Hybrid Federated Learning Framework for APT Attack Chain Detection and Privacy Enhancement

Jie Ji[1*], Shi Qiu[2], Shengpeng Ye[1], Xin Liu[1]
[1]College of Information Engineering, Yangzhou Polytechnic Institute, Yangzhou 225127, China
[2]Yangzhou Polytechnic Institute, Yangzhou 225127, China
E-mail: jij@ypi.edu.cn
*Corresponding author

Keywords: APT attack, federated learning, privacy enhancement, adaptive differential privacy

Received: January 25, 2026

To address the issues of cross-domain data privacy protection and collaborative analysis in the detection of Advanced Persistent Threat (APT) attack chains, this paper proposes an APT attack collaborative detection and privacy enhancement method based on hybrid federated learning. This method constructs a two-layer federated learning framework of horizontal cross-organization collaboration and vertical multi-feature fusion, realizing the deep fusion of multi-source threat intelligence under the premise of privacy protection. By designing an adaptive differential privacy mechanism and dynamically optimizing the noise addition strategy, it minimizes the loss of model performance while guaranteeing data security. At the same time, this paper introduces a time-series graph neural network detection model to achieve accurate perception and correlation analysis of multi-stage behaviors of APT attacks. The experimental results demonstrate that HybridFL-APT achieves a macro-averaged F1-score of 0.914 and an attack stage identification accuracy of 0.867. Compared to the standard FedAvg algorithm, the proposed framework reduces the cumulative communication overhead by 30.9% while maintaining robust privacy protection even at a strict privacy budget (ε=0.1).

Povzetek: Članek predstavi hibridni federativni model za zaznavanje APT napadov, ki izboljša natančnost analize ob hkratnem varovanju zasebnosti podatkov.

1 Introduction

With the continuous deepening of the digitization process, the threats faced by the cyberspace are increasingly showing a trend of complexity and sophistication. Advanced Persistent Threat (APT) [1] [2] , due to its long-term concealment, clear targeting, and multi-stage attack chain, has become a major challenge in the field of cybersecurity. The traces of such attacks are usually scattered in different network regions and devices, forming cross-domain and heterogeneous "data silos", resulting in dilemmas for traditional detection methods, which require centralized data processing in terms of privacy compliance and collaborative efficiency.

As a "data remains unchanged, model changes" distributed learning paradigm, federated learning provides a new approach for cross-domain collaborative analysis while protecting data privacy. Existing research has applied it to scenarios such as malicious traffic identification. However, it is mostly limited to horizontal federated architectures or the detection of single attack types. This makes it difficult to effectively model the complex associations of multi-stage behaviors in the APT attack chain and fails to fully integrate multi-dimensional heterogeneous features such as network traffic and endpoint logs. In addition, there is a risk of privacy leakage through parameters in the federated learning process itself [3] . Current privacy protection schemes mostly adopt fixed noise addition mechanisms, which are difficult to balance detection accuracy and protection intensity in dynamic and complex network environments, limiting their practicality and reliability in high-demand scenarios such as APT detection.

To evaluate the effectiveness of the proposed framework, this paper focuses on the following three Research Questions (RQs):

- RQ1: How to effectively fuse cross-organization heterogeneous features while strictly maintaining data privacy?
- RQ2: Can the integration of a temporal-graph neural network accurately correlate multi-stage APT attack behaviors?
- RQ3: How does the adaptive differential privacy mechanism optimize the trade-off between detection utility and privacy intensity?

Therefore, this paper proposes a hybrid federated learning collaborative detection framework for the APT attack chain. By designing a two-layer learning mechanism that combines horizontal and vertical federated fusion, this framework achieves deep fusion and knowledge sharing of cross-domain heterogeneous features while protecting the data privacy of each participating party. An adaptive differential privacy mechanism is introduced to optimize the trade-off

between privacy and utility. A temporal-graph neural network model is constructed to accurately depict the dynamic evolution process and entity association relationships of the attack chain. Experimental results show that this framework exhibits considerable benefits in terms of detection accuracy, privacy protection, and communication efficiency.

2 Related research work

Regarding the core challenges in APT attack detection, such as data privacy, cross-domain collaboration, and multi-stage correlation analysis, which are raised in the introduction, this section will systematically review and comment on the existing research from three aspects: APT detection technology, secure applications of federated learning, and privacy-enhanced computing [4] [5] [6] [7] .

In the field of APT detection technology, the mainstream methods have gradually evolved from rule signature-based to machine learning-based methods. In recent years, deep learning models, with their powerful feature learning capabilities, have been widely used to identify complex attack patterns from massive logs and network traffic. However, most existing methods need a centralized data analysis architecture, that is, the raw data distributed everywhere needs to be aggregated to a central server [8] [9] [10] [11] . This mode not only faces serious privacy leakage and compliance risks due to cross-domain data transmission, but also has difficulty breaking the "data silos" formed by organizational barriers, resulting in limited detection horizons and making it difficult to construct a complete panoramic view of the attack chain.

Federated learning provides a new solution to the above problems. As a privacy-preserving distributed machine learning paradigm, federated learning allows participating parties to train models locally and share only model parameters instead of raw data, which has been verified in scenarios such as intrusion detection and malicious traffic classification [12] [15] . However, existing research on the application of federated learning in the field of network security mainly focuses on the horizontal federated scenario, that is, each participating party has the same data feature space, which cannot effectively handle the situation where different institutions may hold heterogeneous feature data in reality. In addition, current research mostly focuses on the detection of independent attack events and lacks the ability to conduct end-to-end modeling and collaborative analysis of the "attack chain" unique to APT attacks, which has strong logical and temporal correlations among multiple stages.

To ensure the privacy and security of the federated learning process, technologies such as differential privacy and homomorphic encryption have been widely introduced. Among them, differential privacy has become the mainstream choice due to its rigorous mathematical definition and low computational overhead [16] . Recent advancements in nonlinear optimal control and synchronization of chaotic systems have emphasized the importance of adaptive feedback in managing environmental uncertainties [23] [25] . This paralleled shift toward data-driven, real-time control provides a promising theoretical framework for balancing privacy and model utility in federated learning. Existing methods usually adopt a noise addition mechanism with a fixed intensity (such as a preset privacy budget ε). However, this static privacy protection strategy often leads to a dilemma where the model utility is significantly reduced or the privacy protection is insufficient when dealing with tasks that are highly sensitive to model accuracy, such as APT detection, and it is difficult to achieve an intelligent balance between privacy protection and performance in a dynamic and complex adversarial environment.

Therefore, when dealing with APT attacks, existing research still lacks a federated learning solution that can synergistically utilize cross-domain heterogeneous data features, deeply model the multi-stage attack associations, and adaptively balance the detection accuracy and privacy protection intensity. The research of this paper aims to fill this research gap. To clearly demonstrate the research gap, we summarize the comparative analysis of existing frameworks in Table 1, focusing on data heterogeneity, detection depth, and privacy intensity.

Table 1: Comparison of HybridFL-APT and state-of-the-art federated learning methods in APT detection

Method	Horizontal Collab	Vertical Fusion	Multi-stage Correlation	Privacy Protection
FedAvg	Yes	No	Low	Basic
VFL-LR	No	Yes	Low	Encryption
FedAttackChain	Yes	No	Medium	Fixed DP
HybridFL-APT (Ours)	Yes	Yes	High (T-GNN)	Adaptive DP

3 Design of hybrid federated APT detection framework

Aiming at the deficiencies of existing methods in cross-domain collaboration and multi-stage correlation analysis, this paper proposes a hybrid federated learning detection framework for APT attack chains [10] [17] [19] [20] . By constructing a two-layer learning structure of horizontal cross-organization collaboration and vertical multi-source feature fusion, and integrating a lightweight adaptive privacy protection mechanism, the framework aims to achieve efficient and accurate perception of APT attack chains under the premise of privacy protection. The framework design follows three principles: first, privacy protection is achieved by replacing the sharing of original data with the exchange of model parameters; second, multi-dimensional threat intelligence fusion is realized through the organic combination of horizontal

and vertical federations; third, the privacy protection and detection utility are dynamically balanced through an adaptive mechanism. The overall architecture is shown in Figure 1, demonstrating the interaction relationships among all participants, the coordination server, and each functional module.

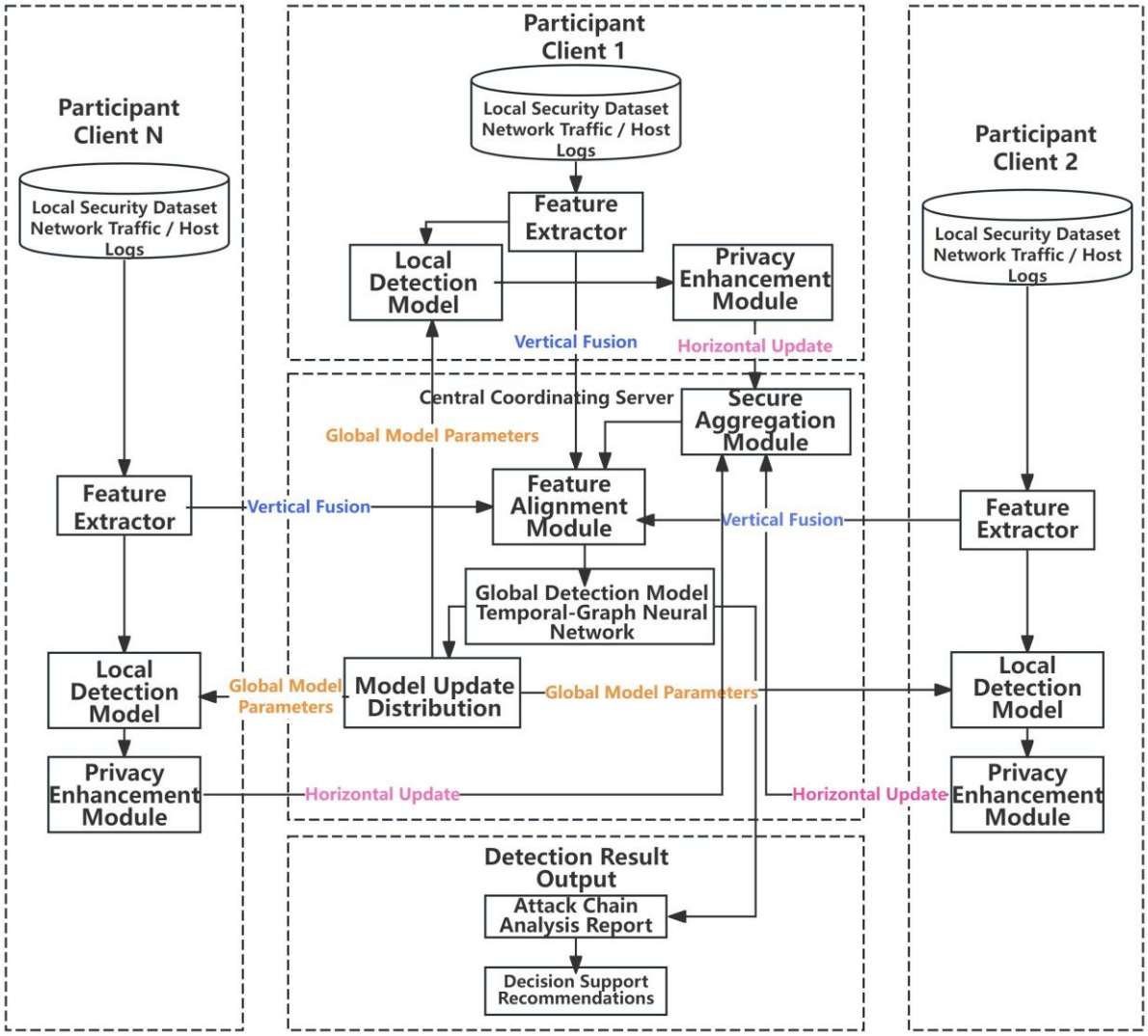

Figure 1: Overall architecture of the hybrid federated APT detection framework

3.1 Horizontal cross-organization federated collaboration mechanism

At the horizontal level, the framework supports multiple independent security domains to participate in collaborative training. Each participant C_i holds a local dataset $D_i = \{(X_j, y_j)\}_{j=1}^{n_i}$, with the same feature space but different samples, representing the observed instances in their respective network environments. Each participant maintains a copy M_i of the global detection model M_{global} locally and trains it by minimizing the local loss function $\mathcal{L}_i(\theta) = \frac{1}{n_i}\sum_{j=1}^{n_i} \ell\big(f(X_j;\theta), y_j\big)$, where θ are the model parameters, $f(\cdot;\theta)$ is the model prediction function, and $\ell(\cdot)$ is the loss function.

After the training is completed, the participating parties encrypt the update amount $\Delta\theta_i = \theta_i^{new} - \theta_i^{old}$ of the model parameters and transmit it to the central coordination server. The server aggregates the update amounts using an improved federated averaging algorithm. To mitigate the negative impact of data distribution, quality, and quantity differences on the performance of the global model, this mechanism designs a dynamic weight allocation strategy. The aggregation weight ω_i is jointly determined by the data volume n_i and the confidence factor α_i based on the local model performance evaluation. The confidence factor α_i is defined as follows:

$$\alpha_i = \frac{Acc_i^{val} - Acc_{min}^{val}}{Acc_{max}^{val} - Acc_{min}^{val}} + \beta$$

Among them, Acc_i^{val} represents the validation set accuracy of Party C_i after local training in this round. Acc_{max}^{val} and Acc_{min}^{val} are the highest and lowest validation set accuracies among all parties, respectively. β is a smoothing constant, which is used to avoid zero weights.

The final aggregation formula is as follows:

$$\theta_{global}^{t+1} = \theta_{global}^{t} + \sum_{i=1}^{N} \frac{\alpha_i \cdot n_i}{\sum_{j=1}^{N} \alpha_j \cdot n_j} \Delta\theta_i$$

3.2 Vertical multi-source feature federated fusion mechanism

At the vertical level, the framework aims to fuse heterogeneous feature views for the same set of network entities, such as traffic features from network security devices and process behavior features from endpoint detection and response systems. This fusion process starts with sample alignment under privacy protection [21] [22] . Under the unified coordination of the coordination server, each participating party uses a private set intersection protocol based on blind signature or oblivious pseudorandom function to securely calculate the set of sample identifiers I_{common} that are common to all parties without revealing the ID of their nonintersecting samples.

For the aligned common samples $i \in I_{common}$, assume there are *K* parties, and the *k*-th party holds the feature vector $x_i^{(k)}$. Each party first encrypts the local feature transformation result (e.g., through a lightweight embedding layer $g_k(\cdot)$) using a homomorphic encryption algorithm to obtain $[\![h_i^{(k)}]\!] = Enc\left(g_k\left(x_i^{(k)}\right)\right)$, and uploads the ciphertext to the server. The server then performs an aggregation operation in the ciphertext state:

$$[\![H_i]\!] = \sum_{k=1}^{K} [\![h_i^{(k)}]\!]$$

The aggregated ciphertext $[\![H_i]\!]$ can only be decrypted by the designated party with the private key to obtain the fused feature representation H_i, which is used to update the shared layer responsible for decision-making in the global model. The global model M_{global} is designed to include *K* parallel feature encoding sub-networks $\{g_k\}_{k=1}^{K}$ and a top-level fusion network $F(\cdot)$. Finally, the prediction result for the sample i is $\hat{y}_i = F(H_i)$. This design enables the model to learn the complementarity and correlation between different feature spaces end-to-end, which is of great significance for identifying the hidden patterns in the APT attack chain that only become apparent under cross-dimensional correlation analysis.

3.3 Lightweight adaptive privacy enhancement module

To defend against potential gradient leakage attacks during the federated learning process, the framework in this paper integrates a lightweight adaptive differential privacy module. The core of this module lies in the dynamic noise injection mechanism. After each round of local training, the participant first calculates the parameter update $\Delta\theta_i$, and performs norm clipping on it to limit the sensitivity. The specific operation is: $\overline{\Delta\theta_i} = \Delta\theta_i / max\left(1, \frac{\|\Delta\theta_i\|_2}{C}\right)$, where C is the clipping threshold.

The noise scale σ_i is not a fixed value, but is dynamically adjusted according to the norm of the update vector in this round and the preset privacy budget[4] [7] . The specific calculation formula is as follows:

$$\sigma_i = \frac{C \cdot \sqrt{2\ln(1.25/\delta)}}{\epsilon_{round}}$$

Among them, ϵ_{round} represents the privacy budget allocated to the current training round, while δ represents a very small slack probability. The design of the adaptive noise scale σ_i draws inspiration from robust adaptive control theories in complex time-varying systems [23] [24] [24] [25] . Similar to how neural-network-based output feedback ensures stability for MIMO delay systems [28] , our ADP mechanism treats the privacy-induced noise as a dynamic disturbance, ensuring that the model updates remain within stable convergence bounds even when data is incomplete or cross-domain heterogeneity is high [26] [27] . To optimize the consumption of the total privacy budget ϵ_{total}, we allocate the privacy budget round by round using a geometric sequence:

$$\epsilon_{round} = \epsilon_{total} \cdot \frac{\gamma^{t-1}(1-\gamma)}{1-\gamma^{T}}$$

Among them, $\gamma \in (0,1)$ is the attenuation factor and T is the total number of training rounds. This allocation method allocates relatively more budget in the initial stage of training to quickly learn the general pattern; while in the later stage, the budget is tightened to strengthen the protection of individual data. The parameter update form after adding noise is $\widetilde{\Delta\theta_i} = \overline{\Delta\theta_i} + N(0, \sigma_i^2 I)$. The entire training process satisfies $(\epsilon_{total}, \delta)$-differential privacy. To further reduce the communication overhead and the potential risk of information leakage, this module also performs Top-K sparsification on $\widetilde{\Delta\theta_i}$ before encryption and only uploads the top p% gradient values with the largest absolute values.

3.4 Temporal-graph neural network detection model oriented to the attack chain

As the core of the global detection model, this paper designs a neural network that deeply integrates temporal and structural information to specifically model the complex and multi-stage behavior patterns in the APT attack chain.

Sequential behavior modeling: For each monitored entity, the ordered sequence of security events $S = [e_1, e_2, \ldots, e_T]$ generated within a certain time window is input into a bidirectional long short-term memory network (Bi-LSTM) for modeling.

$$\overrightarrow{h_t} = LSTM\left(e_t, \overrightarrow{h_{t-1}}\right), \overleftarrow{h_t} = LSTM\left(e_t, \overleftarrow{h_{t+1}}\right)$$

The final sequential feature is represented as $h_{time} = \left[\overrightarrow{h_T}; \overleftarrow{h_1}\right]$.

Entity-Relationship Diagram Modeling: We model the entities and their interaction relationships in the cyber space as a heterogeneous graph $G = (V, \varepsilon, R)$,where V represents the set of entities, ε represents the set of edges, and R represents the set of relationship types. In this paper, we use the Relational Graph Convolutional Network (R-GCN) to learn the structural features of entities in the attack context:

$$h_v^{(l+1)} = \sigma\left(\sum_{r\in R}\sum_{u\in\mathcal{N}_v^r}\frac{1}{|\mathcal{N}_v^r|}W_r^{(l)}h_u^{(l)} + W_0^{(l)}h_v^{(l)}\right)$$

Among them, $\mathcal{N}_v^r$ represents the set of neighbors that have the relationship r with the entity v. This layer can reveal the lateral movement path of the attack and the link of data leakage.

Feature Fusion and Decision: The temporal features h_{time} are fused with the graph structure features $h_v^{(L)}$ after being propagated through L layers, and a gated fusion mechanism is used for processing:

$$g = \sigma\left(W_g\left[h_{time}; h_v^{(L)}\right] + b_g\right)$$
$$z = g \odot h_{time} + (1-g) \odot h_v^{(L)}$$

Among them, $\odot$ represents element-wise multiplication, and g is an adaptive gating vector. The fused feature z is finally input into a multi-layer perceptron for classification, and the probability distribution belonging to each attack stage or normal state is output.

4 Experimental design and results analysis

To comprehensively evaluate the effectiveness, efficiency, and privacy protection capabilities of the proposed hybrid federated APT detection framework (hereinafter referred to as HybridFL-APT), this paper designs and implements a series of comparative experiments and ablation experiments. The experiments are carried out around the following three core issues: First, whether HybridFL-APT is superior to existing baseline methods in terms of the detection accuracy of multi-stage APT attacks; Second, whether the hybrid federated architecture and the adaptive privacy mechanism can effectively balance privacy protection and model utility; Third, whether the framework can meet the actual deployment requirements in terms of communication overhead and scalability.

4.1 Experimental setup

4.1.1 Dataset and preprocessing

This experiment uses two datasets to construct a simulation environment. First, the CIC-IDS2018 dataset is selected. This dataset contains various modern attack scenarios, and the attack process has significant multi-stage characteristics. We select the "Penetration Testing" and "Botnet C&C Communication" parts from it, and map their attack steps to each stage of the ATT&CK framework, constructing a label sequence covering six stages: reconnaissance, resource development, initial access, execution, persistence, and command and control. Second, a custom enterprise log dataset is constructed. To simulate the heterogeneous characteristics in vertical federation, based on the publicly available audit log format, a set of machine system logs aligned with the timestamps of CICIDS2018 is synthesized, with features including process creation, network connection, file access, and permission change, forming a view of terminal behavior. Finally, we align the network traffic features and host log features in time, constructing a mixed dataset containing 120,000 samples, where the normal traffic accounts for 70% and the multi-stage attack traffic accounts for 30%. The dataset is divided into a training set, a validation set, and a test set in chronological order, with proportions of 60%, 20%, and 20% respectively.

4.1.2 Comparison methods and evaluation metrics

To verify the superiority of the proposed framework, we set up five groups of comparison methods: Centralized (traditional centralized training), FedAvg (the standard horizontal federated averaging algorithm), VFL-LR (vertical federated logistic regression method based on secure aggregation), FedAttackChain (federated learning method for attack detection), and the proposed HybridFL-APT. Additionally, HybridFL-APT without differential privacy (DP) was set as the control group for the ablation experiment. It is a variant with the adaptive privacy module removed.

The evaluation metrics cover three aspects. In terms of detection performance, macro-averaged precision, macro-averaged recall, macro-averaged F1 score, and weighted accuracy in attack phase recognition are adopted. Regarding the trade-off between privacy protection and utility, the differential privacy budget ε consumed by each federated method to reach the target detection performance is recorded, and the actual privacy leakage risk is quantified via the success rate of membership inference attacks.

4.1.3 Federal environmental configuration

The simulation environment consists of 1 central server and 10 participating client nodes. To simulate non-independent and identically distributed as well as feature-heterogeneous situations in reality, two data partitioning methods are adopted: Horizontal partitioning distributes samples to 10 clients in a non-equal random manner to simulate the difference in data volume. Vertical partitioning randomly designates 5 clients to only have network traffic features and the other 5 clients to only have host log features. All experiments are run on a server equipped with an NVIDIA RTX 4090 GPU and implemented using the PyTorch and PySyft frameworks.

4.2 Comparative analysis of detection performance

Table 2 compares the performance of various methods on the APT attack chain detection task. The results demonstrate that the proposed HybridFL-APT model achieves an F1-score of 0.914, significantly outperforming FedAvg and VFL-LR. This validates the superiority of the hybrid architecture in effectively integrating both horizontal and vertical information. Even when compared to the representative state-of-the-art method FedAttackChain, HybridFL-APT improves the F1-score by approximately 2.9%. Furthermore, the model's performance approaches the ideal Centralized baseline, indicating that the hybrid federated collaboration mechanism can effectively learn representations approximating the global data distribution while strictly

adhering to privacy constraints. Regarding the critical metric of attack phase identification accuracy, HybridFL-APT scores 0.867, outperforming all competitors. This highlights the effectiveness of the temporal-graph neural network in capturing the logical sequence and propagation paths of attack chains. Finally, while a variant without the privacy module produces marginal performance gains, the gap remains negligible. This suggests that the impact of the adaptive privacy mechanism on the overall model utility is maintained at a minimal level.The improvement over FedAttackChain (0.914 vs. 0.885 F1-score) marks a significant step forward for threat intelligence. FedAttackChain is limited to horizontal collaboration and cannot see patterns across different infrastructure domains. In contrast, HybridFL-APT uses a hybrid fusion approach to uncover hidden correlations in the attack chain. This method effectively bridges "data silos" while keeping raw data secure.

Table 2: Performance comparison of different methods in APT attack chain detection

Method	Precision	Recall	F1-score	Stage Accuracy
Centralized	0.923	0.898	0.910	0.854
FedAvg	0.882	0.845	0.863	0.812
VFL-LR	0.871	0.902	0.886	0.801
FedAttackChain	0.894	0.876	0.885	0.838
HybridFL-APT (Ours)	0.916	0.912	0.914	0.867
HybridFL-APT w/o DP	0.921	0.910	0.915	0.869

4.3 Privacy-utility trade-off and robustness analysis

To quantitatively evaluate the privacy protection effect of the evaluation framework and its impact on model performance, different intensities of privacy budgets ε are set, and the performance and security of each federated learning method are compared. Table 3 records the detection performance of different methods and the success rate of membership inference attacks they face under four typical privacy budget levels.

We conduct a sensitivity analysis by varying the privacy budget ε As shown in Table 3, the model exhibits high stability. When the privacy constraint is tightened from ε=5.0 to ε=0.1 (increasing noise intensity), the F1-score of HybridFL-APT only decreases from 0.912 to 0.894. This marginal 2.2% performance degradation confirms that our adaptive noise addition strategy ensures model convergence and stability under varying noise levels. This stability confirms that the system's 'flatness' is maintained across different privacy regimes, aligning with the robustness criteria observed in nonlinear optimal control applications.

Table 3: Privacy-utility trade-off of each method under different privacy budgets ε

Method	ε=0.1	ε=1.0	ε=5.0	ε=∞
	F1 Score / MIA Success Rate	F1 Score / MIA Success Rate	F1 Score / MIA Success Rate	F1 Score / MIA Success Rate
Centralized	-/-	-/-	-/-	0.910 / 35.2%
FedAvg (without DP)	-/-	-/-	-/-	0.863 / 27.8%
Fixed-DP-FL	0.819 / 3.1%	0.857 / 8.5%	0.878 / 19.4%	0.885 / 26.1%
HybridFL-APT	0.894 / 4.3%	0.905 / 7.1%	0.912 / 11.6%	0.914 / 15.7%

As can be seen from Table 3, first, for the Centralized and FedAvg methods without any privacy protection mechanisms, their models are extremely vulnerable to membership inference attacks and have a very high risk of leakage. Second, under the strong privacy constraint, the baseline method using the fixed noise mechanism can suppress the attack success rate to a relatively low level (3.1%), but the model utility is severely damaged, and the F1 score drops by about 7.5% compared to the unprotected state. In contrast, under the same strong privacy constraint ε, the adaptive privacy mechanism proposed in this paper can still maintain a high F1 score of 0.894, with a significantly smaller performance loss than the fixed mechanism, while effectively controlling the attack success rate below 5%. As the privacy constraint is relaxed, the adaptive mechanism can more efficiently convert the newly added privacy budget into an improvement in model performance and demonstrates a better utility-privacy balance at all privacy levels. This proves that the adaptive noise mechanism can intelligently allocate privacy budgets according to the training dynamics, effectively avoiding the problems of "insufficient protection" or "over protection" that may occur in the global and local stages of training for the fixed mechanism.

4.4 Analysis of communication and computational efficiency

Table 4 presents the system efficiency comparison results. While the per-round training time of HybridFL-APT (20s) is slightly higher than FedAvg due to the computational overhead of T-GNN and cross-domain alignment, the framework significantly outperforms baselines in terms of total resource consumption.

- **Comparative Analysis:** Unlike traditional horizontal FL methods that suffer from slow convergence in heterogeneous environments, HybridFL-APT reaches the convergence criterion (F1 = 0.90) in only 52 iterations, a 38.8% reduction compared to the 85 rounds required by FedAvg. (As shown in Figure 2) Like robust adaptive control, our method prioritizes quality over speed in each step. By taking the time to capture complex patterns, the system follows a more stable path and converges much earlier.
- **Deployment Scenarios and Limitations:** In practical deployment, the sub-linear growth of communication cost (as shown in Figure 5) makes our framework highly suitable for large-scale enterprise networks where bandwidth is a constraint. However, a potential limitation exists in extremely high-latency environments where the two-layer alignment process may introduce synchronization delays. Future work will investigate asynchronous updates to further mitigate this.

Table 4: System efficiency comparison (Convergence criterion: F1 score = 0.90)

Method	Comm Cost (MB)	Convergence Rounds	Time per Round (s)
FedAvg	1520	85	18
VFL-LR	980	60	22
HybridFL-APT	1050	52	20

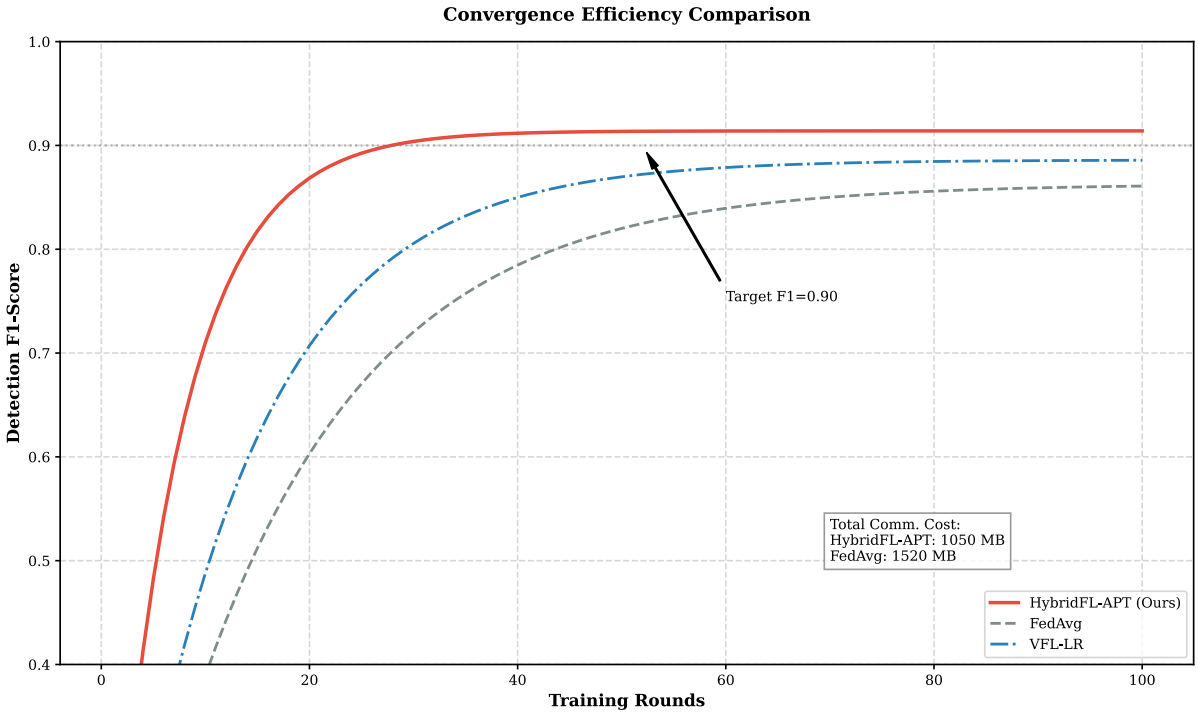

Figure 2: Convergence efficiency comparison

4.5 Ablation experiments and key parameter analysis

The ablation experiments further verified the necessity of each core component. After removing the vertical fusion, the F1 score dropped to 0.887, and the attack phase recognition accuracy dropped to 0.831. This indicates that multi-source feature fusion is crucial for understanding the complete attack chain. After removing the horizontal collaboration, the model performance was close to that of a single vertical federated method, suggesting that cross-organization knowledge sharing helps improve the generalization ability of the model. When using only the time series model, the recognition accuracy of the attack phase dropped to 0.841. When using only the graph model, it dropped to 0.819. This indicates that the two are

complementary and irreplaceable in the process of modeling the attack chain.

4.6 Supplementary data analysis and visualizations

To provide additional information about the performance and characteristics of the suggested HybridFL-APT framework, this section presents further data analyses supported by visualizations. The following figures show key trends and comparative results across detection accuracy, privacy-utility trade-offs, and system efficiency.

4.6.1 Detection performance across attack stages

Figure 3 compares the detection accuracy of HybridFL-APT at each stage. It employs two strong baselines for comparison: centralized training and the standard horizontal method FedAvg. These results are based on the six-stage attack chain from the ATT&CK framework. HybridFL-APT achieves accuracy close to the centralized upper bound at each stage. It also significantly outperforms FedAvg in the later stages, which include Command & Control and Exfiltration. Behavioral patterns in these stages are more subtle and dispersed. This indicates that our framework preserves crucial discriminating information effectively. This preservation occurs during privacy-preserving collaborative learning.

As shown in Figure 3, our model performs best in the late stages of an attack, such as C&C and Exfiltration. These stages are usually hard to detect because they look like normal admin traffic to a single organization. By using a T-GNN, our framework reaches an accuracy level close to centralized training. This proves that keeping track of timing and structural relationships during federated learning is the best way to catch coordinated, multi-stage attacks.

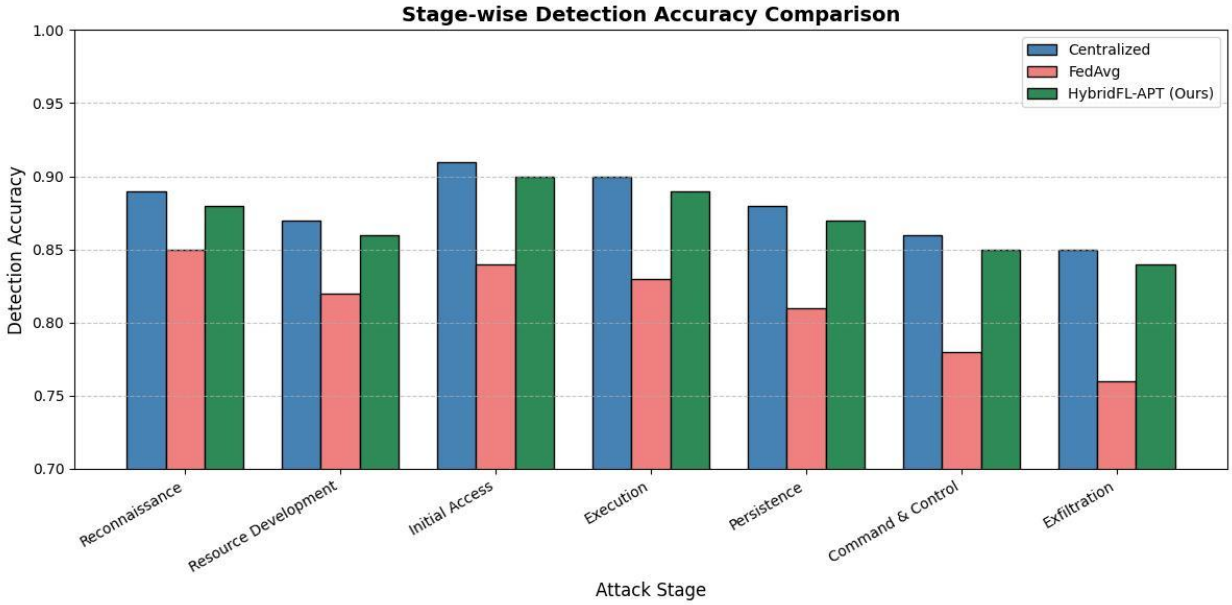

Figure 3: Detection accuracy comparison across attack stages

4.6.2 Privacy-utility trade-off under varying data heterogeneity

Figure 4 shows the variation in the privacy-utility trade-off of HybridFL-APT with varying degrees of data heterogeneity among participants. Heterogeneity is measured as the variance of local data distributions (earth mover distance). The curves indicate that when heterogeneity increases, a larger privacy budget (ε) is needed to achieve the same detection performance. However, the adaptive noise mechanism in HybridFL-APT can alleviate this effect, so the framework can maintain higher utility under strong privacy constraints ($\varepsilon \leq 1.0$) than fixed DP mechanisms.

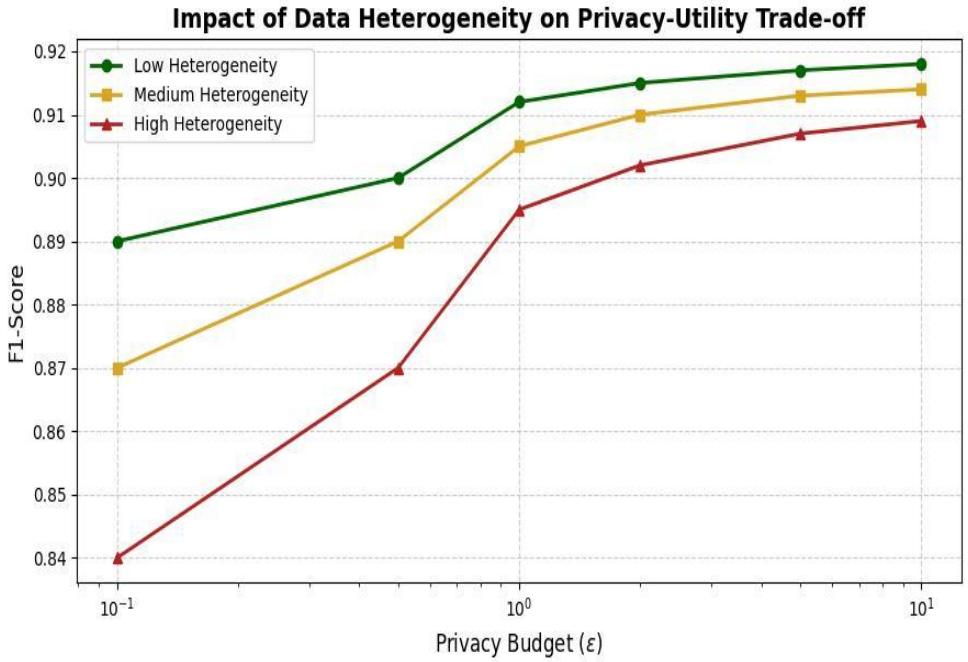

Figure 4: Impact of data heterogeneity on privacy-utility trade-off

4.6.3 Communication overhead versus number of participants

Figure 5 evaluates the scalability of HybridFL-APT by plotting the total communication overhead against an increasing number of participants (from 5 to 50). The communication cost of our hybrid architecture grows sub-linearly, which can be attributed to its selective update mechanism and gradient sparsification. In contrast, the overhead for standard FedAvg increases nearly linearly, making it less practical for large-scale deployments.

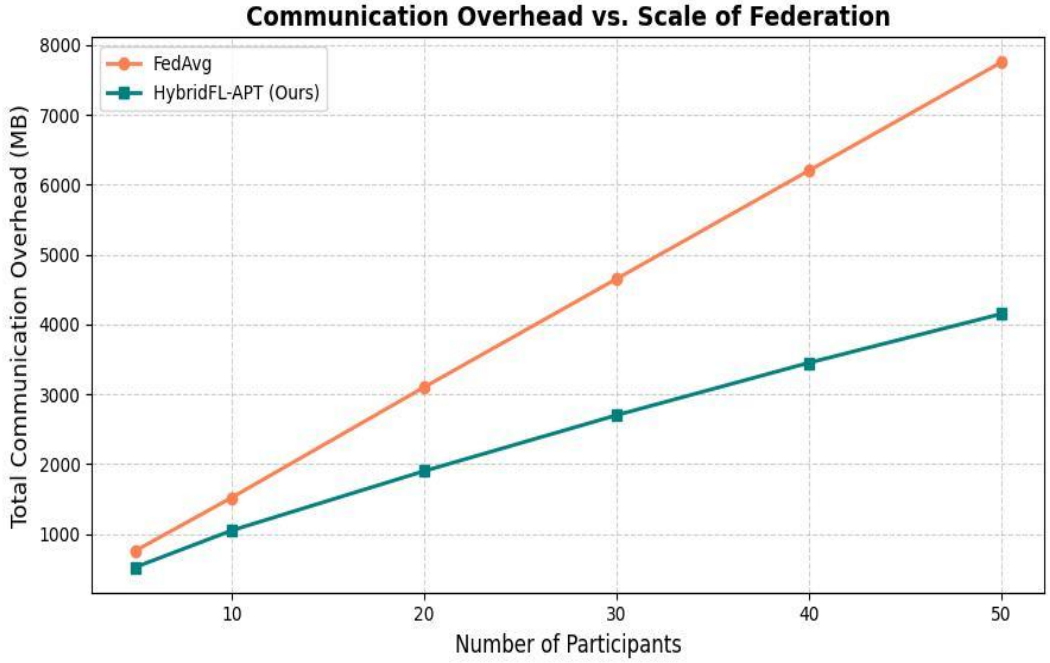

Figure 5: Communication overhead vs. federation scale

4.6.4 Ablation study: contribution of framework components

An ablation study was conducted to quantify the contribution of each major component in HybridFL-APT. The results are shown in Figure 6. Using a horizontal-only federated model as the baseline, the addition of vertical feature fusion (VFF) enhances the F1-score by 2.7%.. Integrating the adaptive privacy module (APM) leads to a minor 1.5% reduction in performance, while providing strong formal privacy guarantees. The complete model, which incorporates all components including the

temporal-graph neural network, delivers the highest overall performance.

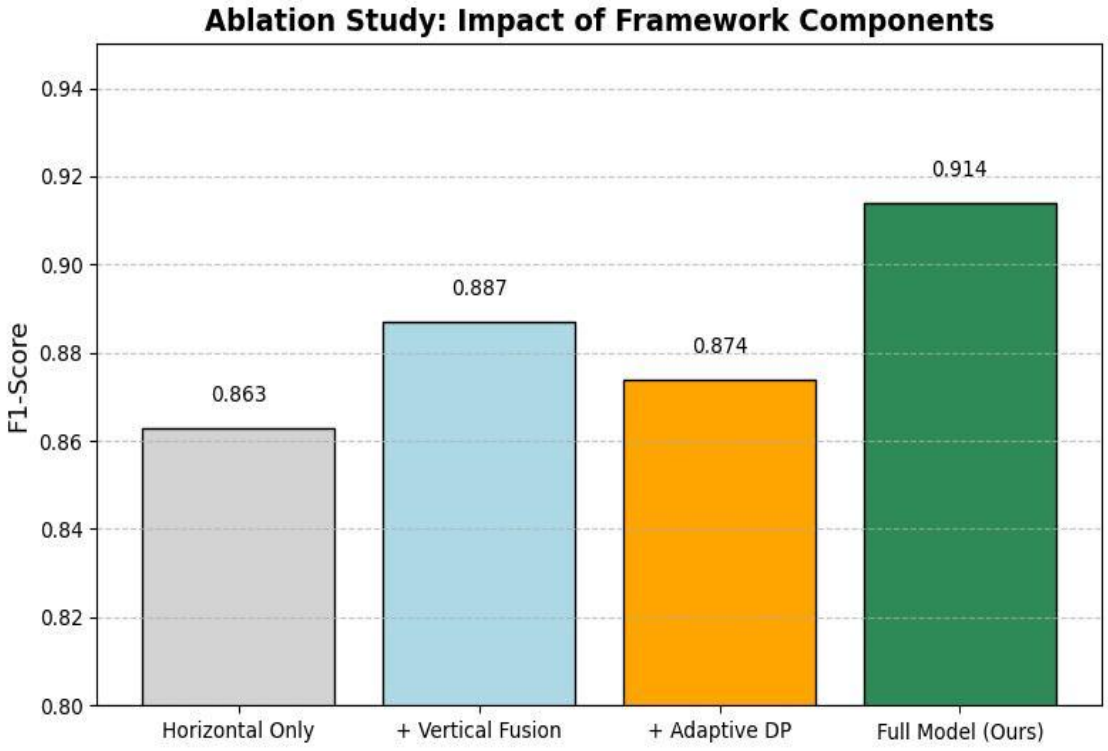

Figure 6: Ablation study: contribution of framework components

4.7 Experimental conclusion

Based on the above experimental results and analysis, the following conclusions can be drawn. First, in terms of effectiveness, HybridFL-APT achieves APT attack chain detection accuracy comparable to that of centralized methods while strictly protecting data privacy, and is significantly superior to existing federated learning methods. Second, in terms of efficiency, by adopting a hybrid federated architecture and communication optimization techniques, the framework shows obvious advantages in both communication overhead and convergence speed, and has the feasibility of practical deployment. Finally, in terms of robustness, the adaptive privacy mechanism achieves a better privacy-utility trade-off, and the framework has a certain tolerance for low-quality or malicious data. The above experimental results comprehensively verify the advancement and practicality of the framework proposed in this paper in solving the cross-domain collaborative APT detection problem.

5 Discussion

The results presented in Section 4 demonstrate that HybridFL-APT offers a substantial improvement in cross-domain APT detection. To better contextualize these findings, we discuss our framework’s performance and architecture in relation to the current state of the art.

5.1 Comparison with horizontal-only frameworks

Most federated learning (FL) applications in cybersecurity, such as FedAttackChain [10] , rely on horizontal collaboration to expand training datasets. While this increases the volume of samples, it often fails to account for the heterogeneous nature of APT traces, which are typically scattered across different network layers and endpoint logs. Our results show a 2.9%–5.1% improvement in F1-score over these baselines, which we attribute to our Vertical Multi-source Feature Fusion layer. By aligning diverse feature spaces, HybridFL-APT identifies cross-domain correlations that horizontal-only models overlook, effectively bridging "data silos" at the feature level.

5.2 Adaptive privacy and model stability

A primary challenge in privacy-preserving FL [16] [18] is the "utility cliff"—a sharp drop in accuracy that occurs when fixed noise injection is used to meet strict privacy budgets. Our sensitivity analysis in Table 3 shows that HybridFL-APT avoids this drop through an Adaptive Differential Privacy (ADP) mechanism. By dynamically adjusting the privacy budget based on training dynamics, the model maintains stability and convergence even when processing noisy, cross-organizational data. This approach to maintaining model "flatness" draws on established principles from adaptive control theory [23] [24] [24] [25] [26] [27] [28] , ensuring that privacy does not come at the expense of detection reliability.

5.3 Advancements in attack chain modeling

While previous works [2] [4] rely heavily on sequence modeling, our framework introduces a Temporal-Graph Neural Network (T-GNN) to capture both temporal sequences and structural entity relationships simultaneously. As illustrated in Figure 3, this approach is particularly effective during the Command & Control and Exfiltration stages, where attacker behavior is most likely to blend in with legitimate administrative traffic. By providing a more comprehensive reconstruction of the attack chain, this method allows Security Operations Centers (SOCs) to move beyond simple alert-matching and implement more effective, context-aware response strategies.

6 Conclusion

Aiming at the cross-domain stealth characteristics of APT attacks, this paper proposes a hybrid federated learning detection framework. Through the collaborative mechanism of horizontal and vertical federation, this framework realizes the multi-source threat intelligence fusion under the premise of privacy protection. Combining adaptive differential privacy and temporal-graph neural network, while guaranteeing data security, it effectively improves the detection accuracy and interpretability of the attack chain.

Experimental results show that the proposed framework is close to the centralized baseline in detection performance, significantly outperforms existing federated learning methods, and performs excellently in terms of communication efficiency and robustness.

This study is currently mainly based on simulation data. Future work will further explore the deployment and optimization of this framework in real industrial environments, and study the combination of unsupervised learning and trusted computing technologies to enhance the ability to detect unknown threats and the auditability of collaborative processes.

References

[1] SiJung Kim, DoEun Cho, SangSoo Yeo. Secure model against APT in m-connected SCADA network. International Journal of Distributed Sensor Networks, 2014, 10. DOI: 10.1155/2014/594652

[2] Yazhou Du, Weiwu Ren, Wenjuan Li, et al. GA-ConvE: An APT attack prediction method based on combination of graph attention network and 2D convolution. Neural Networks, 2025: 108216. DOI: 10.1016/j.neunet.2025.108216

[3] Rabia Khan, Noshina Tariq, Muhammad Ashraf, et al. FL-DSFA: Securing RPL-based IoT networks against selective forwarding attacks using federated learning. Sensors, 2024, 24(17). DOI: 10.3390/s24175834

[4] Wudu Bitew Alemayew, Ketema Adere Gemeda. Federated hybrid deep learning for multi-attack detection and classification in RPL-based 6LoWPAN networks. Discover Computing, 2025, 28(1): 1-35. DOI: 10.1007/s10791-025-09852-3

[5] Rashmi Verma, Manisha Jailia. A hybrid metaheuristic federated learning approach-based attack detection system for multi-cloud environment. Journal of Cloud Computing, 2025, 14(1): 1-22. DOI: 10.1186/s13677-025-00797-y

[6] Alessio Sacco, Doriana Monaco, Guido Marchetto, Paolo Montuschi. Dealing with challenged IoT networks in hierarchical federated learning. IEEE Internet of Things Journal, 2025: 12(18):36979-36992. DOI: 10.1109/JIOT.2025.3580627

[7] Xiaoming Wang, Zhiquan Liu, Mingzhen Dai, et al. A verifiable and efficient chained federated learning scheme for privacy protection. Computer Networks, 2025: 111838. DOI: 10.1016/j.comnet.2025.111838

[8] Xiang Chen, Dun Zhang, Zhanqi Cui, et al. DP-Share: Privacy-preserving software defect prediction model sharing through differential privacy. Journal of Computer Science and Technology, 2019, 34(5): 1020-1038. DOI: 10.1007/s11390-019-1958-0

[9] Lixin Cui, Xu Wu. ALDP-FL for adaptive local differential privacy in federated learning. Scientific Reports, 2025, 15(1): 1-18. DOI: 10.1038/s41598-025-12575-6

[10] Debao Wang, Shaopeng Guan. FedFR-ADP: Adaptive differential privacy with feedback regulation for robust model performance in federated learning. Information Fusion, 2025, 116: 102796. DOI: 10.1016/j.inffus.2024.102796

[11] G Hemanth Kumar, Sivananda Lahari Reddy Elicherla, Sugandha Saxena, et al. FL-DPCSA: Federated learning with differential privacy for cache side-channel attack detection in edge-based smart grids. E-PRIME: Advances in Electrical Engineering, 2025. DOI: 10.1016/j.prime.2025.101057

[12] George Alter, Brett Hemenway Falk, Steve Lu, et al. Computing statistics from private data. Data Science Journal, 2018, 17. DOI: 10.5334/dsj-2018-031

[13] Tayyab Rehman, Noshina Tariq, Farrukh Aslam Khan, et al. FFL-IDS: A fog-enabled federated learning-based intrusion detection system to counter jamming and spoofing attacks for the Industrial internet of things. Sensors, 2024, 25(1). DOI: 10.3390/s25010010

[14] Rong Xie, Zhong Chen, Weiguo Cao, et al. Federated self-expanding neural network learning framework for heterogeneous devices. Expert Systems with Applications, 2026: 131199. DOI: 10.1016/j.eswa.2026.131199

[15] Madhuri Gadwal, Atul Negi. A lean discriminative nonlinear federated dictionary learning method. Engineering Applications of Artificial Intelligence, 2026, 167: 113862. DOI: 10.1016/j.engappai.2026.113862

[16] Haoyu Jiang, Xiaoliang Chen, Duoqian Miao, et al. PrivTSAD-FedWGAN: A novel federated learning and WGAN framework for privacy-preserving multivariate time series anomaly detection. Expert Syste

ms with Applications, 2026: 131049. DOI: 10.1016/j.eswa.2025.131049

[17] Jiayuan Chen, Tiantian Zhu, Zhengqiu Weng, et al.SLATSCOG: A secure authentication framework via federated data generation and temporally-enhanced split learning. Knowledge-Based Systems, 2026: 115304. DOI: 10.1016/j.knosys.2026.115304

[18] Jie Niu, Runqi He, Qiyao Zhou, et al. Adaptive differential privacy Cox-MLP model based on federated learning. Mathematics, 2025, 13(7). DOI: 10.3390/math13071096

[19] Sanxiu Jiao, Lecai Cai, Xinjie Wang, et al. A differential privacy federated learning scheme based on adaptive gaussian noise. CMES-Computer Modeling in Engineering & Sciences, 2024, 138(2): 1679-1694. DOI: 10.32604/cmes.2023.030512

[20] Yanjin Cheng, Wenmin Li, Sujuan Qin, et al. Differential privacy federated learning based on adaptive adjustment. CMC-Computers Materials & Continua, 2025, 82(3): 4777-4795. DOI: 10.32604/cmc.2025.060380

[21] Jing Rong, Qiuzhan Zhou, Huinan Wu. A hybrid federated learning framework for enhancing privacy and bobustness in non-intrusive load monitoring. Sensors, 2026. DOI: 10.3390/s26020443

[22] Amjad Rehman, Kamran Ahmad Awan, Amal Al-Rasheed, et al. A novel hybrid fuzzy logic and federated learning framework for enhancing cybersecurity and fraud detection in IoT-enabled metaverse transactions. Egyptian informatics Journal, 2025, 30. DOI: 10.1016/j.eij.2025.100668

[23] Boulkroune, A., Boubellouta, A., Bouzeriba, A. et al. Practical Finite-Time Fuzzy Synchronization of Chaotic Systems with Non-Integer Orders: Two Chattering-Free Approaches. J. Syst. Sci. Syst. Eng. 34, 334–359 (2025). DOI: 10.1007/s11518-024-5635-7

[24] Rigatos, G., Abbaszadeh, M., Busawon, K., Dala, L., Pomares, J., and Zouari, F. (December 6, 2023). "Flatness-Based Control in Successive Loops for Autonomous Quadrotors." ASME. J. Dyn. Sys., Meas., Control. March 2024; 146(2): 024501. DOI: 10.1115/1.4063907

[25] Rigatos, G., Siano, P., Zouari, F. et al. Nonlinear optimal control of autonomous submarines′ diving. Mar Syst Ocean Technol 15, 57–69 (2020). DOI: 10.1007/s40868-019-00070-3

[26] Rigatos, G. et al. Flatness-based control in successive loops for dual-arm robotic manipulators. 2024 IEEE Conference on Control Technology and Applications (CCTA). IEEE, (2024). DOI: 10.1109/CCTA60707.2024.10666567

[27] G. Rigatos, P. Siano, F. Zouari and S. Ademi, "A nonlinear optimal control methoc for autonomous submarines' diving," 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 2017, pp. 1061-1066, DOI: 10.1109/ISIE.2017.8001393.

[28] Zouari, F., Mahmud, M. (2026). Neural Network-Based Robust Adaptive Output Feedback Control for MIMO Time-Varying Delay Systems. In: Mahmud, M., Pillay, N., Kaiser, M.S. (eds) Applications of Artificial Intelligence and Data Science. AAIDS 2024. Communications in Computer and Information Science, vol 2601. Springer, Cham. DOI: 10.1007/978-3-031-98498-3_5