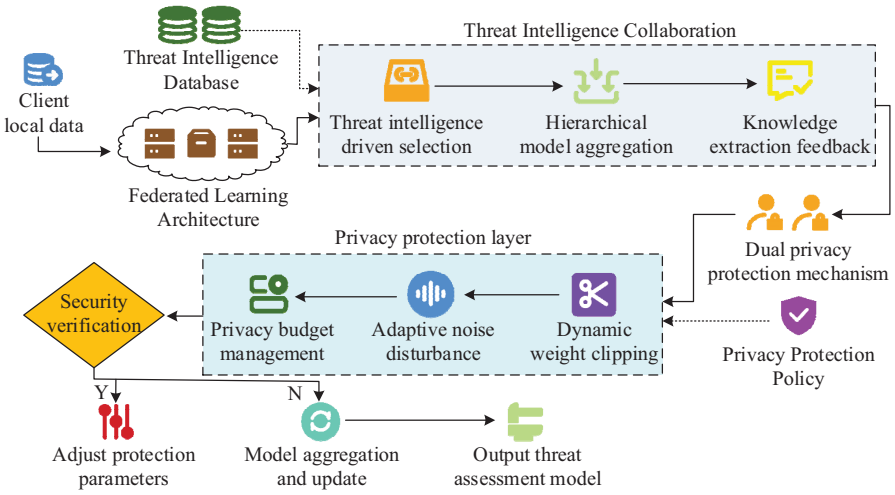

Figure 6 Privacy protection collaboration model.

ability of cyber threats, and provide a technical foundation for building a safe and trustworthy collaborative defense system.

3 Results

3.1 Quantitative Evaluation of Model Performance and Privacy Protection

To quantitatively evaluate the comprehensive effectiveness of the privacy protection collaboration model proposed in the study, the study adopts a comparative method to conduct experiments. The experimental environment is built based on the PyTorch framework and the Flower federated learning library, simulating a federated learning network composed of 1 coordination node and 50 clients. The data base uses the public CIC-IDS-2018 network intrusion detection data set and divides it into each client in a non-independent and identically distributed manner to simulate the differences in threat types and data volumes faced by different organizations in the real world. The experiment selected the Federated Averaging algorithm (FedAvg), the Federated Proximal Optimization (FedProx), and the Differential Privacy-Federated Averaging (DP-FedAvg) that combines fixed gradient clipping and Gaussian noise as the baseline model. All models use the same long short-term memory network as the infrastructure of the local threat assessment model and are trained under the same hyperparameter settings. The total number of federated learning rounds is set to 100 rounds. The evaluation indicators include accuracy, precision, recall and F1-Score. All indicators are calculated on an independent test set composed of threat samples that did not participate in training. The evaluation of privacy protection level strictly follows the differential privacy theory and uses Rényi differential privacy to calculate the cumulative privacy budget consumption. The performance results of each model are presented in Figure 7.

From Figure 7(a), the accuracy of FedAvg, FedProx and DP-FedAvg was 88.5%, 89.3% and 84.1%, respectively, while the proposed model reached 91.8%. In Figure 7(b), the precision of the three was 85.2%, 86.7% and 80.5%, respectively, while the precision of the proposed model was 89.4%. In Figure 7(c), the recall rates of FedAvg, FedProx and DP-FedAvg were 86.1%, 87.0% and 79.8%, respectively, and the proposed model was 90.2%. The results show that the proposed dynamic weight clipping and adaptive privacy budget mechanism achieves a better balance between model utility and privacy protection. To explore the specific impact of the privacy protection mechanism, the study analyzed the performance changes and final

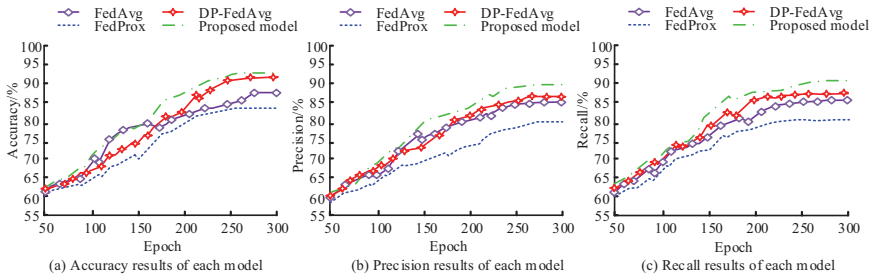

Figure 7 Performance results of each model.

Table 1 Model performance and privacy guarantee under different privacy budget configurations

Privacy Budget (ε) Configuration	Test Accuracy (%)	Rényi Privacy Loss ($\alpha = 2$)	Effective Noise Multiplier (Mean)
$\varepsilon = 1.0$ (Strong Protection)	89.5	0.98	1.28
$\varepsilon = 3.0$ (Medium Protection)	91.8	2.89	0.75
$\varepsilon = 8.0$ (Weak Protection)	92.5	7.65	0.32

privacy protection level of the privacy protection collaboration model under different privacy budget constraints. The experiment set up three different global privacy budgets, corresponding to strong, medium and weak privacy protection intensities, and recorded the final performance and privacy loss of the model under the corresponding configurations, as shown in Table 1.

From Table 1, when configured for strong privacy protection ($\varepsilon = 1.0$), the amount of noise added to the model was the largest, and its accuracy decreased by 2.3% relative to the neutral configuration ($\varepsilon = 3.0$). However, even at this strongest protection level, the accuracy of the proposed model remained at 89.5%, which demonstrated the advantages of the adaptive mechanism. As the privacy budget is relaxed, model performance gradually improves, but the final privacy loss also increases. Under the default neutral configuration ($\varepsilon = 3.0$), the proposed model achieved a threat detection accuracy of more than 91% with provable and limited privacy loss (2.89). This configuration can be considered to achieve a good balance between security and usability in practice. To evaluate the detection capabilities of each model on different types of threats in more detail, the study calculated the F1-Score of each model against four types of typical cyber-attacks on the test set, as shown in Table 2.

The results in Table 2 showed that the F1-Score of the proposed model in four types of cyber-attack detection was better than that of the baseline

Table 2 Comparison of F1-Score of each model on different cyber-attack types (%)

Attack Type	FedAvg	FedProx	DP-FedAvg	Proposed Model
DoS	90.2	90.8	83.1	92.5
Probe	84.6	85.1	76.9	87.8
R2L	78.4	79.0	70.5	82.6
U2R	73.2	74.0	62.8	77.5
Weighted Average	85.6	86.2	77.3	89.1

model, with a weighted average of 89.1%. Especially in R2L and U2R attacks, which were difficult to detect, the model improved by 4.2% and 4.3%, respectively, compared with FedAvg, proving that it could effectively learn complex threat patterns. However, the performance of DP-FedAvg was significantly degraded due to the fixed noise mechanism, which highlighted the advantages of the adaptive privacy protection mechanism.

3.2 Verification of Collaborative Analysis Effectiveness in Cyber Attack and Defense Scenarios

To verify the value of the proposed model in an environment closer to the real world, the study designed a set of dynamic cyber-attack and defense simulation scenarios to evaluate its effectiveness in collaborative threat analysis. This scenario simulates a 30-day continuous attack and defense cycle. The attacker uses a variety of attack tactics including port scanning, brute force cracking, vulnerability exploitation, and intranet lateral movement. The attack strategy evolves dynamically over time. The defenders consisted of five independent organizations, each deploying a client instance of the proposed model, with different local network environments and types of initial attacks. The experiment specially set up an APT attack chain. Each stage of the attack chain will appear dispersedly in the networks of different organizations. It is difficult for any single organization to discover a complete attack picture based on its own data. The core goal of the experiment is to test whether the federated learning collaborative mechanism can enable all participants to identify such complex attacks earlier and more accurately through knowledge sharing than when learning in isolation. The experiment first compared the overall detection rate and average detection time of the APT attack chain by the defender in different modes, as shown in Figure 8.

Figure 8(a) shows the overall detection rate results of APT attacks. When each organization adopted the isolated learning mode, the overall detection rate was the lowest at 71.3%. When using the FedAvg collaboration mode,

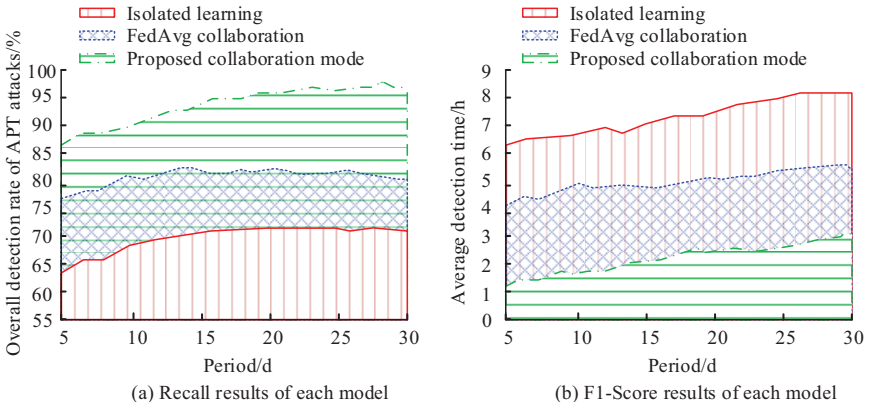

Figure 8 Comparison of detection performance of APT attack chains under different collaboration modes.

Table 3 Comparison of training rounds in each attack stage of APT when it is first detected

APT Attack Stage	Isolated Learning Mode	FedAvg Collaboration	Proposed Collaboration
Initial Reconnaissance	7	5	3
Vulnerability Exploitation	15	11	6
Execution	22	16	9
Lateral Movement	19	13	8
Data Exfiltration	25	18	11

the detection rate was 83.5%. When using the proposed collaboration mode, the detection rate was the highest at 95.6%. Figure 8(b) shows the average detection time results. In the isolated learning mode, the average time from the beginning of the attack to the time it was recognized was the longest, at 8.2 h. In the FedAvg collaboration mode, the average detection time was 5.1 h. In the proposed collaboration mode, the average detection time was the shortest, only 2.8 h. The experiment further analyzed the early detection capabilities of different models at each key stage of the attack chain, as shown in Table 3.

From Table 3, for the initial reconnaissance stage, all models could be discovered relatively quickly, but the proposed model only needed three rounds to sense, which was faster than other modes. In the more critical stages of attack execution and lateral movement, the advantages of the collaboration model was apparent. For example, on “lateral movement”, the isolated learning pattern was not discovered until the 19th round, but the proposed

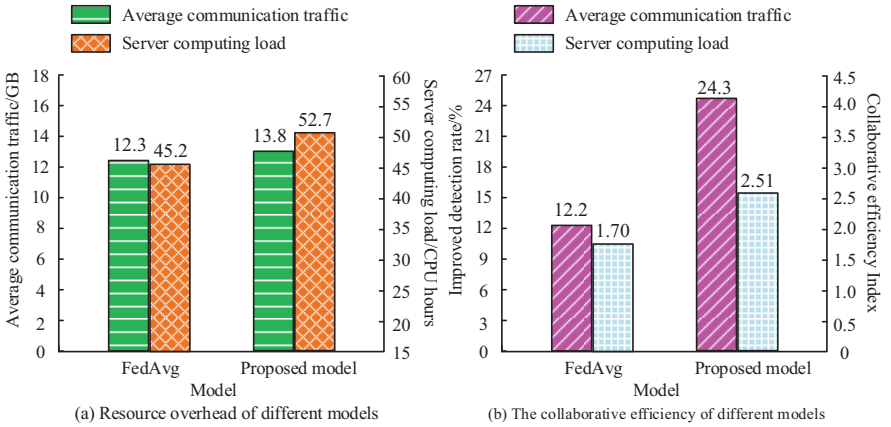

Figure 9 Comparison of resource overhead and collaborative efficiency of different models.

model issued an alert in the 8th round, gaining valuable response time for the defender. This shows that the client selection mechanism based on threat intelligence can preferentially activate those clients with relevant attack puzzles, thereby quickly forming an understanding of new threat patterns in the global model. Finally, the study counted the average network communication traffic generated by each client participating in federated learning and the computing load of the central server, as shown in Figure 9.

Figure 9(a) shows the resource overhead of different models. The average communication traffic of the FedAvg collaboration model was 12.3 GB and the server computing load was 45.2 CPU hours. The average communication traffic of the proposed model was 13.8 GB and the server computing load was 52.7 CPU hours. Figure 9(b) shows the collaborative efficiency of different models. The FedAvg collaboration model increased the detection rate by 12.2%, and the collaborative efficiency index reached 1.70. The proposed model increased the detection rate by 24.3%, and the collaborative efficiency index reached 2.51. This shows that the additional computing resources consumed by the proposed model bring about more cost-effective improvements in security capabilities and have greater advantages in actual operating environments with limited resources.

4 Discussion

To improve the cyber threat collaborative analysis capability under multi-party collaborative while strictly protecting data privacy, this study proposes

a federated learning privacy protection collaboration model that integrates threat intelligence-driven selection and hierarchical aggregation. Experimental results showed that in the threat detection task, the proposed model had an accuracy rate of 91.8%, a precision rate of 89.4%, a recall rate of 90.2%, and an F1-Score of 89.8%, all of which were better than those of the baseline model. In the simulated APT attack and defense scenario, the proposed model increased the overall detection rate of the attack chain to 95.6% and shortened the average detection time to 2.8 hours, which was better than those of the isolated learning mode and FedAvg collaboration mode. The proposed model effectively improves the effectiveness of collaborative defense while ensuring provable privacy. Compared with other research in the same field, the design ideas have obvious differences in problem focus and technical approaches. The federated learning privacy protection framework for edge computing proposed by Wang et al. [21] aims at limited device resources and unstable network in medical IoT scenarios, using lightweight secret sharing and random mask technology to replace computationally intensive homomorphic encryption. Its core lies in balancing privacy protection and computing efficiency through periodic average training strategies to achieve stable operation in edge environments. In contrast, the research is oriented towards cyber threat analysis and judgment scenarios, which are characterized by dynamic evolution of threat patterns and highly heterogeneous data. The hierarchical aggregation strategy designed in this study is able to distinguish basic features from advanced threat patterns, which is in contrast to the unified treatment proposed by Wang et al. This difference reflects the differentiated needs of federated learning architectures in different application fields: Medical IoT focuses more on stability and efficiency, while threat detection requires the ability to deal with concept drift and rare attack patterns. The research by Jagarlamudi et al. [22] revealed a key issue in privacy protection in federated learning from the methodological level: the lack of privacy measurement. Through multi-index joint evaluation, they found that blindly enhancing privacy protection will lead to a significant decline in model performance. This result provides theoretical support for using adaptive privacy budget allocation mechanisms. The proposed model allocates more privacy budget in the early stage of training and gradually tightens it in the later stage. This dynamic strategy is consistent with the security and performance imbalance problem pointed out by Jagarlamudi et al. Experimental results showed that under the medium protection intensity of $\varepsilon = 3.0$, the proposed model only produced a Rényi privacy loss of 2.89, while maintaining a detection accuracy of 91.8%, which proves that the adaptive mechanism can indeed achieve a

better balance between privacy protection and model utility. In addition, Li and Chang [23] used the Takagi Sugeno fuzzy model to model the privacy protection tracking control problem of nonlinear networked control systems. By dynamically quantifying the network transmission pressure and introducing a privacy function protection reference model that converges over time, the synergistic implementation of privacy protection and tracking control dual objectives was verified on a nonlinear mass spring damping system. The idea of coupling privacy mechanisms and control strategies in dynamic environments in this study is inherently consistent with the concept of achieving privacy utility dynamic balance through adaptive differential privacy in this paper. The FedPerGNN federated graph neural network framework proposed by Wu et al. [24] integrates high-order graph information while ensuring privacy through a privacy preserving model update mechanism and graph extension protocol. Compared with existing methods, it reduces errors by 4.0% to 9.6% on six personalized datasets. The design concept of improving model utility through information structure optimization is similar to the client selection mechanism driven by threat intelligence in this article, both reflecting an effective path to achieve performance improvement through information enhancement under privacy protection constraints.

However, due to limitations of the experimental environment and data set size, the federated network constructed by the research only contains 50 clients and relies on a single public data set for performance evaluation. Future research can try to deploy verification in cross-domain and cross-platform real operational networks and explore more efficient asynchronous aggregation mechanisms and lightweight privacy protection technologies to optimize communication and computing overhead.

5 Conclusion

The study proposed a federated learning privacy protection collaboration model for cyber threat analysis. Threat intelligence-driven client selection and hierarchical aggregation strategies improve threat detection capabilities while protecting data privacy. The research shows that the accuracy and comprehensive performance indicators of the proposed model in threat detection tasks are better than traditional federated learning methods. In simulated attack and defense scenarios, the detection rate of APT attack chains has been significantly improved, and the average detection time has been significantly shortened. The model successfully controls privacy loss to a low level under medium privacy protection intensity, achieving a good balance

between privacy protection and model utility. These achievements provide a reliable technical foundation for building a safe and trustworthy collaborative defense system.

Fundings

This research is supported by the National Key Research and Development Program of China under Grant No.2024YFB2906601.

References

[1] Fang X, Chen W, Hu T, Chen Z, Zeng Q, Li J. Federated learning empowered microgrids: Lithium battery state-of-health prediction and multi-node co-optimisation. Journal of Energy Storage, 2025, 139(PB):118892. doi:10.1016/J.EST.2025.118892.

[2] Li X, Zhang J. Multi-source data fusion for real-time cybersecurity situational awareness and visualization. Journal of Cyber Security and Mobility, 2025, 14(4): 955–980. doi:10.13052/jcsm2245-1439.1448.

[3] Zhang H, Meng F, Wang Q. Computer network security system optimization based on improved neural network algorithm and data search. Journal of Cyber Security and Mobility, 2025, 14(1): 75–100. doi:10.13052/jcsm2245-1439.1414

[4] Shen C, Zhang W, Zhou T, Zhang Y, Zhang L. An efficient and secure privacy-preserving federated learning framework based on multiplicative double privacy masking. Computers, Materials & Continua, 2024, 80(3): 4729–4748. doi:10.32604/cmc.2024.054434.

[5] Yang YN, Ao XU, Zhang CJ, Xie T. Research on privacy protection in smart classrooms based on federated multi-task learning. Modern Educational Technology, 2024, 34(9): 123–132. doi:10.3969/j.issn.1009-8097.2024.09.012.

[6] Xu Y, Mao Y, Li S, Li J, Chen X. Privacy-preserving federal learning chain for Internet of Things. IEEE Internet of Things Journal, 2023, 10(20): 18364–18374. doi:10.1109/JIOT.2023.3279830.

[7] Yang Q, Huang A, Fan L, Chan CS, Lim JH, Ng KW, et al. Federated learning with privacy-preserving and model IP-right-protection. Machine Intelligence Research, 2023, 20(1): 19–37. doi:10.1007/s11633-022-1343-2.

[8] Zhu M, Yuan J, Wang G, Xu Z, Wei K. Enhancing collaborative machine learning for security and privacy in federated learning. Journal of Theory

and Practice of Engineering Science, 2024, 4(02): 74–82. doi:10.53469/jtpes.2024.04(02).11.

[9] Tissot H. FormulAI: Designing rule-based datasets for interpretable and challenging machine learning tasks. Artificial Intelligence and Applications. 2025, 3(1): 72–82. doi:10.47852/bonviewAIA42021781.

[10] Zhang K, Li P. Federated learning optimizing multi-scenario ad targeting and investment returns in digital advertising. Journal of Advanced Computing Systems, 2024, 4(8): 36–43. doi:10.69987/JACS.2024.40806.

[11] Wang G, Zhou L, Li Q, Liu X, Wu Y. FVFL: A flexible and verifiable privacy-preserving federated learning scheme. IEEE Internet of Things Journal, 2024, 11(13): 23268–23281. doi:10.1109/JIOT.2024.3385479.

[12] Xie Q, Jiang S, Jiang L, Huang Y, Zhao Z, Khan S. Efficiency optimization techniques in privacy-preserving federated learning with homomorphic encryption: A brief survey. IEEE Internet of Things Journal, 2024, 11(14): 24569–24580. doi:10.1109/JIOT.2024.3382875.

[13] Akash TR, Lessard NDJ, Reza NR, Islam MS. Investigating methods to enhance data privacy in business, especially in sectors like analytics and finance. Journal of Computer Science and Technology Studies, 2024, 6(5): 143–151. doi:10.32996/jcsts.2024.6.5.12.

[14] Lyu M, Ni Z, Chen Q, Li F. Edge-DPSDG: An edge-based differential privacy protection model for smart healthcare. IEEE Transactions on Big Data, 2024, 11(1): 21–34. doi:10.1109/TBDATA.2024.3366071.

[15] Sharma P, Sharma SK, Dani D. Edge-assisted federated learning for anomaly detection in diverse IoT network. International Journal of Information Technology, 2025, 17(5): 3035–3045. doi:10.1007/s41870-024-01728-x.

[16] Xie H, Zhang Y, Zhongwen Z, Zhou H. Privacy-preserving medical data collaborative modeling: A differential privacy enhanced federated learning framework. Journal of Knowledge Learning and Science Technology, 2024, 3(4): 340–350. doi:10.60087/jklst.v3.n4.p340.

[17] Ali W, Din IU, Almogren A, Rodrigues JJPC. Federated learning-based privacy-aware location prediction model for internet of vehicular things. IEEE Transactions on Vehicular Technology, 2024, 74(2): 1968–1978. doi:10.1109/TVT.2024.3368439.

[18] Kuang Y, Jiang B, Cui X, Li S, Liu Y, Song H. Flexible differential privacy for internet of medical things based on evolutionary learning. IEEE Internet of Things Journal, 2024, 11(9): 16954–16968. doi:10.1109/JIOT.2024.3366889.

[19] Sinthiya C. Federated learning architectures for privacy-preserving collaborative intelligence in distributed networks. International Journal of Computer Science and Engineering Innovations, 2025, 1(1): 27–37. doi:10.64137/31079458/IJCSEI-V1I1P104.

[20] Wei M, Yang J, Zhao Z, Zhang X, Li J, Deng Z. Defedhdp: Fully decentralized online federated learning for heart disease prediction in computational health systems. IEEE Transactions on Computational Social Systems, 2024, 11(5): 6854–6867. doi:10.1109/TCSS.2024.3406528.

[21] Wang R, Lai J, Zhang Z, Li X, Vijayakumr P, Karuppiah M. Privacy-preserving federated learning for internet of medical things under edge computing. IEEE Journal of Biomedical and Health Informatics, 2022, 27(2): 854–865. doi:10.1109/JBHI.2022.3157725.

[22] Jagarlamudi GK, Yazdinejad A, Parizi RM, Pouriyeh S. Exploring privacy measurement in federated learning. The Journal of Supercomputing, 2024, 80(8): 10511–10551. doi:10.1007/s11227-023-05846-4.

[23] Li M, Chang X. Fuzzy tracking control for discrete-time nonlinear network systems with privacy protection and dynamic quantization. International Journal of Fuzzy Systems, 2023, 25(3): 1227–1238. doi:10.1007/s40815-022-01436-3.

[24] Wu C, Wu F, Lyu L, Qi T, Huang Y, Xie X. A federated graph neural network framework for privacy-preserving personalization. Nature Communications, 2022, 13(1): 3091. doi:10.1038/s41467-022-30714-9.

Biographies

Rui Zhuang obtained her Ph.D. in Cyber Science and Technology in 2024 from the University of Science and Technology of China. Presently, she is

working as an Engineer in the Department of Fundamental Network Technology, China Mobile Research Institute.

Xiling Yang is an Associate Professor in the College of Business Administrations, Henan Finance University, Zhengzhou, China.