Telecommunication Network Fraud Early Warning Based on Blockchain and Federated Learning

Haoliang Lan, Zhikang Xiang
Department of Computer Information and Network Security, Jiangsu Police Institute, Nanjing, China

Abstract: In response to the challenges of passive case analysis and disposal, insufficient utilization of data value mining, and low degree of collaborative sharing of intelligence and information technology in the crackdown and governance of telecommunication network fraud crimes, the application of blockchain combined with federated learning in telecommunication network fraud early warning aims to achieve automatic analysis and disposal of fraud cases, collaborative sharing and analysis of intelligence information, and value fusion mining of multi-source data. The paper based on blockchain and federated learning, a "end-chain-end" decentralized layered distributed architecture and related functional modules were designed, and the data processing, model training, chain collaboration, and chain application involved in the functional modules were elaborated in detail. At the same time, a specific analysis was conducted on the advantages of the system by combining security, efficiency, and persistence. The various functional modules operate independently and collaborate closely, which helps to achieve data security sharing, collaborative modeling, and accurate early warning under the participation of multiple institutions, providing a new theoretical and application paradigm for telecommunication network fraud early warning.

Keywords: Telecommunication Network Fraud; Early Warning; Blockchain; Federated Learning; Hierarchical Distributed Architecture

1. Introduction

The convenience, diversity, universality, and inclusiveness exhibited by the derivative development of telecommunication network technology and related new business forms are transforming and improving people's daily lives and work, and facilitating the country's innovative development. However, they also provide conditions for criminals to engage in telecommunication network fraud. Telecommunication network fraud often exhibits characteristics such as tight organization, diverse methods, strong concealment, complex and difficult-to-trace fund flows. It not only causes serious damage to personal, family, and social property, but also leads to the loss of the original function of finance and the disorder of national financial order, thereby inducing systemic financial risks and social trust crises [1]. To this end, it is necessary to adhere to the people-centered principle, fully implement measures for prevention, control, and management, and resolutely curb the frequent and high occurrence of telecommunication network fraud crime. Correspondingly, how to promptly and effectively detect and identify telecommunication network fraud crime has become the key to comprehensive governance of such crimes [2].

The telecommunication network fraud early warning utilizes big data intelligent analysis technology as a carrier, comprehensively leveraging internal police data as well as diverse social data from areas such as people's livelihood, government affairs, and finance to conduct early warning analysis and research on fraud-related types, personnel, accounts, funds, and other information, which can provide modular component support for building an integrated governance platform for telecommunication network fraud crimes. However, the privacy protection and trusted sharing involved in massive intelligence data processing can lead to issues such as data silos, increased costs, uncontrollable data, and inability to ascertain ownership, thereby reducing the willingness and extent of data sharing, as well as the insufficient exploration and utilization of data value, affecting the effectiveness of early warning for telecommunication network fraud crime.

As an emerging distributed storage and processing technology based on cryptography,

blockchain possesses significant advantages such as decentralization, traceability, anonymity, and incentive-compatibility, which are crucial in addressing issues faced by early warning system for telecommunication network fraud, such as data silos, uncertainty of ownership, and enhancing the willingness of participating parties to share data. However, on the other hand, in the face of massive heterogeneous fraud-related data and complex scenarios, performance bottlenecks may easily arise in consensus, audit, and parameter transfer. For massive data scattered across different organizations and institutions, federated learning only requires uploading the local updates (gradient information) from local model training to a central server for model aggregation, to obtain the final global model. This approach not only reduces costs and improves performance but also ensures data controllability, thereby better safeguarding data privacy and security [3]. In light of this, this paper combines the principles and characteristics of blockchain and federated learning to construct a telecommunication network fraud early warning system based on blockchain and federated learning, providing a new theoretical and application paradigm for comprehensively enhancing the effectiveness of telecommunication network fraud early warning.

2. Literature Review

As a typical representative of non-contact crime, telecommunication network fraud faces the difficulties of difficult investigation, tracing the source, and recovering losses, which also poses severe challenges to the current forms of economic crime and the market economic order. Against this backdrop, compared to the lagging passive crackdown and governance, the telecommunication network fraud early warning based on the limited nature of criminal subjects, the regularity of criminal activities, and the predictability of criminal behavior has important practical significance. In the early 1990s, the emergence of CompStat, a public security information management system based on statistical analysis of map data, marked the birth of predictive policing [4]. Subsequently, in 2013, the RAND Corporation systematically elaborated on the categories, technologies, and processes involved in predictive policing in "Predictive Policing: The Role of Crime Prediction in Law Enforcement Applications" [5]. In recent years, some targeted research has also been conducted on non-contact crime early warning. Jenga et al. [6] evaluated state-of-the-art crime prediction techniques that are available in the last decade, discussed possible challenges, and provide a discussion about the future work that could be conducted in the field of crime prediction. Zhu et al. [7] utilized target group index analysis and adata technology to conduct victim profiling analysis on 2,747 reported cases of telecommunication network fraud in Chinese Mainland and Hong Kong, and based on this, designed and developed a fraud message classification and filtering reminder device using data mining, machine learning, and natural language processing algorithms, ensuring that users receive as few fraudulent messages as possible, thereby curbing the high incidence of telecommunication network fraud at its source. Chen et al. [8] utilized big data analysis and artificial intelligence technology to design the TGAI-FPF, which can capture fraudulent behavior profiles in telecommunications networks, and thus laying a solid foundation for constructing behavior profiles to combat telecommunications network fraud. Focusing on the non-emotional features of cybercrime-related texts, Wei et al. [9] constructed a cyber crime early warning model based on social stance and text features through three steps: top-level ontology construction, media discourse collection, and dissonance calculation, which demonstrated good early warning performance in terms of precision, recall, and F1 value under text data, but its early warning capability for multimodal features consisting of text, audio, and video needs further improvement. Furthermore, in the construction of a telecommunication network fraud early warning model, facing massive, multi-source, and heterogeneous data from various parties, it is particularly crucial to ensure data controllability while maintaining privacy security. The distributed training and aggregation of federated learning enables the training of the early warning model to be completed locally on the client side with privatized data, and only requires the transmission of gradient information from the local model in each training round, eliminating centralized data sharing, which reduces overhead to some extent and avoids potential data uncontrollability and privacy leakage. Khan et al. [10] introduced a privacy-preserving federated learning algorithm for financial crime detection,

strategically emploied differential privacy and secure multiparty computation to guarantee the privacy of training data. To simultaneously address model heterogeneity and data heterogeneity, and improve the applicability of federated learning model, Huang et al. [11] designed a novel adaptive heterogeneous federated learning method through logical layer distribution alignment output, collaborative knowledge sharing, and weight adaptive updating, which has better intra-domain accuracy and cross-domain generalization ability, providing a benchmark for the future development of heterogeneous federated learning. Arora et al. [12] combined federated learning with multi-party computation and noisy aggregates inspired by differential privacy to propose a privacy-preserving financial crime detection scheme which enables financial institutions to jointly train accurate anomaly detection models.

Although federated learning can provide support for telecommunication network fraud early warning to a certain extent, it also faces three major challenges in practical application: centralized processing, lack of incentive mechanism, and low robustness [13]. Correspondingly, the unique technical features of blockchain related to decentralization can effectively alleviate the shortcomings faced by federated learning. Furthermore, its mechanisms such as consensus, anonymity, traceability, auditability, and immutability can further enhance the security, privacy, and reliability of the early warning model for telecommunication network fraud crime, which provides strong guarantees for effectively addressing issues like centralized single-point failures, inability to confirm data ownership, and data forgery and fraud, thus ensuring the effectiveness of early warning against telecommunication network fraud crime. Meanwhile, the current extensive research on blockchain in areas such as distributed processing, incentive mechanisms, and robustness can provide references for the construction of the telecommunication network fraud early warning architecture in this paper [14-17].

In summary, the integration of blockchain and federated learning facilitates the realization of data ownership confirmation, data controllability, and privacy protection. Meanwhile, it inherently possesses advantages in enhancing the willingness to share data, avoiding potential centralized single-point failures, and preventing data forgery and fraud. Therefore, for telecommunication network fraud early warning, the combination of blockchain and federated learning can achieve complementary advantages, thereby helping to build a multi-source heterogeneous data fusion and processing infrastructure to alleviate the current predicaments this field faces. Next, this paper conducts detailed design and analysis on the architecture, process, implementation, and characteristics of the telecommunication network fraud early warning system by integrating the principles and features of blockchain and federated learning.

3. Related Technologies

This paper comprehensively adopts blockchain and federated learning technologies to realize the mining, sharing, analysis, and modeling of early warning intelligence information, thereby constructing a systematic, comprehensive, and sustainable telecommunication network fraud early warning system.

3.1 Federated Learning

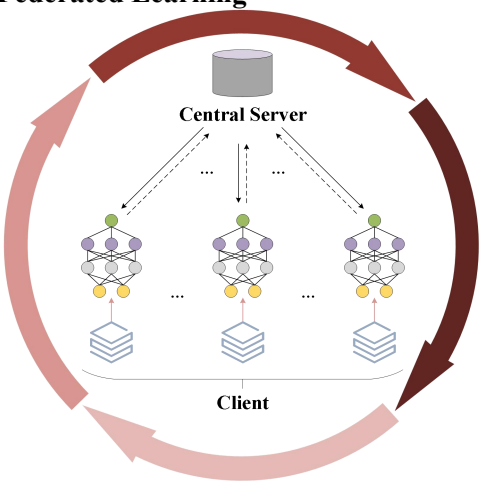

Figure 1. Federated Learning Process

As a distributed machine learning paradigm, the learning process of federated learning is illustrated in Figure 1. First, each participant uses local data to train the initial model distributed by the server. After that, the client uploads the gradients of the trained local model to the central server for global model aggregation. This process iterates repeatedly for n rounds until the global model converges or meets the requirements of practical application. Currently, federated learning can be further categorized into horizontal federated learning, vertical federated learning, and federated

transfer learning based on the degree of overlap between samples and sample features [18]. Considering the characteristics of users and their data related to telecommunication network fraud, vertical federated learning—designed for scenarios where there is high overlap between samples but low overlap between sample features—is relatively more suitable for the telecommunication network fraud early warning. Specifically, the training of the early warning model can be completed by vertically splitting the training data based on the dimension of sample features. In addition, by examining the process of federated learning, for the modeling of telecommunication network fraud early warning which involves multiple entities such as civil services, government affairs, finance, and public security, federated learning not only helps address issues related to costs, privacy, and controllability in the transmission and processing of massive amounts of data but also facilitates overcoming the "data barriers" and "data silos" formed by privacy protection measures between departments or industries, thereby realizing the effective cross-departmental sharing and collaboration of early warning intelligence information and computing resources for telecommunication network fraud.

3.2 Blockchain

The layered architecture of blockchain is shown in Figure 2, which consists of the data layer, network layer, consensus layer, incentive layer, contract layer, and application layer from bottom to top.

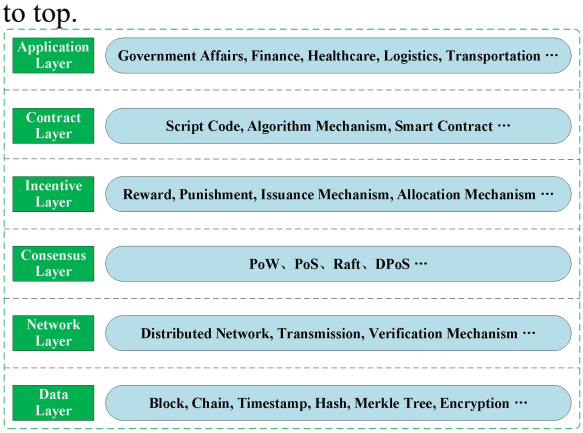

Figure 2. Blockchain Architecture

For the telecommunication network fraud early warning, since federated model training data is only stored on the local clients of each participant, the only information exchanged with the central server is gradient information. Therefore, the quantity and quality of data from each participant will directly affect the accuracy of the global model, and the effectiveness of model aggregation also depends on the enthusiasm and credibility of all parties in participating in training [3]. In addition, the overall security and availability of the model rely on the central aggregation server, and a single-point failure of this server will have a global impact. Correspondingly, the layered architecture of blockchain is shown in Figure 2, which consists of the data layer, network layer, consensus layer, incentive layer, contract layer, and application layer from bottom to top, the principles and characteristics of each layer of blockchain not only well make up for the shortcomings of federated learning but also possess unique advantages of its own:

Peer-to-peer networks, consensus mechanisms, and distributed data storage ensure the decentralization of blockchain, avoiding the centralized single-point failure of federated learning. Furthermore, the immutability and auditability formed by timestamping, encryption technologies, chain structures, and smart contracts built on this foundation effectively prevent data forgery and fraud, thereby further enhancing the security of the model;

Hash functions, asymmetric encryption, distributed storage, ring signatures, and consensus mechanisms enable data and identity anonymity, enhancing the level of privacy protection;

Reward and punishment incentive measures based on issuance mechanisms and distribution mechanisms promote the willingness and enthusiasm of all parties to participate in data sharing and interaction;

The chain structure, timestamp, hash function, combined with distributed data storage, consensus mechanisms, and smart contracts, ensure the traceability of data, thereby solving the challenge of data ownership confirmation.

Currently, blockchain is mainly categorized into private blockchains, public blockchains, and consortium blockchains based on the degree of decentralization, coverage scope, and specific application scenarios. Among them, the number of nodes in a consortium blockchain is strictly selected and restricted, and it is usually jointly participated in and maintained by specific organizations or institutions with a certain level of trust relationship. In addition, consortium blockchains can rely on access control

permissions to impose strict restrictions and management on data access and usage. On the premise of meeting relevant legal and regulatory requirements, they maintain a certain degree of decentralization, thereby striking a balance between privacy, security, supervision, efficiency, and performance. While demonstrating credibility and reliability, they more comprehensively realize business collaboration, information sharing, and risk control. Therefore, compared with private blockchains and public blockchains, the access mechanism and security settings of consortium blockchains can better align with the application scenarios of telecommunication network fraud early warning.

4. Telecommunication Network Fraud Early Warning Architecture

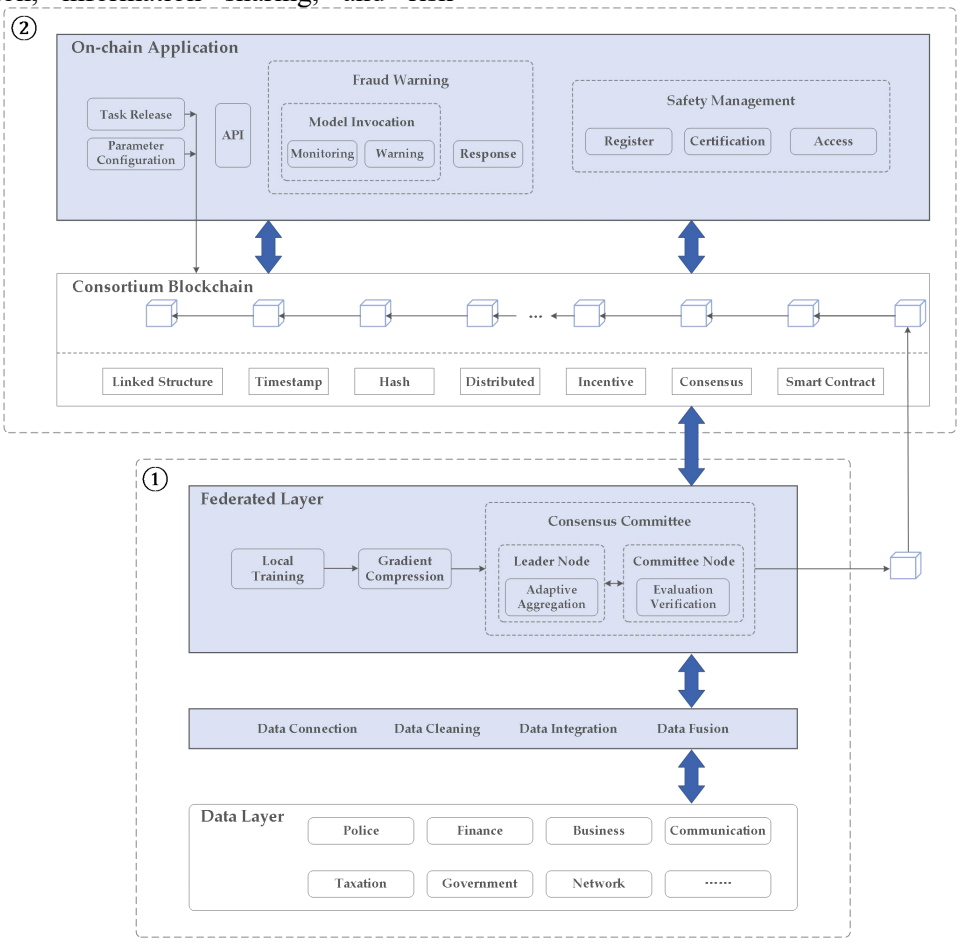

Figure 3. Telecommunication Network Fraud Early Warning Architecture Consist of ① Participant's Local Client and ② Blockchain Business End

The telecommunication network fraud early warning architecture based on blockchain and federated learning is shown in Figure 3. It can be seen that the system adopts a "terminal-chain-terminal" layered structure, including the terminal data resource layer, terminal federated learning layer, and chain-end business layer. Next, the functional modules involved in each layer will be designed and elaborated in detail.

4.1 Data Resource Layer

For the telecommunication network fraud early warning of involving multiple entities, the diversity of data sources and the heterogeneity of data structures lead to inconsistencies in data format, structure, and quality [19]. To this end, drawing on the work in literature [20], the standardized processing on massive fraud-related data is realized through data connection, data cleaning, data integration, and data fusion as illustrated in Figure 4. This enables the leveraging of the advantages of a multi-source integrated data infra-structure to ensure that the telecommunication network early warning fraud achieves the expected effectiveness.

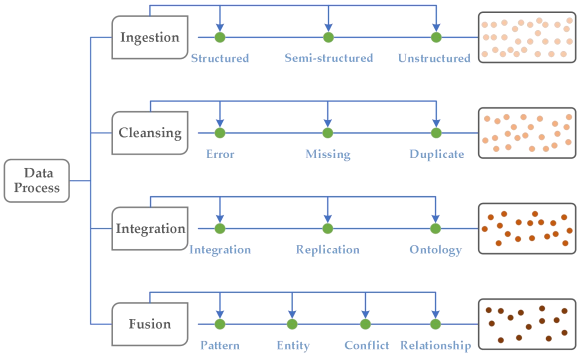

Figure 4. Big Data Processing

The diversity of data sources results in data exhibiting structured, semi-structured, and unstructured characteristics, which makes unified storage and utilization inconvenient. During the data preprocessing phase for telecommunication network fraud early warning, various terminal systems can uniformly convert multi-source data into data with structured representations that are easy for mining and utilization through data connection. For structured data, incremental or full-volume extraction methods can be adopted, where the required data items are extracted by recording the maximum extraction time each time. For semi-structured data, an ontology knowledge base can be used to decouple data content from structure, thereby converting semi-structured data into structured data for subsequent utilization. For unstructured data, its multimodal features prevent it from directly providing usable information; thus, manual annotation or intelligent algorithms [21] need to be used for data markup, so as to generate usable structured labels for unstructured data.

For the telecommunication network fraud early warning, the data collection process is affected by equipment failures, transmission interference, and specification differences, resulting in data errors, missing values, and duplicates. Data cleaning helps identify, handle, and eliminate such "dirty data," which improves data quality while providing a guarantee for subsequent data integration and fusion. For erroneous data, distance-based outlier detection can be used to identify and remove it. For missing data, screening can first be conducted based on blank fields, special symbols, or time-slice functions, and then manual supplementation or regression prediction functions can be used to fill in the missing data. For duplicate data, the BigMatch detection tool [22] is first used for entity alignment to identify duplicate data records, and on this basis, a priority queue algorithm is applied to eliminate them.

For each participating terminal in the telecommunication network fraud early warning, due to the diversity of data sources, the data remains scattered even after data connection and data cleaning are completed. This gives rise to issues such as identical names assigned to different entities, inconsistent data granularity, and failure to aggregate telecommunication fraud-related data. To address this, each participating terminal needs to select and adapt one method from schema integration, data replication, and ontology-based methods to conduct data integration, based on its own data characteristics. Specifically, for scenarios involving large-scale data with frequent updates, a mediated schema can be adopted to access the interfaces of various data sources. This establishes mappings between data sources and the mediator, thereby simplifying terminal operations. For scenarios where data sources are stable, data replication can be used to copy data from each source to a unified data warehouse for terminal utilization. For scenarios with frequent semantic conflicts, a shared vocabulary can be leveraged to resolve such conflicts between data sources, building on the ontology descriptions of each data source.

Data integration primarily addresses the issue of multi-source dispersion of fraud-related data. In contrast, data fusion, building on the foundation of data integration, further leverages a hybrid fusion strategy [23] to perform schema matching, entity alignment, conflict resolution, relationship inference, and entity fusion on the data. This process generates information gain and constructs a solid multi-source data foundation for mining intelligence information in the telecommunication network fraud early warning.

4.2 Federated Learning Layer

The federated learning layer is based on methods such as Loss Function, Stochastic Gradient Descent, and Federal Average, and combines multiple rounds of local model training, gradient compression, and global aggregation to realize the iterative construction of the telecommunication network fraud early warning model. Also, the aggregated model parameters that have gone through the consensus process are stored on the alliance chain in the form of a distributed ledger, thereby achieving

cross-departmental sharing and collaboration of intelligence information and computing resources for telecommunication network fraud early warning.

4.2.1 Local Model Training

Before the local training, the model requester first issues a federated learning task. Subsequently, it initializes the global model for telecommunication network fraud early warning, packages the model into a genesis block, and publishes the block to the alliance chain. Correspondingly, during the local model training phase, each participant downloads the global model from the bound blockchain consensus node based on the specific task, and completes local model training by combining parameters such as federated training rounds, learning rate, and gradient descent batch size.

For the modeling of telecommunication network fraud early warning, the goal of local training is to minimize the loss function of the early warning model on the local privatized dataset through multiple rounds of iteration, namely:

$$\min_{w_i^k \in W} F_i\left(w_i^k\right)=\frac{1}{N_i}\sum_{(x,y)\in D_i} l\left[w_i^k,(x,y)\right] \quad (1)$$

Where, w_i^k represents the model parameters of participant i in the k^{th} round of training; W denotes the value range of the parameters; $F_i\left(w_i^k\right)$ stands for the objective function of participant i, which is used to evaluate the error between predicted results and actual results, and its form depends on the specific task; N_i indicates the size of the local private dataset; (x,y) represents a training sample; $l\left[w_i^k,(x,y)\right]$ signifies the loss function, and D_i denotes the local dataset.

4.2.2 Gradient Compression

To improve training efficiency, mini-batch gradient descent can be used for training the local early warning model [24]. The gradient of the early warning model on the local training dataset is expressed as:

$$\nabla f\left(w_i^k;D_{bs}\right)=\frac{1}{bs}\sum_{n=1}^{bs}\frac{\partial l\left[w_i^k,(x,y)\right]}{\partial w_i^{k-1}} \quad (2)$$

where, bs represents the batch size of mini-batch gradient descent, and D_{bs} denotes the local training dataset. In this mode, the recursion manner of w_i^k is as follows:

$$w_i^k=w_i^{k-1}-\eta\nabla f\left(w_i^{k-1};D_{bs}\right) \quad (3)$$

where, η represents the learning rate, while after participant i completes K rounds of iterative training, a gradient sparsification method based on residual accumulation is adopted for gradient compression.

4.2.3 Global Aggregation

After completing the model local training, the client performs gradient compression, then signs the compressed gradient with its private key and uploads it to the blockchain consensus node in the form of a blockchain transaction. Upon receiving the gradient information uploaded by the client, the consensus node verifies the digital signature using the specific client's public key. After confirming its validity, it forwards the transaction to other consensus nodes and simultaneously calculates a contribution score for the client. Nodes with contribution scores ranking in the top 50% form the new consensus committee for the next round, and the node with the highest contribution score within the consensus committee automatically becomes the new leader node. The contribution score of client i in the k-th round of model training is calculated based on the quality of the gradient it submitted [25], and the quality scoring formula is as follows:

$$scr_i^k=\frac{1}{\frac{1}{S_v}\sum_{(x^n,y^n)\in D_v} L\left(f\left(\widetilde{w_i^k},x_n\right),y_n\right)} \quad (4)$$

where, D_v and S_v are the validation dataset and its size respectively; $L\left(f\left(\widetilde{w_i^k},x_n\right),y_n\right)$ is the model loss value; $\widetilde{w_i^k}$ is the model parameter, and its update method is as follows:

$$\widetilde{w_i^k}=w_i^{k-1}-\nabla\tilde{f}\left(w_i^k\right) \quad (5)$$

In practical application of telecommunication network fraud early warning, the intermittency and availability of client devices may lead to dynamic changes in the consortium blockchain network [26]. To address this, after all client nodes participating in training have submitted their gradient updates, the leader node can perform adaptive global gradient aggregation by combining the Federated Averaging algorithm [27]and dynamic parameters [28], so as to better adapt to the dynamic changes of the consortium blockchain network. On this basis, the leader node packages all valid transactions into a new block in the form of a Merkle tree. After being verified and approved by two-thirds of the consensus committee nodes, the new block is broadcast to other nodes in the consortium blockchain and stored on the chain, achieving tamper-proof distributed reliable storage under the premise of ensuring traceability and auditability. In addition, poisoning attack, as a common form of attack in federated learning, will seriously affect the quality of the final

aggregated model [29]. To this end, during the model aggregation phase, it is necessary to incorporate the selection rate S and only aggregate the gradient information submitted by the top S% of clients with higher contribution scores, thereby protecting the global model from the erroneous gradient information uploaded by malicious clients to a certain extent.

4.3 Business Layer

The chain-end business layer further enhances and ensures the security, robustness, and sustainability of telecommunication network fraud early warning by open-source blockchain distributed ledger technologies such as consensus mechanisms, incentive mechanisms, and smart contracts. Moreover, to implement access control, protect data security and privacy, and prevent potential malicious node attacks, a consortium blockchain is adopted to guarantee that only authorized terminals can participate in federated learning and access relevant content.

4.3.1 Blockchain Structure

As mentioned earlier, the telecommunication network fraud early warning involves a total of four roles: the model requester that initiates federated learning, the client node that performs local model training, the consensus node that verifies and forwards client models, and the leader node that packages transactions and produces blocks. Correspondingly, combined with the iterative construction process of the federated model described in Section 3.2, the designed block structure is shown in Figure 5. The block header mainly contains information such as the hash value of the previous block, the hash value of the current block, the leader's public key, the leader's signature, the timestamp, and the global gradient of this round. The block body, on the other hand, stores all legitimate transactions related to model training and aggregation in the form of a Merkle tree. And, to further enhance the decentralization degree and avoid potential single-point attacks, a newly elected leader node is not allowed to be re-elected as a leader node within a subsequent random time interval.

The telecommunication network fraud early warning involving multiple departments requires the processing of massive amounts of data. Compared with mining-proof algorithms such as Proof of Work (PoW) and Proof of Stake (PoS), the Practical Byzantine Fault Tolerance (PBFT) consensus mechanism [25] can better meet the throughput requirements of telecommunication network fraud early warning and federated learning. Indeed, before each round of consensus, the PBFT also dynamically elects consensus committee nodes and leader nodes based on the node contribution value, which can ensure the credibility and security of the consensus process. Therefore, the PBFT is used as the consensus mechanism of the telecommunication network fraud early warning proposed in this paper.

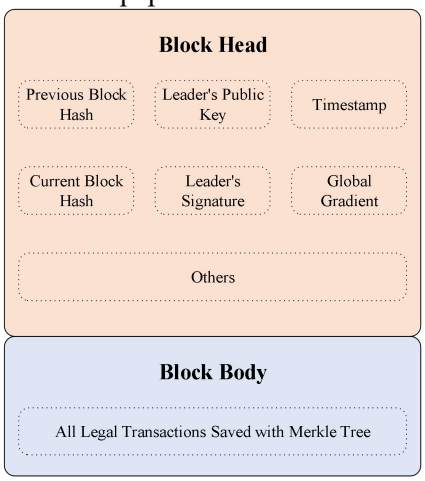

Figure 5. Blockchain Structure

The design of the incentive mechanism should not only effectively encourage high-quality members to make continuous contributions to the sound operation of telecommunication network fraud early warning, but also prevent participants with high historical contribution levels from slacking off, thereby ensuring the efficiency and sustainability of the early warning model. To this end, the incentive mechanism should consider to provide participants who continuously supply high-quality data for model training with as many rewards as possible. Specifically, in addition to considering the participants' contribution scores in the current round, their historical contribution scores should also be taken into account when calculating the reward score. By calculating the weighted average of these two scores, it can not only effectively encourage high-quality members to make continuous contributions to the sound operation of the telecommunication network fraud early warning, but also prevent participants with high historical contribution levels from slacking off.

4.3.2 Blockchain Collaboration

Consortium blockchain, leveraging chain collaboration built on consensus mechanisms, incentive mechanisms, and smart contracts,

avoids trust deficits, data silos, and ownership disputes in cross-departmental collaboration, thereby providing efficient, secure, and sustainable support for telecommunication network fraud early warning. The specific collaboration process is as follows:

✧ Access control: Based on a decentralized Certificate Authority (CA) organization, the consortium blockchain provides node identity registration, authentication, as well as transaction authorization and access control. Meanwhile, it adopts a permission isolation mechanism, only opening criminal evidence data to public security departments, which ensures privacy while optimizing storage efficiency.

✧ Consensus confirmation: After each node collects a sufficient amount of legitimate information, it verifies data consistency and adopts a Byzantine Fault Tolerance (BFT) consensus mechanism. Before each round of consensus, consensus committee nodes and leader nodes are dynamically elected based on node contribution values. In this way, verifiable data sharing is realized in the form of a distributed ledger, ensuring the credibility and efficiency of the consensus process.

✧ Block generation: The leader node packages all legitimate transactions into a new block. After being verified and approved by consensus committee nodes, the block is broadcast to other nodes on the consortium blockchain and stored on-chain. On the premise of ensuring traceability and auditability, tamper-proof distributed reliable storage is achieved.

✧ Intelligent analysis: Combined with smart contracts, collaborative analysis and modeling are conducted on cross-departmental data to build a telecommunication network fraud early warning model with privacy protection, improving the capability of telecommunication network fraud early warning. On this basis, smart contracts are used to quantify and record the data contribution values of each department, ensuring the transparency and fairness of the incentive mechanism and enhancing the participation enthusiasm of all parties.

✧ Early warning decision-making: It records experts' review and feedback on early warning signals, realizes responsibility traceability in human-machine collaboration, and ensures the accuracy of early warning.

4.3.3 Blockchain Application

The workflow of telecommunication network fraud early warning is shown in Figure 6. In the model training phase, each participating department first completes model training by combining early warning algorithm with local intelligence feature data. Then, the model parameters are shared with the central server for iterative integration to obtain a global early warning model. During this process, in addition to uploading the global model to the blockchain, the contribution values of each terminal are calculated and uploaded to the blockchain simultaneously, providing a reference for subsequent incentive schemes; In the model early warning phase, the model is used to conduct early warning monitoring on objects to be monitored. In this process, it is necessary to integrate early warning signals and expert wisdom, and the decision-making structure conducts comprehensive decision-making to achieve early warning response. At the same time, the synchronous extraction and on-chain certification of fraud-related data are completed; In the early warning response phase, the implementation of the early warning response will be tracked and feedback will be collected, so as to further feed back into the telecommunication network fraud early warning. Specifically, for positive feedback, it is necessary to combine the early warning results to complete the extraction and on-chain certification of relevant fraud-related feature data, so as to enrich and expand the training sample library and improve the accuracy of model training and early warning, while for negative feedback, it is necessary to integrate the wisdom and opinions of relevant experts to perform reverse correction on the early warning model.

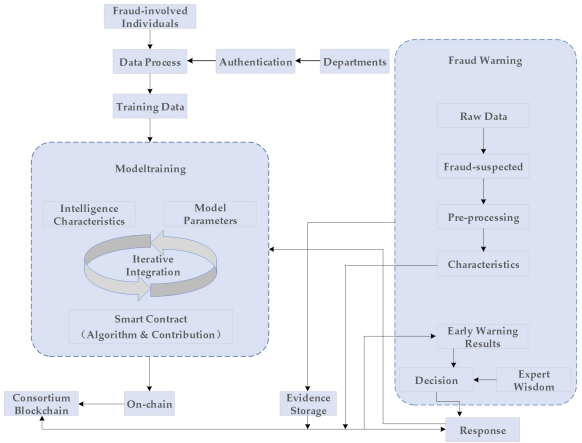

Figure 6. Network Telecommunication Fraud Early Warning Process

5. Telecommunication Network Fraud Early Warning Architecture

Next, the technical characteristics of telecommunication network fraud early warning under blockchain and federated learning will be analyzed from three aspects: security, efficiency, and durability, to demonstrate its effectiveness.

5.1 Security

Security is mainly guaranteed by the mechanism features of the consortium blockchain (including identity authentication, immutability, decentralization, and encryption algorithms) and the distributed training mode of federated learning. In terms of data privacy, data in federated learning is only trained and processed in a decentralized way locally, with only the trained model parameters uploaded; meanwhile, the encryption algorithms of blockchain provide security for data transmission and storage. This dual mechanism can effectively prevent the leakage of fraud-related data. In terms of data credibility, the chain structure, hash algorithm, and consensus mechanism of blockchain ensure that every data operation can be recorded and cannot be tampered with, thus safeguarding the authenticity, integrity, and credibility of fraud-related data. In terms of attack resistance, blockchain adopts digital signature and identity authentication mechanisms to strictly verify the identities of nodes participating in federated learning, ensuring that only authorized terminals participate in the construction of the early warning model and preventing malicious node intrusion. Additionally, the decentralized architecture of blockchain has no single target for attacks; combined with the distributed training of federated learning, it can significantly enhance the system's ability to resist attacks.

5.2 Efficiency

Efficiency is mainly guaranteed by the smart contracts, distributed architecture, and consensus mechanism of blockchain, as well as the distributed collaborative training of federated learning. In terms of automatic execution, smart contracts can automatically implement processes such as task allocation, data verification, and model training, reducing manual intervention and processing delays, and making the telecommunication network fraud early warning process more efficient and smooth. In terms of parallel processing, the distributed architecture of blockchain supports multiple terminals to participate in federated learning simultaneously. Each node can process local fraud-related data in parallel, avoiding bottlenecks in centralized data transmission and processing, and greatly improving the efficiency of data processing and model training. In terms of model iteration, the distributed training and parameter update mechanism of federated learning can make full use of the computing resources of each participating terminal; at the same time, only model parameters (rather than raw data) need to be transmitted and exchanged. Combined with the timestamp and consensus mechanism of blockchain, it can quickly integrate the training results of all parties, realize the rapid iteration of the early warning model, and promptly identify new fraud patterns. In terms of institutional collaboration, based on blockchain and federated learning, different institutions can conduct collaborative modeling without transferring data out of their local systems, share model optimization results, realize the efficient mining and utilization of cross-institutional data value, and break down data barriers and data silos between institutions.

5.3 Durability

Durability is mainly guaranteed by the incentive mechanism of blockchain, as well as the scalability of the combination of blockchain and federated learning, and expert wisdom. In terms of participation, an incentive mechanism is designed using blockchain smart contracts to reward institutions that provide high-quality data and actively participate in model training and optimization. This mobilizes the enthusiasm of all parties, continuously enriches data sources, and optimizes the early warning model. In terms of scalability, the architecture of blockchain and federated learning has good scalability, allowing for easy access to new institutions and data sources, continuously improving the early warning model to adapt to the increasingly complex situation of telecommunication network fraud. In terms of continuous optimization, based on the historical data records of blockchain and the experience accumulated from the continuous training of federated learning, expert wisdom is introduced to conduct in-depth analysis and reverse correction of the early warning model, promoting the evolutionary iteration and

upgrading of the model.

6. Conclusions

As an effective measure for the comprehensive governance of telecommunication network fraud crime, telecommunication network fraud early warning is of practical significance for protecting public property, maintaining social stability, and curbing the high incidence of such crimes. To address the shortcomings in anti-fraud work including insufficient mining and utilization of data value, inefficiencies in the research, judgment and disposal of fraud early warnings, and difficulties in the collaborative integration of terminal institutions, a "terminal-chain-terminal" decentralized hierarchical distributed architecture based on blockchain and federated learning has been constructed. Meanwhile, functional modules such as distributed storage, verification and authentication, model training, and automatic early warning have been designed. These efforts have promoted the secure and efficient cross-institutional collaborative training of multi-source intelligence information, providing new theoretical methods and application paradigms for further advancing the construction of the telecommunication network fraud early warning system. Also, this paper mainly elaborates on the functional architecture and operation mechanism of telecommunication network fraud early warning, without covering specific coding implementation; subsequent research will be carried out around this work.

References

[1] Godfred Yaw Koi-Akrofi, Joyce Koi-Akrofi, Daniel Adjei Odai, and Eric Okyere Twum. Global telecommunications fraud trend analysis[J]. International Journal of Innovation and Applied Studies, 2019, 25(3): 940-947.

[2] Veronica M. White, Joel Hunt, and Brannon Green. A discussion of current crime forecasting indices and an improvement to the prediction efficiency index for applications[J]. Security Journal, 2024, 37(1): 47-64.

[3] Hai Lan, Qian Wang, Jing Xu, Yan Xue, Bin Zhang. Review of research on blockchain-based federated learning[J]. Network and Information Security, 2024, 24: 1643-1654.

[4] Xin Lv. The characteristics of the criminal intelligence predictive analytical technology in the united states: an analysis from rand's report[J]. Journal of Intelligence, 2016, 35: 7-12.

[5] R. B. Santos. Theoretical foundations of crime analysis[M]. In Crime Analysis with Crime Mapping, 5th ed.; Sage Publications: Thousand Oaks, California, United States, 2016, pp. 22–45.

[6] Kebede Jenga, Cagatay Catal, Gokmen Kar. Machine learning in crime prediction[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14: 2887-2913.

[7] Chunjin Zhu, Chenlu Zhang, Renke Wang, Jingwen Tian, Ruoxuan Hu, Jingtong Zhao, Yaxin Ke, and Ning Liu. Building of safer urban hubs: insights from a comparative study on cyber telecom scams and early warning design[J]. Urban Governance, 2023, 3(3): 200-210.

[8] Dong Chen, and Yang Wu. Research on the use of communication big data and AI artificial intelligence technology to construct telecom fraud prevention behavior portrait[J]. Intelligent Decision Technologies, 2024, 18(3): 2589-2605.

[9] Moji Wei, Yanqing Zhao, Shiwei Zhu, and Chen Li. Early warning of cyber-crime based on social viewpoint modeling[J]. Computer Engineering and Science, 2021, 43(1): 151-160.

[10] Md. Saikat Islam Khan, Aparna Gupta, O. Seneviratne, and Stacy Patterson. Fed-RD: Privacy-preserving federated learning for financial crime detection[C]. Proceedings of IEEE Symposium on Computational Intelligence for Financial Engineering and Economics, Hoboken, NJ, USA, 2024.

[11] Wenke Huang, Mang Ye, and Bo Du. Adaptive heterogeneous federated learning[J]. Journal of Image and Graphics, 2024, 29: 1849-1860.

[12] Sunpreet S. Arora, Andrew Beams, Panagiotis Chatzigiannis, Sebastian Meiser, Karan Patel, Srinivasan Raghuraman, Peter Rindal, Harshal Shah, Yizhen Wang, Yuhang Wu, Hao Yang, and Mahdi Zamani. Privacy-preserving financial anomaly detection via federated learning & multi-party computation[C]. Proceedings of Annual Computer Security Applications Conference Workshops, Honolulu, HI, USA, 2024.

[13] Youyang Qu, Md. Palash Uddin, Chenquan Gan, Yong Xiang, Longxiang Gao, and John Yearwood. Blockchain-enabled federated learning: a survey[J]. ACM Computing Surveys, 2022, 55: 1-35.

[14] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. Blockchain challenges and opportunities: a survey[J]. International Journal of Web and Grid Services, 2018, 14: 352-375.

[15] Jingyu Zhang, Siqi Zhong, Tian Wang, H. Chao, and Jin Wang. Blockchain-based systems and applications: a survey[J]. Journal of International Technology, 2020, 21: 1-14.

[16] Qiangqiu Gan, Raymond Y. K. Lau, and Jin Hong. A critical review of blockchain applications to banking and finance: a qualitative thematic analysis approach[J]. Technological Analysis and Strategic Management, 2025, 37: 387-403.

[17] Yuanjian Zhou, Tianci Zhao, Zhengjun Jing Quanyu Zhao, Yongkang Zhu. A blockchain-based privacy-preserving data aggregation scheme with robustness in smart grids[J]. Journal of Supercomputing, 2025, 81: 675-675.

[18] Mang Ye, Wei Shen, Bo Du, E. Snezhko, Vassili Kovalev, P. Yuen. Vertical federated learning for effectiveness, security, applicability: A survey[J]. ACM Computing Surveys, 2025, 57: 1-32.

[19] Zeyu Ren, Zhenchao Wang, Zunwang Ke, Zhe Li, and Silamu Wushour. A review of multimodal data fusion[J]. Computer Engineering and Applications, 2021, 57: 49-64.

[20] Jiahe Yan, Honghui Li, Ying Ma, Zhen Liu, Dalin Zhang, Zhouxian Jiang, and Yuhang Duan. Multi-source heterogeneous data fusion technologies and government bigdata governance system[J]. Computer Science, 2024, 51: 1-14.

[21] Jia Liu, Tianrui Li, Peng Xie, Shengdong Du, Fei Teng, and Xin Yang. Urban big data fusion based on deep learning: An overview[J]. Information Fusion, 2020, 53: 123-133.

[22] William E. Yancey. BigMatch: A program for extracting probable matches from a large file for record linkage[J]. Computing, 2002, 1: 1-8.

[23] Michael Benedikt, Bernardo Cuenca Grau, and Egor V. Kostylev. Logical foundations of information disclosure in ontology-based data integration[J]. Artificial Intelligence, 2018, 262: 52-95.

[24] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent[J]. Cited on, 2012, 14: 2.

[25] S. Pahlajani, A. Kshirsagar, and V. Pachghare. Survey on private blockchain consensus algorithms[C]. Proceedings of the 1st International Conference on Innovations in Information and Communication Technology, Chennai, India, 2019.

[26] Liang Gao, Li Li, Yingwen Chen, Chengzhong Xu, and Ming Xu. FGFL: A blockchain-based fair incentive governor for federated learning[J]. Journal of Parallel and Distributed Computing, 2022, 163: 283-299.

[27] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas. Communication-efficient learning of deep networks from decentralized data[C]. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, Florida, USA, 2017.

[28] Hanzhong Peng, Zhujun Zhang, Liyue Yan, and Chenglin Hu. Research on intrusion detection mechanism optimization based on federated learning aggregation algorithm under consortium chain[J]. Network and Information Security, 2023, 23: 76-85.

[29] A.R. Short, H.C. Leligou, M. Papoutsidakis, and E. Theocharis. Using blockchain technologies to improve security in federated learning systems[C]. Proceedings of the 44th Annual Computers, Software, and Applications Conference, Madrid, Spain, 2020.

