International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 206-217

https://doi.org/10.70517/ijhsa46416

Privacy Protection and Collaborative Defense System for Software Development Security Detection in Power Industry under Artificial Intelligence Federated Learning Framework

Lina Chen[1] and Heng Xia[1,*]
[1] Digital Operation Center of Guangxi Power Grid Co., LTD., Nanning, Guangxi, 530000, China
Corresponding authors: (e-mail: 18767164710@139.com).

Abstract With the accelerated digital transformation of power systems, traditional protection measures can no longer cope with complex attack methods. In this paper, for the problem of data privacy protection and security defense in software development in the power industry, a privacy protection and collaborative defense system based on artificial intelligence federated learning framework is proposed. The study adopts differential privacy technology to protect client data privacy, designs a differential privacy federation learning method based on knowledge distillation, and constructs a collaborative DNS defense system based on blockchain technology. The experimental results show that the proposed method achieves 86.7% and 95.9% security detection accuracy on MNIST and Fashion Mnist datasets, respectively, which is 4.6 and 4.5 percentage points higher compared to the FedMatch method; in terms of the accuracy rate of different types of samples, the accuracy rate of the attack event, natural event, and no-event type reaches 82%, 81%, and 95%, an improvement of 5, 8 and 5 percentage points over the FedMatch method, respectively; and significantly outperforms the traditional differential privacy mechanism in terms of model convergence speed. The study provides a new idea for data security and efficient circulation in software development in the power industry, which can effectively deal with cyber security threats and guarantee the stable operation of the power system.

Index Terms federated learning, differential privacy, knowledge distillation, collaborative defense, power software security, blockchain

I. Introduction

In recent years, the cybersecurity situation has become increasingly severe, seriously threatening the safe and stable operation of power systems. As a kind of data closely related to national security and economic development, power grid data has stricter restrictions on its use to ensure that events threatening national security, such as information leakage, do not occur [1]-[3]. Ensuring the data security of the new power system is an important part of power system network security protection [4].

Due to the power system network security with dynamic, confrontational, chain characteristics, its network security protection is also different from the open interconnected network or management system, the network attack sample size is small, once the impact of the damage is huge, the attack chain is long [5]-[8]. At the same time, the network attack on the power system means hidden, so that the defense system for the power system protection surface is large, many risk points, universal protection technology is difficult to adapt [9], [10]. At present, the power system network security protection to isolate, detect, check and kill the main protection measures rely on security equipment passive protection [11], [12]. In the face of the number of tens of thousands and rapid growth of border threats, on the one hand, the security equipment to receive a large amount of information related to network security and protection equipment information is scattered, the resources are not effectively integrated [13]. On the other hand, the rapid expansion of the malicious attack feature library, a large number of unknown alarms can not be matched, resulting in inefficient protection and high consumption of system resources [14], [15]. This indicates that the existing security protection system has been unable to adapt to the constantly evolving network technology, and there are various problems such as protection technology lagging behind the means of attack, security functions constraining the business functions, and protection measures affecting the control of business real-time and so on [16]-[18]. For this reason, the application of artificial intelligence technology in the security detection scenario of the power grid can provide some references for the subsequent research and development of privacy computing as well as the implementation of the digitalization of the power grid, so as to provide a safe and efficient flow of data in the process of related electric power software development.

In recent years, the cybersecurity situation has become increasingly severe, seriously threatening the safe and stable operation of power systems. As a kind of data that is closely related to national security and economic development, power grid data has stricter restrictions on its use to ensure that events that threaten national security, such as information leakage, do not occur. Ensuring the data security of the new power system is an important part of power system network security protection. Due to the power system network security with dynamic, confrontational, chain characteristics, its network security protection is also different from the open interconnected Internet or management system, network attacks on the sample size is small, once the impact of destruction is huge, the attack chain is long. At the same time, the power system for the network attack means hidden, so that the defense system for the power system to protect the surface, the risk of multiple points, general protection technology is difficult to adapt. At present, the power system network security protection to isolate, detect, check and kill the main protection measures rely on security equipment passive protection. In the face of the number of tens of thousands and rapid growth of border threats, on the one hand, the security equipment to receive network security-related information and protection equipment information is scattered, the resources have not been effectively integrated. On the other hand, the rapid expansion of the malicious attack feature library, a large number of unknown alarms can not be matched, resulting in low protection efficiency, the system resource consumption. This shows that the existing security protection system has been unable to adapt to the constantly developing network technology, there are protection technology lagging behind the means of attack, security features constraints on the business functions, protection measures affect the control of real-time business and other aspects of the problem. For this reason, the application of artificial intelligence technology to the security detection scenario of the power grid can provide some references for the subsequent research and development of privacy computing as well as the implementation of the digitization of the power grid, so as to promote the safe and efficient flow of data in the process of the development of related electric power software.

This study first analyzes the characteristics of Internet-based power group intellectualization software development, and explores the trusted mechanism of blockchain technology in power software development. Aiming at the problem of power data privacy protection, a differential privacy federation learning method based on knowledge distillation is proposed, which introduces unlabeled public datasets and considers the difference in data volume between clients, and designs a dedicated privacy protection scheme for power software development scenarios. Meanwhile, a DNS collaborative defense system is constructed, which contains three layers of defense strategies, namely, edge resolution, traffic cleaning, and service mirroring, and a federation chain model is established by using the super ledger to realize the trust mechanism and collaborative defense among all related organizations. Through experiments on MNIST, Fashion Mnist and REDD datasets, the proposed method is verified to be superior in terms of security detection accuracy, model stability and communication efficiency, providing a new technical path and solution for software development security detection in the power industry.

II. Differential Privacy-based Privacy Protection for Power Software Development

II. A.Internet-based power group intellectualization software development

As the Internet continues to evolve and software development technologies become increasingly innovative, software scale and complexity grow and rise to a whole new level. Today's software has evolved from a closed, static change to a new paradigm of continuous change in demand, continuous evolution of the system and continuous growth in scale. Traditional software development methods such as automated software development methods and engineering software development methods have been unable to effectively solve the many intricate problems in the various stages of software development in the Internet environment under the new software situation, and the traditional software development methods are facing great challenges. Internet-based group intelligence software development utilizes the advantages of the Internet in data processing and dissemination rate, effectively pooling the capabilities of developers around the world to improve the quality and efficiency of software development, later referred to as “group intelligence software development”.

The complete process of Crowd Intelligence software development usually involves software requirement publishers, developers and Internet-based Crowd Intelligence software development platforms, and individually may involve third-party payment institutions, such as banks, Alipay or WeChat.

(1) Demand publisher: It can be an individual or a company that organizes its own demand for software and then publishes it on the Group Intelligence software development platform, which can effectively shorten the cycle, reduce the cost and improve the efficiency.

(2) Internet-based Group Intelligence software development platform: responsible for the management of the entire process of software development, including user registration, user release of software development requirements, matching with the appropriate developers, receiving and sending solutions, as well as the collection

and payment of honorariums and the provision of communication platforms, etc., which is a key link in the development of Group Intelligence software.

(3) Developer: It can be an individual developer or a company or a team, who chooses the software development tasks suitable for him/her according to his/her skills and interests from the Group Intelligence software development platform.

II. B.Blockchain-based Trusted Mechanism for Power Software Development

In this paper, considering the transaction process under group intellectualization software development and the related process of software development, we propose a blockchain-based security mechanism for group intellectualization software development by taking advantage of the blockchain's features of decentralization, tampering, and traceability, and by combining the trust problems encountered in the software development process. The software development process is divided into steps to release important information to the blockchain, weakening the role of the trust center of the platform in the traditional group intellectualization system.

In traditional group intellectualization software development, developers on the Internet have to take the platform as the center, and data and privacy are stored in this centralized server. In this paper, the introduction of blockchain technology, all nodes on the Internet as equal nodes, no longer need to take the platform as the center of trust, each node joins the blockchain network to save a complete copy of the blockchain, with the help of the consensus mechanism, the nodes reach a consensus among themselves.

The system architecture for users to participate in the development of a specific task is divided into multiple processes, including user registration, publishing requirements, receiving tasks, submitting solutions, receiving and accepting solutions, payment of honorarium, program arbitration, service compensation and task rollback. All processes are implemented and deployed on the blockchain through the form of Ethernet smart contracts, which are triggered as a piece of program code and run automatically by nodes on the blockchain, generating transaction behaviors between nodes.

II. C.Differential Privacy Preservation Based on Knowledge Distillation in Federated Learning

Differential privacy techniques, as a privacy preserving method [19], [20], have been widely applied in the field of federated learning. Existing studies on differential privacy applied to federated learning either do not consider the existence of unlabeled public data or the difference in the amount of data between clients, so they limit its application in real-world scenarios. A differential privacy federated learning approach based on knowledge distillation is proposed, which introduces unlabeled public datasets and takes into account the difference in data volume between clients, and a dedicated privacy protection scheme is designed for this scenario.

Differential privacy technique is adopted to ensure the data privacy of clients, which is divided into two classes according to the degree of privacy of client data, and unequal differential privacy budgets are assigned to clients of different classes in the federation training phase to realize the hierarchical allocation of privacy budgets. In addition, to limit the total consumption of privacy budget, the privacy budget in the federation training phase is set as a fixed value, and the privacy budget in the pseudo-label adding phase is adjusted according to the client's need for privacy and the nature of the parallel combination of privacy budgets. The scheme is extensible, and the highly flexible privacy budget allocation enables it to meet complex privacy requirements.

II. C. 1) Algorithm overview

It is assumed that all clients involved in training are trustworthy and do not attack other clients. The server is honest and curious and may steal the privacy of the clients, in view of which differential privacy techniques are introduced to secure the data of each client.

Public data with unknown labels is known to exist, as opposed to private data with known labels that is stored locally at each client. The purpose of each client is to jointly train a global model and guarantee the privacy of the local data in the process. There are large differences in the amount of data between the clients, and based on the differences in the amount of data, the clients are divided into two categories: large-data-volume clients $C^{'} = \{c_1^{'}, c_2^{'}, c_3^{'} ... c_m^{'}\}$, general client $C^{s} = \{c_1^{s}, c_2^{s}, c_3^{s} .... c_m^{s}\}$. The algorithmic framework is shown in Fig. 1.

The proposed knowledge distillation-based federation learning method for differential privacy is divided into two phases: the first phase is the pseudo-labeling of the public dataset; the second phase is the federation training. In the first phase, through knowledge distillation, the teacher model obtained from the training of the large-data-volume client is used to add pseudo-labels to the public unlabeled data, so as to obtain the labeled public dataset, and to establish a new client, i.e., the “special client”. In the second phase, the general client and the special client are trained together in the federation.

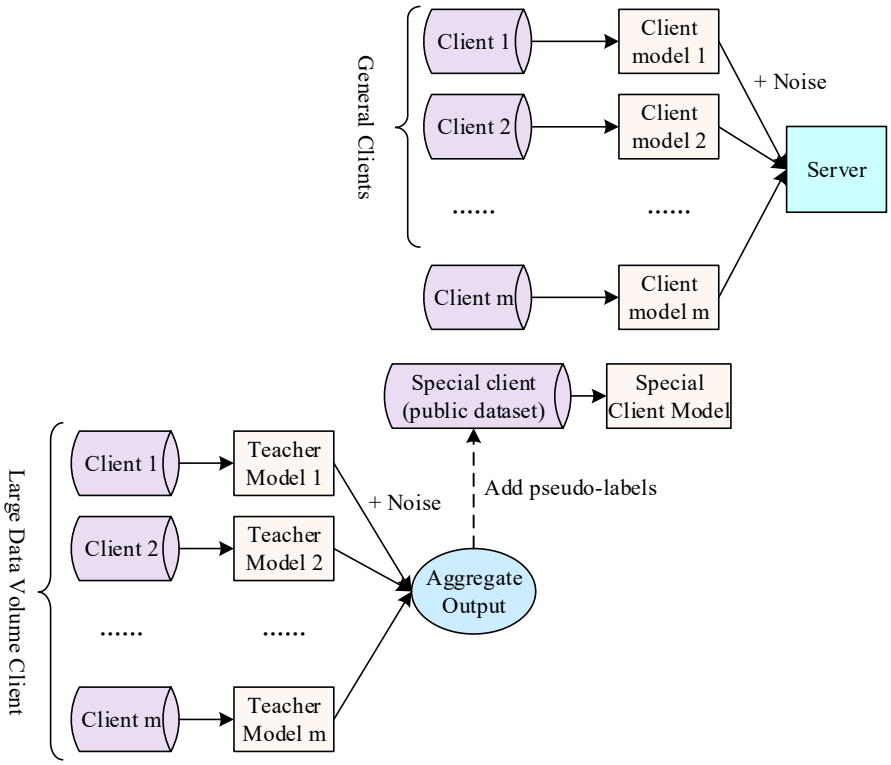

Figure 1: Algorithm structure

Differential privacy techniques are applied in both phases to protect the data privacy of the clients. In federated learning, different differential privacy budgets are assigned to different classes of clients according to the differences in data privacy, realizing the hierarchical allocation of privacy budgets. The privacy budget allocation in each phase of the algorithm is adjusted according to the clients' need for privacy and the parallel combinatorial nature of the privacy budget.

II. C. 2) Algorithm design

(1) Pseudo-labeling

First, the clients are grouped into “big data clients” and “general clients” according to the size of data, and the teacher models trained on the big data client cluster are used to add labels to the unlabeled public dataset.

Definition 1 (Label Counting): Assuming that the number of labeled categories is k, each teacher model T_i outputs its prediction $f_i(\tilde{x})$ of the labels of the public data $\tilde{x}$. For a given category $j \in [k]$ and data $\tilde{x}$, the number of teachers who think that the label of $\tilde{x}$ is j is denoted as $n_j(\tilde{x})$, with the expression shown in equation (1):

$$n_j(\tilde{x}) = \left| \{ i : i \in [m], f_i(\tilde{x}) = j \} \right| \tag{1}$$

where $\tilde{x}$ is the public data and $f_i(\cdot)$ is the prediction result.

Definition 2 (Pseudo-labeling): the category with the highest count of $n_j(\tilde{x})$ is used as the pseudo-labeling of this data, and noise is introduced in consideration of data privacy security. Thus, the pseudo-label expression for the data is shown in equation (2):

$$f(x) = \arg\max_j \left\{ n_j(\tilde{x}) + noise \right\} \tag{2}$$

(2) Federated Learning

The federated learning process is shown in Fig. 2, where the global model is first downloaded by each client and trained on local data, and then the local model obtained after training is uploaded to the central server, which is responsible for aggregating the model and sending the global model in the next communication round. In the algorithm design of this paper, the algorithm used by the central server to aggregate the models is the federated average algorithm that is widely used in federated learning [21].

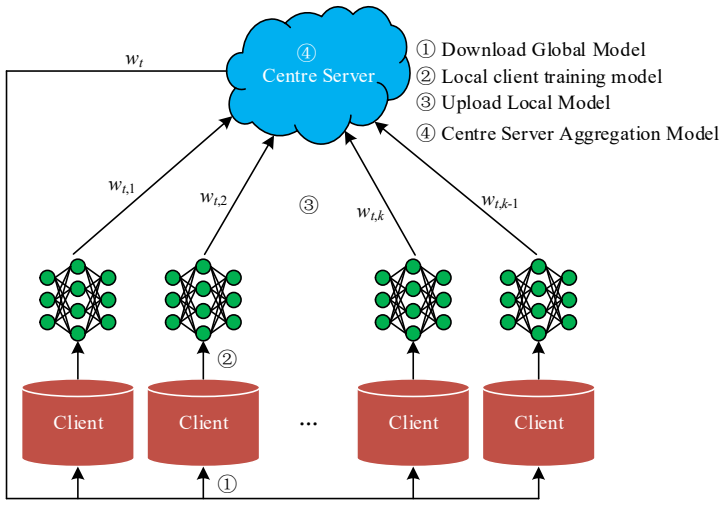

Figure 2: Federal learning process

Definition 3 (Federated Averaging Algorithm): the central server initializes the model parameters, training is executed for a number of rounds, and at least 1 to as many as K clients are selected to participate in the training in each round, and each selected client k simultaneously trains locally with local data to obtain the local model w_t based on the model $w_{t+1,k}$ for the current round issued by the server, and upload $w_{t+1,k}$ back to the server. The server aggregates the collected models from each client based on the number of samples from each party using weighted average to get the next round of models w_{t+1} as in equation (3):

$$w_{i+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{i+1,k} \tag{3}$$

where w is the model parameter and n is the number of samples.

In federation training, in order to avoid data overlapping, this paper does not simply add the public dataset to each client participating in federation training, but treats the public data as a special client c^{pnb} participating in federation training, i.e., the general client $C^s = \{c_1^s, c_2^s, c_3^s ....c_m^s\}$ and the special client c^{pnb} together for federated training.

In each round of training, some clients are randomly sampled for training. Considering that the data of the special client comes from the public dataset and is shared, this paper sets the special client to participate in each round of training. Differential privacy technique is introduced in federated learning, adding noise at each client model parameter to ensure the security of client data.

(3) Privacy budget allocation

In order to ensure the data privacy in the training process, this paper adopts the differential privacy technique to ensure the security of client data.

Definition 4 (Differential Privacy): for any two neighboring datasets D and D' and any subset S_φ of P_φ, an algorithm M is said to provide ε-differential privacy protection if the algorithm M satisfies the following equation (4), where the parameter ε is called privacy protection budget:

$$\Pr[M(D) \in S_M] \leq \exp(\varepsilon) \times \Pr[M(D') \in S_M] \tag{4}$$

where M is a randomized algorithm, P_M is the set consisting of all possible outputs of M, and D is the data set.

The allocation of the privacy budget follows the following properties:

Property 1 (Sequential Combination Property): algorithms $M_1, M_2, M_3 \ldots M_k$ satisfy $\varepsilon_1 - DP$, $\varepsilon_2 - DP$, $\varepsilon_3 - DP \ldots \varepsilon_k - DP$, respectively, for the same dataset D, the combinatorial algorithm $M(M_1(D), M_2(D), M_3(D) \ldots M_k(D))$ consisting of these algorithms satisfies $\left(\sum_{i=1}^{k} \varepsilon_i\right) - DP$.

Property 2 (Parallel Combinatorial Property): algorithms $M_1, M_2, M_3 \ldots M_k$ can be combined if they satisfy $\varepsilon_1 - DP$, $\varepsilon_2 - DP$, $\varepsilon_3 - DP \ldots \varepsilon_k - DP$, and datasets $D_1, D_2 \ldots D_k$ are disjoint, and the combined algorithm $M(M_1(D_1), M_2(D_2), M_3(D_3) \ldots M_k(D_k))$ consisting of these algorithms satisfies $\left(\max_{i \in \{1,\ldots,k\}} \varepsilon_i\right) - DP$.

III. DNS Cooperative Defense System

III. A. Overall architecture

When software suffers from security hijacking or saturated traffic attack, the key point of defense is usually not in itself, and the factors affecting the availability and integrity rate involve multiple organizations such as authoritative resolution service organizations, Internet service providers (ISPs), public recursive service organizations, and application service providers, etc. However, there is currently no supporting trust and synergy mechanism between the relevant organizations. Superledger is a permission chain that, on the basis of inheriting the core concepts such as decentralization of traditional blockchain technology, significantly optimizes the transaction performance for enterprise-level application scenarios, strengthens the member identity management mechanism, and adopts the channel architecture and backing strategy to provide data privacy protection mechanisms at different granularities. This section proposes a DNS collaborative defense system, and the overall architecture of the defense system is shown in Figure 3, in which related organizations form an alliance chain to form a joint defense against security hijacking and saturation traffic attacks by sharing information and resources.

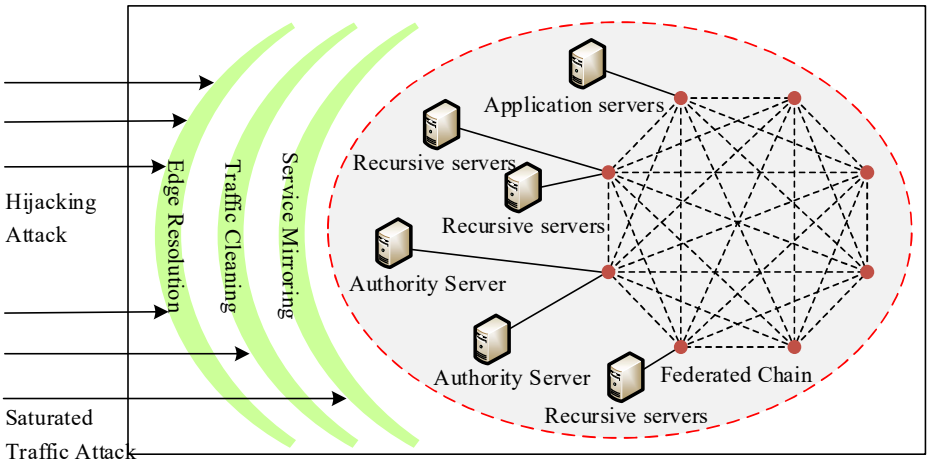

Figure 3: The overall architecture of the DNS cooperative defense system

DNS collaborative defense system mainly contains edge resolution, traffic cleaning, service mirroring 3-layer defense strategy, respectively, from different perspectives against saturated traffic attacks and security hijacking attacks to carry out defense.

III. B. Alliance Chain Model

The core of DNS collaborative defense system is its federation chain [22], as a security protection application scenario, the throughput and confirmation time of on-chain transactions, the confidentiality of transactions and data need to be taken into account.Fabric is designed as an application service platform, which has a higher transaction performance and a more complete privacy protection mechanism compared with other blockchain technologies, and it is suitable for domain name-related organizations to share information among themselves and to collaborative defense. This section establishes a federation chain model based on Fabric as shown in Figure 4.

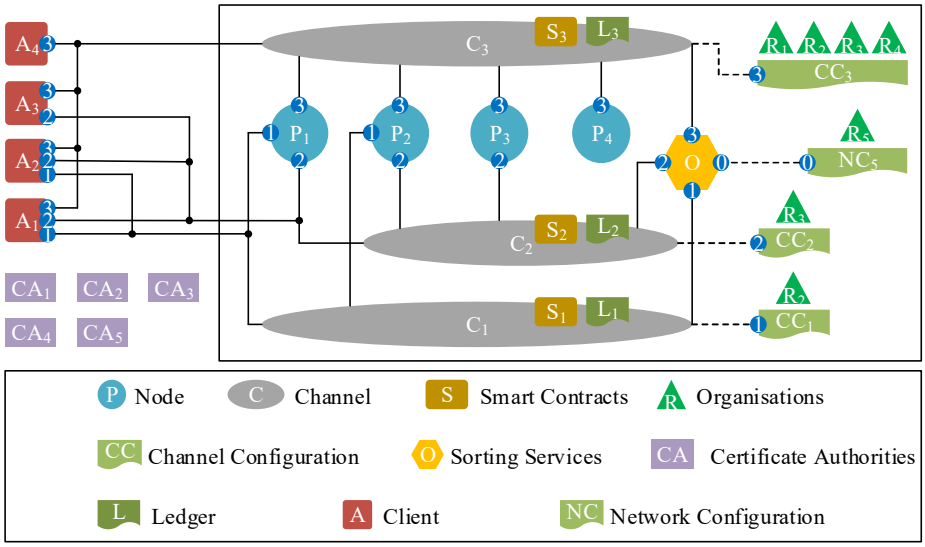

Figure 4: Alliance chain model schematic

Institutions in the Alliance Chain model contain ISP R1, Public Delivery Service R2, Authoritative Resolution Service R3, Application Service Provider R4, and MSP R5. R5 is elected by all the institutions in the Alliance Chain and is authorized to establish a network with members from multiple institutions.

The federation chain model inherits the characteristics of Fabric license chain, based on CA and MSP for member identity and authority management, and each organization has a CA for authentication and management of users. The model adopts multi-channel for business isolation to protect transaction and data privacy, and jointly protects data consistency and network efficiency through endorsement strategy and sequencing service. Currently Fabric supports three consensus algorithms, Solo, Raft and Kafka, of which the Solo consensus algorithm only supports a single sorting node, and the Kafka consensus algorithm has not yet achieved true decentralization. Therefore, MCC members establish sorting nodes individually and form O5 through the Raft consensus algorithm to provide sorting services for the alliance chain, achieving a balance between decentralization and transaction performance.

Fabric alliance chain can be composed of one to multiple alliances, and this section defines three types of alliance models around the collaborative defense system, namely, service mirroring, edge resolution, and traffic cleansing: 1) Service mirroring alliance is a one-to-one alliance composed of a recursive service organization and an Internet service provider, which can effectively safeguard the privacy of the related services. The recursive service organization is responsible for managing the alliance, and the Internet service provider is responsible for providing the mirroring service. The recursive service provider can establish mutually isolated service mirroring alliance with different ISPs to defend against direct saturation traffic attacks. 2) Edge resolution alliance is a one-to-many alliance consisting of an authoritative service provider, multiple public recursive service providers and ISPs, which protects the privacy of the data of the authoritative service provider. The authority service organization is responsible for managing the coalition, and the other organizations are responsible for edge resolution and defending against domain hijacking attacks and direct saturation traffic attacks against the authority. 3) Traffic cleaning coalition consists of all the organizations and is jointly managed by all the organizations, and the domain name service organization is responsible for traffic cleaning and defending against reflection amplification attacks. Since any organization may become the target of reflection amplification attacks, and all domain name service providers may become the springboard for reflection amplification, cleaning such attack traffic requires the cooperation of all organizations to achieve effective defense.

MCC R5 establishes service mirroring alliance (R1, R2), edge resolution alliance (R1, R2, R3) and traffic cleaning alliance (R1, R2, R3, R4) by modifying network configuration NC5. The service mirroring channel C1, the edge resolution channel C2 and the traffic cleaning channel C3 are created by the management organizations in the alliance, and the queuing nodes as well as the nodes and clients of the alliance members are added to the channels according to the rules specified in the channel configurations NC1, NC2, and NC3 to complete the establishment of the alliance chain network.

Institutions have their own nodes and clients, nodes are responsible for maintaining the ledger, executing smart contracts, providing endorsements and other affairs, and clients connect to the domain name servers or application servers of institutions, interact with the relevant servers for data, initiate proposals requesting collaborative defense, and control the servers to execute the collaborative defense proposals of other institutions in accordance with the

relevant strategies. Smart contracts and ledgers are conceptually subordinate to the channel and physically have copies in each node.

IV. Experimental results and analysis

IV. A. Security Detection Performance

In order to verify the security detection performance of the model constructed in this paper, the model in this paper is compared with other 3 methods on two datasets, MINIST and Fashion Mnist, and the comparison results are shown in Table 1. The other three comparison methods selected in this paper are supervised learning method (SL), FedAvg-FixMatch and FedMatch method.

As shown in Table 1, the model in this paper performs well in the 2 dataset detection accuracy, reaching 86.7% and 95.9% respectively, while the accuracy of the SL method is relatively low, 72.3% and 78.5% respectively. Relative to other methods, the accuracy of this paper's model on the MINIST dataset is improved by 14.4, 11.9, and 4.6 percentage points, respectively, and on the Fashion Mnist dataset by 17.4, 7.0, and 4.5 percentage points, respectively. This is because the FedAvg-FixMatch and FedMatch methods fail to fully utilize the local unlabeled data, and the SL method even uses only the labeled dataset for model training, failing to utilize the valuable information in the unlabeled data, which results in a relatively low model accuracy. The false alarm rate and omission rate of this paper's model in MINIST and Fashion Mnist datasets are low, which are 1.8%, 6.4% and 0.2%, 0.3%, respectively. This indicates that the model in this paper can reduce the cases of misidentifying attack samples as normal samples to a certain extent, and also can effectively avoid the cases of misidentifying normal samples as attack samples.

Table 1: Comparison of various methods (%)

Method	MINIST			Fashion Mnist		
	Accuracy	False rate	Leakage rate	Accuracy	False rate	Leakage rate
SL	72.3	6.2	14.5	78.5	5.8	8.9
FedAvg-FixMatch	74.8	4.3	11.6	88.9	2.2	2.7
FedMatch	82.1	3.8	8.8	91.4	1.7	1.5
Ours	86.7	1.8	6.4	95.9	0.2	0.3

Table 2: Evaluation results of various methods and categories

Method	Event type	Evaluation index		
		Accuracy/%	Recall/%	F1/%
SL	Attack event	74	48	59
	Natural event	68	71	70
	No event	70	94	84
FedAvg-FixMatch	Attack event	70	63	68
	Natural event	69	75	73
	No event	88	93	96
FedMatch	Attack event	77	72	75
	Natural event	73	77	75
	No event	90	93	94
Ours	Attack event	82	78	78
	Natural event	81	80	80
	No event	95	99	98

The evaluation results of different methods for different categories are shown in Table 2. The model in this paper also has improved precision and recall for different types of samples. Compared with the FedAvg-FixMatch method, the precision rate is improved by 12, 12, and 7 percentage points, and the recall rate is improved by 15, 5, and 6 percentage points, respectively. Compared with the FedMatch method, the precision rate is improved by 5, 8, and 5 percentage points, and the recall rate is improved by 6, 3, and 6 percentage points, respectively. The F1 value is more intuitive to reflect that the model in this paper is better. This is because the model constructed in this paper can make full use of local unlabeled data through pseudo-labeling and consistency regularization, and mitigate the impact on model performance. The model's aggregation method based on knowledge distillation also further

improves the model performance. The method in this paper shows better detection performance in the security threat detection problem.

IV. B. Accuracy analysis

Among the methods based on privacy in federated learning, differential privacy method is one of the more widely cited, which can protect the private data by adding noise information to the model parameter information that will be transmitted after the trained parametric model. Although the differential privacy method can protect the model parameter information of each participant to a certain extent with low communication overhead, this method can cause model errors and inaccuracies, so the model accuracy will be affected.

In the experimental design of this module, firstly, the model privacy protection using the traditional differential privacy method by adding Gaussian noise is used as a control group, and the accuracy experiment is compared with the method of this paper, and secondly, the error value of the objective function of the control group as well as the method of this paper is collected for each round in federated learning, which is selected as a measure of the error function MAPE. The experiment compares the convergence process of a participant in the training process under this paper's method and the traditional differential privacy mechanism, and the results are shown in Figure 5. From the experimental results, it can be seen that the differential privacy protection method based on knowledge distillation in this paper can reach a better convergence target faster, and the overall performance of the model is also more stable.

And in the process of this experiment power system node participants for experimental comparison, in the process of adding noise, the noise intensity is set to 1, 2, respectively, and the experimental results are tabulated as shown in Table 3 and Table 4. From the experimental results, the performance of this paper's method is more stable, and for different noise interference can maintain a better stability of prediction results. The experiments compare the results of this paper based on knowledge distillation of differential privacy protection and traditional differential privacy mechanism, from the noise intensity set to 1, 2 respectively, the comparison results can be seen, the method can ensure better accuracy and stability in the training task, the method of this paper and the difference between traditional differential privacy scheme to improve the accuracy of the comparison of 2%-5%, 5%-9%, respectively. This method does not interfere and process the model parameters transmitted by each participant, so it has certain advantages in terms of prediction accuracy and overall stability of the model, and can ensure the accuracy of the prediction of each distributed stage.

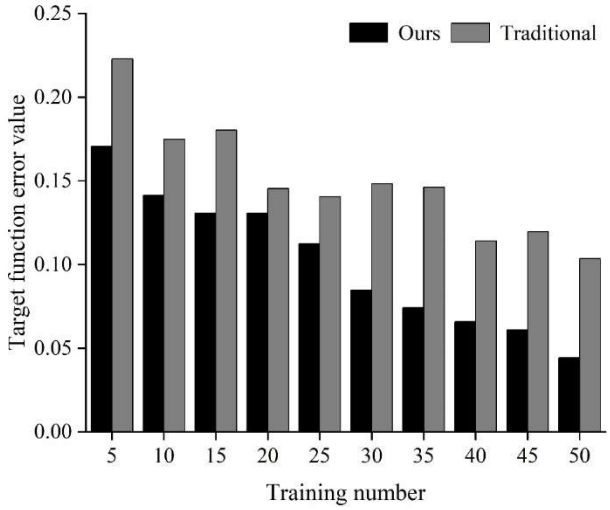

Figure 5: The process of convergence of our model and traditional differential privacy mechanism

Table 3: Accuracy comparison when the noise intensity is 1

MAPE	1	2	3	4	5
Traditional	9.84%	8.26%	7.54%	8.44%	5.88%
Ours	5.78%	5.89%	4.12%	5.34%	3.69%
MAPE	6	7	8	9	10
Standalone	9.59%	8.85%	8.65%	8.06%	9.06%
Ours	7.63%	5.83%	6.47%	5.02%	4.19%

Table 4: Accuracy comparison when the noise intensity is 2

MAPE	1	2	3	4	5
Traditional	11.06%	12.63%	11.79%	13.92%	9.63%
Ours	5.36%	5.49%	5.42%	5.27%	4.54%
MAPE	6	7	8	9	10
Standalone	12.34%	13.44%	15.43%	12.71%	12.96%
Ours	7.28%	4.55%	6.84%	5.42%	4.78%

IV. C. Communications overhead analysis

In order to analyze the communication efficiency of this paper's method, this experiment uses the communication optimized Fedsel (Federated SGD under Local Differential Privacy), CMFL (A Decentralized Framework with Committee Mechanism) this paper into the method for comparison and used FedAvg algorithm as a benchmark, experiments were conducted on the datasets MNIST, Fashion Mnist and REDD dataset respectively, the number of communication rounds was fixed and set to 100. the horizontal coordinate was set to the number of clients involved in the communication and the vertical coordinate was set to the communication overhead in terms of bytes, and the results of the experiments are shown in Fig. 6~Fig. 8.

From Fig. 6~Fig. 8, it can be seen that the communication overhead of each algorithm increases as more terminal data concentrators are involved in federated learning training, and the FedAvg algorithm does not take any communication optimization from the MNIST and Fashion Mnist datasets, as a baseline, the ware communication overhead is the largest. This is followed by the Fedsel algorithm, which is a two-stage local differential privacy framework that performs sole selection, and the communication overhead is roughly 73% of the FedAvg algorithm. This is followed by the CMFL algorithm, which screens irrelevant clients with coarseness and fineness for communication optimization, with a communication overhead of roughly 62% of the FedAvg algorithm, while the smallest overhead is the method proposed in this paper, with a communication overhead of roughly 47% of the FedAvg algorithm, and even less, with a communication overhead of only 25-45% of that of the FedAvg on the power dataset REDD.

The experimental results on different datasets show that the method proposed in this chapter has less communication overhead between the grid center and the terminal data concentrator during the federated learning training process, which saves more cost for the grid software with large and complex data volume, and at the same time speeds up the training time and achieves the final model results faster.

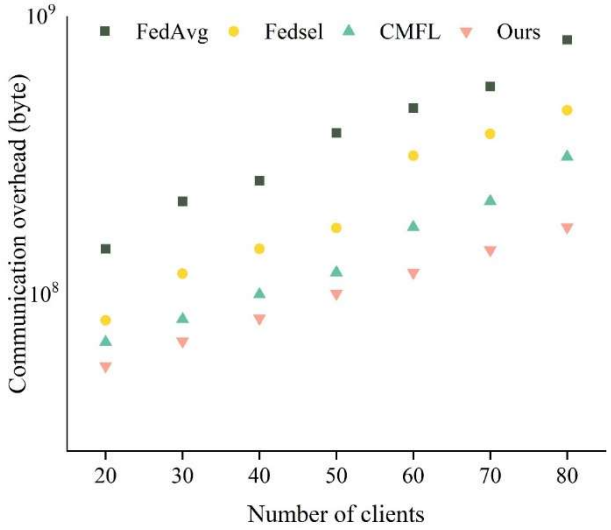

Figure 6: Communication overhead comparison in MNIST dataset

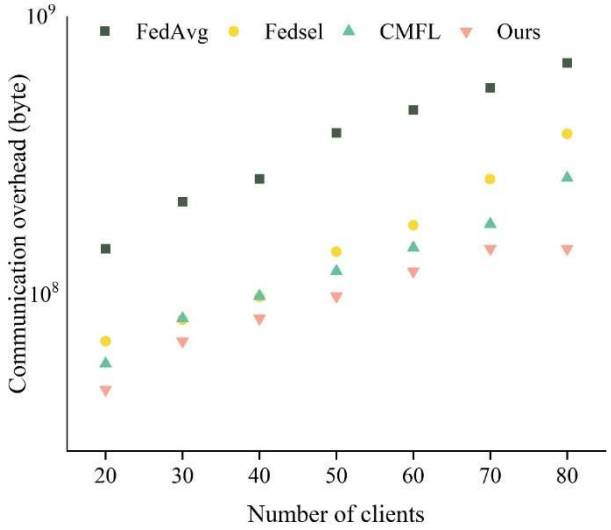

Figure 7: Communication overhead comparison in Fashion Mnist dataset

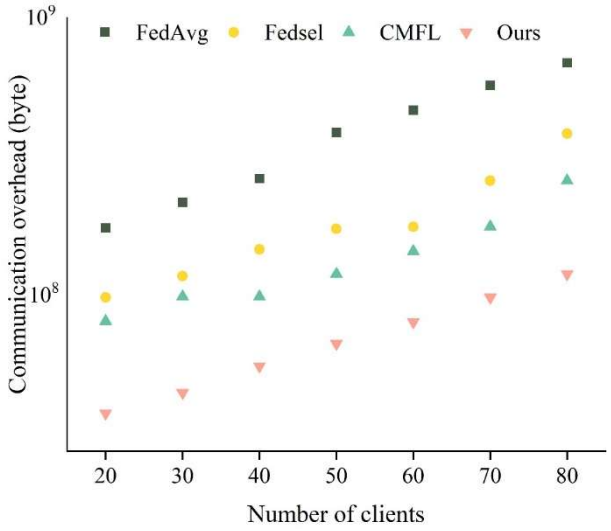

Figure 8: Communication overhead comparison in REDD dataset

V. Conclusion

In this paper, a privacy protection and collaborative defense system for security detection of software development in the power industry under the framework of artificial intelligence federated learning is proposed. The study constructs a differential privacy federation learning method based on knowledge distillation, and also designs a DNS collaborative defense system. The experimental results show that the proposed model has excellent security detection performance, and the false alarm rate and leakage rate on Fashion Mnist dataset are 0.2% and 0.3%, respectively, which are significantly lower than that of SL, FedAvg-FixMatch, and FedMatch methods; when the noise intensity is 2, the average value of the MAPE index of this method in the first five nodes is 5.22%, which is much lower than the 11.81% of the traditional differential privacy method; in terms of communication efficiency, the communication overhead of this method is only 47% of the FedAvg algorithm, which is particularly outstanding on the MNIST dataset. In addition, the super ledger-based federation chain model constructed in this study successfully realizes the trust and collaboration mechanism among multiple organizations, and effectively counteracts threats such as security hijacking and saturated traffic attacks through three types of federation models, namely, service mirroring federation, edge resolution federation, and traffic cleaning federation.

This research provides a new scheme for power software development that takes into account privacy protection and security defense, which is of great value for enhancing the network security protection capability of power systems and can effectively guarantee the safe and stable operation of power systems. Future research will further optimize the model structure and explore the application verification under more scenarios.

References

[1] Sun, C. C., Hahn, A., & Liu, C. C. (2018). Cyber security of a power grid: State-of-the-art. International Journal of Electrical Power & Energy Systems, 99, 45-56.

[2] Kimani, K., Oduol, V., & Langat, K. (2019). Cyber security challenges for IoT-based smart grid networks. International journal of critical infrastructure protection, 25, 36-49.

[3] Gunduz, M. Z., & Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. Computer networks, 169, 107094.

[4] Das, A. K., & Zeadally, S. (2019). Data security in the smart grid environment. In Pathways to a smarter power system (pp. 371-395). Academic Press.

[5] Liang, G., Weller, S. R., Luo, F., Zhao, J., & Dong, Z. Y. (2018). Distributed blockchain-based data protection framework for modern power systems against cyber attacks. IEEE Transactions on Smart Grid, 10(3), 3162-3173.

[6] Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. Energy Reports, 7, 8176-8186.

[7] Pan, K., Palensky, P., & Esfahani, P. M. (2019). From static to dynamic anomaly detection with application to power system cyber security. IEEE Transactions on Power Systems, 35(2), 1584-1596.

[8] Li, X., & Hedman, K. W. (2019). Enhancing power system cyber-security with systematic two-stage detection strategy. IEEE Transactions on Power Systems, 35(2), 1549-1561.

[9] Wei, M. (2017). Research on the construction of network security attack and defense range system in power monitoring system. In Information Technology and Intelligent Transportation Systems (pp. 82-90). IOS Press.

[10] Chatterjee, K., Padmini, V., & Khaparde, S. A. (2017, July). Review of cyber attacks on power system operations. In 2017 IEEE Region 10 Symposium (TENSYMP) (pp. 1-6). IEEE.

[11] Sridhar, S., & Govindarasu, M. (2017). Model-based attack detection and mitigation for automatic generation control. IEEE Transactions on Smart Grid, 5(2), 580-591.

[12] Telukunta, V., Pradhan, J., Agrawal, A., Singh, M., & Srivani, S. G. (2017). Protection challenges under bulk penetration of renewable energy resources in power systems: A review. CSEE journal of power and energy systems, 3(4), 365-379.

[13] Acosta, M. R. C., Ahmed, S., Garcia, C. E., & Koo, I. (2020). Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks. IEEE access, 8, 19921-19933.

[14] Liu, Z., Wei, W., & Wang, L. (2021). An extreme value theory-based catastrophe bond design for cyber risk management of power systems. IEEE Transactions on Smart Grid, 13(2), 1516-1528.

[15] He, Y., Mendis, G. J., & Wei, J. (2017). Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. IEEE Transactions on Smart Grid, 8(5), 2505-2516.

[16] Wen, M., Xie, R., Lu, K., Wang, L., & Zhang, K. (2021). FedDetect: A novel privacy-preserving federated learning framework for energy theft detection in smart grid. IEEE Internet of Things Journal, 9(8), 6069-6080.

[17] Jithish, J., Alangot, B., Mahalingam, N., & Yeo, K. S. (2023). Distributed anomaly detection in smart grids: a federated learning-based approach. IEEE Access, 11, 7157-7179.

[18] Ashraf, M. M., Waqas, M., Abbas, G., Baker, T., Abbas, Z. H., & Alasmary, H. (2022). Feddp: A privacy-protecting theft detection scheme in smart grids using federated learning. Energies, 15(17), 6241.

[19] Pengfei Zhang,Xiang Fang,Zhikun Zhang,Xianjin Fang,Yining Liu & Ji Zhang. (2025). Horizontal multi-party data publishing via discriminator regularization and adaptive noise under differential privacy. Information Fusion,120,103046-103046.

[20] Junyan Ouyang,Rui Han,Xiaojiang Zuo,Yunlai Cheng & Chi Harold Liu. (2025). Accuracy-aware differential privacy in federated learning of large transformer models. Journal of Information Security and Applications,89,103986-103986.

[21] DaoquGeng,ShouzhengWang & YihangZhang. (2024). Multi-Objective Federated Averaging Algorithm. Expert Systems,42(2),e13761-e13761.

[22] Han Yu,JianBin Li,Jimeng Song,Qingle Wang,Rongxin Lai,Yuancheng Li & Ziqi Shen. (2024). Secure and Efficient Multi-keyword Fuzzy Search Over Encrypted Data on Alliance Chain. Recent Advances in Electrical & Electronic Engineering,17(7),652-665.