IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 May 2024, accepted 12 July 2024, date of publication 16 July 2024, date of current version 1 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3429205

RESEARCH ARTICLE

Secure and Scalable Blockchain-Based Federated Learning for Cryptocurrency Fraud Detection: A Systematic Review

AHMED ABDELMOAMEN AHMED, (Member, IEEE), AND OLUWAYEMISI O. ALABI
Department of Computer Science, Prairie View A&M University, Prairie View, TX 77446, USA

Corresponding author: Ahmed Abdelmoamen Ahmed (amahmed@pvamu.edu)

This work was supported in part by the National Science Foundation (NSF) under Grant 2011330 and Grant 2200377.

ABSTRACT With the wide adoption of cryptocurrency, blockchain technologies have become the foundation of such digital currencies. However, this adoption has been accompanied by a surge in cryptocurrency fraud, causing significant losses to financial organizations and individuals. One way to mitigate these losses is to use Federated Learning (FL) techniques to detect fraudulent cryptocurrency transactions. This paper provides an overview of secure, privacy-preserving, and scalable Blockchain-based Federated Learning (BCFL) as a promising solution for slowing the exponential growth of cryptocurrency fraud. BCFL enables multiple entities to collaboratively train machine learning models for detecting fraudulent cryptocurrency transactions without sharing their private data, thus preserving privacy. However, Integrating differential privacy and Secure Multi-party computation (SMPC) models in BCFL presents an additional scalability challenge. This study provides an overview of BCFL, evaluating existing research on its security, privacy, and scalability challenges in detecting cryptocurrency fraud. The review explores existing research and various methodologies, highlighting advancements and challenges in creating effective, privacy-conscious fraud detection solutions for cryptocurrency transactions. We first discuss the current state of BCFL in fraud detection, along with its potential advantages and limitations, and then discuss the existing research gaps. In particular, this paper examines various BCFL frameworks, consensus algorithms, and block architectures, emphasizing their strengths and limitations in the context of cryptocurrency fraud detection to develop scalable and privacy-preserving solutions. We compare various solutions that address scalability and privacy challenges in BCFL, including adopting a geographically distributed cloud computing model that utilizes SMPC and lightweight consensus algorithms and protocols to manage computational overheads.

INDEX TERMS Blockchain, cryptocurrency, fraud, federated learning, scalability, security, privacy, literature review.

I. INTRODUCTION

The exponential growth in cryptocurrency fraud, crypto fraud for short, has caused significant financial losses in recent years [1]. For instance, the FBI's Internet Crimes Complaint Center (IC3) has reported a 183% increase in crypto fraud in the United States in 2022, resulting in a loss of 2.57 billion dollars [2]. Despite the progress in security and privacy solutions, crypto fraud continues to rise, mainly due to the absence of inter-organizational collaboration and privacy concerns.

The growing emphasis on Federated Learning (FL) as a promising solution for crypto fraud detection has inspired and driven a paradigm shift from centralized to decentralized security and privacy architectures [3]. FL was introduced by Google in 2016 as a distributed Machine Learning (ML) paradigm that leverages local computational power and enables multiple entities to collaboratively train an ML model using local data samples without sharing their sensitive data, ensuring privacy [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Peter Langendoerfer.

Traditional machine learning approaches require consolidating all data in a centralized repository, typically a data center managed by cloud service providers [5]. This practice raises privacy concerns for data owners and risks violating data confidentiality regulations among organizations [6]. As a result, the development of privacy-preserving ML techniques has become crucial [7]. The need to protect data privacy and the shift towards decentralized data collection led to FL's emergence [8].

The training of an FL model commences with the distribution of an initial model from a central cloud server to several clients that have the distributed training datasets. Each client computes local training updates, known as Stochastic Gradient Descent (SGD), to be aggregated with the global model's updates. The central server then aggregates these local updates to formulate an updated global model. This iterative process of local update computations and aggregations persists until the model converges to an optimal state [9].

Despite the superior performance capabilities of FL – facilitated by its distributed architecture, optimized bandwidth consumption, and low latency–, it is full of limitations, particularly concerning security and privacy [10]. The system's inherent dependence on a central FL server for aggregating model parameters exposes it to vulnerabilities such as Single Points of Failure (SPoF) and Distributed Denial-of-Service (DDoS) attacks [11]. Furthermore, privacy risks, including inference attacks, remain a significant concern. Such attacks allow adversaries to deduce raw datasets used in the training phase of local and global models. For instance, a privacy leakage scheme introduced in [12] enables adversaries to infer label features from global gradient parameters and reconstruct raw training data. This privacy risk often discourages organization participation in the FL model training process. Therefore, mitigating raw data inference attacks is paramount for fostering trust and encouraging the involvement of multiple crypto entities in the federated process.

In this context, blockchain has emerged as a critical player in the field of FL, primarily due to its inherent support for decentralization [10]. The decentralized nature of blockchain systems makes them less vulnerable to the attacks mentioned above and versatile in various applications, particularly in financial systems where third-party trust is essential to ensure the integrity and immutability of information in a guaranteed manner [13], [14]. Integrating both methodologies, commonly called Blockchain-based Federated Learning (BCFL), has been widely adopted in various applications, including fraud detection. In detecting crypto fraud, the data utilized for model training is sourced from multiple participating entities, such as crypto security exchanges, financial institutions, and organizations, to converge the FL model. These contributing entities possess the capability to improve the quality of model updates. However, such interventions also potentially degrade the model's performance, underscoring the importance of data quality and security in this large-scale joint modeling process. To maintain the integrity and reputation of blockchain networks, the timely detection of fraudulent transactions is paramount. Comparative studies have been conducted to determine the most effective ML algorithms for fraud detection hosted on blockchain networks based on a balance between accuracy and computational speed [15].

The primary motivation behind this study is driven by the critical concerns surrounding data privacy for the participating entities in BCFL model training, particularly in the financial sector, where data security is crucial. These concerns arise from the potential for inference attacks and malicious attacks from participating distributed entities or on central servers, leading to the compromise of the global model or breaches of user privacy [16]. Also, the lack of scalable methods to meet the computational demands of BCFL for fraud detection poses a significant challenge. Existing solutions to address computational demands primarily focus on small-scale FL systems [17].

To address these data privacy concerns, several FL frameworks were proposed in the literature [13], [16], [18], [19], [20], [21], [22] where only trained models are shared with the participating entities instead of the raw datasets, thus preserving data confidentiality and utility while upholding privacy standards. For instance, Qu et al. [10] proposed a fully decentralized FL-based framework, which leverages blockchain technology as the foundation and the Proof-of-Work (PoW) consensus mechanism to confront poisoning attacks. Lu et al. [23] presented a framework to overcome obstacles in data sharing by merging DL with a permission blockchain system. The proposed framework creates secure links between Internet of Things (IoT) devices at the network edge via encrypted logs, which are upheld by central IoT nodes, guaranteeing data confidentiality and availability. Rather than preserving unprocessed data, the framework utilizes a permissioned blockchain to retrieve pertinent data and manage data access, thereby resolving storage limitations and privacy issues.

Crypto fraud detection would require multiple entities to participate in training a global FL model in cross-silo FL settings [3]. The participating entities need to share their crypto exchange data and information about their financial data silos. To coordinate such a data-sharing process, a manager is designated as one of the participating organizations or a trusted third party responsible for coordinating the model training process. However, selecting a trusted manager poses a significant challenge, as organizations may have differing priorities and levels of trust. While a fully decentralized BCFL eliminates the need for a trusted central authority, it introduces complexities in managing communication overhead, ensuring efficient model training, and maintaining system stability. Furthermore, the performance of the aggregation process is bounded by the quantity and quality of local computing resources [4].

Table 1 lists the abbreviations and their meanings used in this article. The contributions of this paper are fourfold. First, this paper reviews the existing BCFL approaches for crypto

TABLE 1. Abbreviations and their meanings used in this paper.

Abbreviation	Meaning
FL	Federated Learning
BCFL	Blockchain-based Federated Learning
HFL	Horizontal Federated Learning
HE	Homomorphic Encryption
VFL	Vertical Federated Learning
FTL	Federated Transfer Learning
PoA	Proof of Authority
PoS	Proof of Stake
PoW	Proof of Work
SpoF	Single Point of Failure
DID	Decentralized Identifier
DDoS	Distributed Denial of Service
DP	Differential Privacy
SMPC	Secure Multi-Party Computation
IC3	Internet Crimes Complaint Center
ML	Machine Learning
SGD	Stochastic Gradient Descent
IoT	Internet of Things
VAE	Variational Auto-Encoders
DAG	Direct Acyclic Graph
DHFA	Homomorphic-encrypted Federated Averaging
MIMF	Market Integrity and Major Frauds
IID	Independent and Identically Distributed
D2C	Data-driven Cognitive Computing
MDP	Markov Decision Process
WGAN	Wasserstein Generative Adversarial Network
P2P	Peer-to-Peer
TFP	Prediction of Traffic Flow
DTN	Digital Twin Networks
2PC	Two-Phase Commit Protocol
ACP	Atomic Commit Protocol
3PC	Three-Phase Commit
ZKP-FL	Zero-Knowledge Proof-based Federated Learning
ECC	Elliptic Curve Cryptography
FLoC	Federated Learning of Cohorts
LDP	Local Differential Privacy
EM	Expectation Maximization
IPFS	Interplanetary File System

fraud detection, including an extensive analysis of BCFL frameworks, consensus algorithms, and block structures, and emphasizes their applicability, strengths, and limitations. Second, we highlight the scalability issues of the existing BCFL solutions for crypto fraud detection and propose a scalable approach capable of effectively managing extensive data from multiple participating financial entities. Third, this paper extensively analyzes the existing security and privacy approaches for BCFL and proposes the best-fit approaches for crypto exchanges of the financial entities participating in the global FL model during the training phase to reduce the risk of inference and malicious attacks. Fourth, we explore the existing BCFL cross-silo settings in the case of eliminating the centralized manager role and employing a fully decentralized approach for managing the communication overhead and maintaining the system's stability.

The rest of the paper is organized as follows: Section II presents related work. Section III presents a comprehensive technical background review of BCFL and a comparison between its existing crypto fraud frameworks. The deployment of scalable and secure BCFL strategies in detecting fraud is presented in Section IV. Section V presents a performance evaluation of the reviewed BCFL solutions. Finally, Section VI summarizes the results of this work.

II. RELATED WORK

The orchestrated and integrated design of FL and blockchain systems (i.e., BCFL) for crypto fraud detection is becoming increasingly important [3]–various studies focused on implementing BCFL solutions using machine learning that incorporate privacy-preserving mechanisms have been presented in academia and industry [3], [10], [24], [25], [26], [27], [28], [29], [30], [31], [32]. For instance, Qammar et al. [33] systematically analyzed the existing challenges of securing FL using blockchain. This study explored BCFL's existing approaches, focusing on security and privacy aspects, including securing financial records and rewards, as well as verification and accountability.

A secure federated learning framework using homomorphic encryption and verifiable computing was presented in [34]. The authors proposed a secure FL model that prevents adversaries from inferring communicated sensitive information. The proposed approach enables collaborative FL model training without exposing raw data to the participating entities. Another work proposed by Baracaldo et al. [25] used homomorphic encryption to provide security capabilities to FL models. The authors, who are IBM researchers, presented a Fully Homomorphic Encryption (FHE) framework that allows computations to be executed directly on the communicated encrypted data without decryption. This implies that the aggregator entity in an FL system does not need to access the training dataset to train the FL model.

Lin et al. [35] proposed an FL approach that leverages Variational Auto-Encoders (VAE) to mitigate the risk of sensitive data being reconstructed through inference. The proposed approach allows participants to utilize their original data to generate synthetic data using VAE, which is then used to train an FL model, ensuring the privacy of the data. Another privacy-preserving FL framework that utilizes differential privacy techniques to protect model updates from inference attacks was presented in [24]. The framework employed a secure aggregation mechanism to combine local data updates while preserving data privacy.

Li et al. [36] proposed a federated gradient boosting decision trees framework that protects data privacy by training decision trees sequentially and combining the final models. This approach ensures that the raw data remains on the local devices, preventing direct exposure to the central server. Wang and Hu [37] presented a review on BCFL technologies and examined how blockchain can be applied to FL, focusing on system composition. The authors analyzed the concrete functions of BCFL from a mechanism design perspective, highlighting the existing concerns that blockchain addresses in the context of FL. Addressing these concerns, Ullah et al. [38] proposed a BCFL-based approach to utilize a Proof of Authority (PoA) consensus algorithm. The proposed framework harnesses the immutability and

transparency of blockchain to ensure data integrity and privacy throughout the FL process.

Various existing works discuss using BCFL to address privacy and security concerns in different domains. For instance, Yuan et al. [39] proposed a two-layer approach for FL in the IoT context. The proposed system divides the network into sub-chains, known as shards, to control and regulate the exchange of sensitive data among the communicating participants. Additionally, the proposed approach employs a main-chain layer based on a Direct Acyclic Graph (DAG) framework to enable asynchronous validation across the deployed shards. Sharding and DAG mechanisms addressed the scalability and efficiency challenges in BCFL. In particular, sharding is a technique for dividing a large blockchain into smaller, more manageable sections called shards to allow the network to process and manage higher data volumes and perform more volume of transactions efficiently.

A different study by Guo et al. [40] proposed a dual-layer blockchain structure for secure traffic prediction based on FL. The lower and upper layers of the blockchain store the aggregated parameters of the local and global models, respectively. The proposed system utilized a Distributed Homomorphic-encrypted Federated Averaging (DHFA) scheme that tackles the issues related to securing the distributed computations. In another related work presented by Aurna et al., [41], a highly accurate and privacy-preserving system for detecting credit card fraud using FL is proposed, in which three deep learning models (i.e., CNN, MLP, LSTM) are used within the FL framework. The proposed approach explored four distinct sampling techniques to address data imbalance, and the proposed framework is compared with other state-of-the-art models.

Significant progress has been made by law enforcement in the field of crypto fraud detection. For example, the Market Integrity and Major Frauds (MIMF) unit [42] has emerged as a force in combating fraudulent activities and market manipulation associated with cryptocurrency since 2019. MIMF has successfully prosecuted cases involving cryptocurrency fraud with intended losses exceeding 2 billion dollars for investors worldwide. Prosecutors employed data analytics and conventional law enforcement methods to identify and bring justice to cryptocurrency investment schemes.

III. TECHNICAL BACKGROUND OF BLOCKCHAIN-BASED FEDERATED LEARNING

This section provides a technical background of BCFL technologies, including FL architectures, workflows, and communication topologies.

A. FEDERATED LEARNING OVERVIEW

FL empowers distributed computing devices to train a global model collaboratively, safeguarding sensitive data by retaining it on the originating devices throughout the training process [43]. Notably, the participating parties who

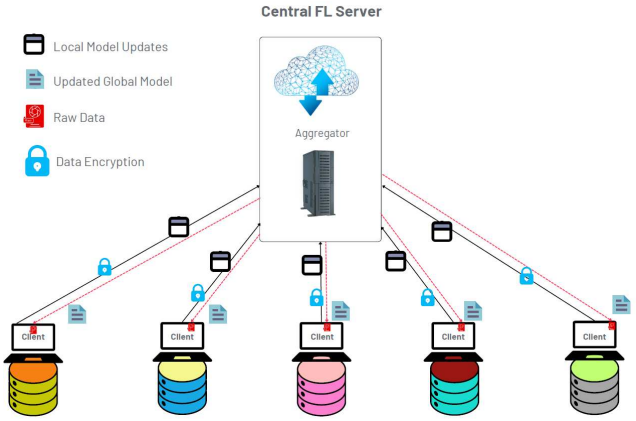

FIGURE 1. Federated learning architecture and workflow.

contribute to federation learning without uploading their training datasets to a cloud-based data warehouse maintain the privacy of each data sample [44]. In FL, the aggregation server identifies and invites a constituency of potential contributors to the training process, and then aggregates the locally trained models when they arrive [37]. Once the training process is completed –by converging the model's gradients and weights– locally at each contributor, the training parameters are sent to the aggregation server.

In each training epoch, denoted by τ, the aggregator server disseminates the initial model updates (i.e., weights) to the selected constituency of FL contributors, denoted by μ. Each contributor computes its local model using the received weights, represented by temporally defined feeds, each with its contextual constraints. Then, these timed model feeds are sent to the server to carry out the required aggregations and transform them into a global FL model [33]. The training FL workflow can be represented formally as follows.

$$\Phi_{\tau+1} = \Phi_{\tau} + \frac{1}{\mu_{\tau}} \sum_{i=1}^{\mu_{\tau}} \pi_i \tag{1}$$

where Φ_{τ} is the aggregated global model at the τ^{th} epoch, and π_i is the current state of the local model of a contributor μ_i. The Vanilla Federated Average (FedAvg) algorithm is used to aggregate of the local model updates π_i at the server [33].

B. FEDERATED LEARNING ARCHITECTURE AND TAXONOMY

Figure 1 illustrates a simplified architecture of FL, which involves the following distributed components.

- **Clients**: Also called contributors, participating entities, or Local Nodes, they are the edge devices that hold the local data samples used for training local models at the network edge.
- **Communication Network**: It facilitates the exchange of model updates between local nodes and the aggregation server.

- **Central Server**: It orchestrates the FL learning and inference processes. It also manages the model initialization, distribution, and aggregation operations. The centralized FL server aggregates the local model updates received from clients to construct the global FL model.
- **Global Model**: It is initialized with random weights, defined as temporally feeds with contextual local updates generated from clients. The aggregation server trains this model on each epoch with clients' local updates.

As shown in Figure 1, the FL training process can be represented as a traditional client-server architecture, where local data samples, such as crypto exchanges in the crypto fraud context, undergo local training process to ensure data privacy. The intermediate gradients produced by each training epoch are desensitized and transferred to a trusted central FL server for aggregation. The aggregated global parameters are then disseminated to local clients to update their local models. The FL workflow can be summarized in the following steps.

- **System Initialization**: The central server initiates a joint FL modeling task and invites clients to participate in the training process.
- **Local Computations**: Following the joint model initialization step, the central server sends some model parameters to instruct the invited contributors to perform their local computations. Each contributor then independently executes a series of computations utilizing their dataset and sends its model updates to the central server, fostering a decentralized computation paradigm and empowering autonomous data-driven contributions.
- **Global Aggregation**: The central server aggregates the received updates from contributors and aggregates them to train the global FL model.

Figure 2 shows a simplified taxonomy of FL that shows its diverse components, including data distribution, model architecture, and system communication architecture. The figure illustrates that FL is built on four pillars: model-centric, data-centric, communication topology, and aggregation approaches. Model-centric approaches focus on initializing, updating, and disseminating FL models centrally. In particular, a central server initiates the model training process and then invites a set of contributors to participate by sending their local updates for aggregation [45]. On the other hand, data-centric approaches focus on data distribution among the participating entities or clients [46]. The system communication architecture is a categorization based on how multiple participating entities communicate with the central server. Next, we describe these four FL components.

C. ALGORITHMIC COMPLEXITY OF BLOCKCHAIN FEDERATED LEARNING

The computational complexity of BCFL algorithms is a critical factor that impacts these systems' scalability and real-world applicability. Several studies have examined the algorithmic complexity of different federated learning approaches in the context of blockchain applications [3], [12].

One of the critical components of a BCFL system is the aggregation of local models trained on distributed data sources. McMahan et al. [5] introduced the Federated Averaging (FedAvg) algorithm, which has a computational complexity of $O(K * N)$, where K is the number of local model updates, and N is the number of participating nodes. This relatively efficient linear complexity allows for scalable model aggregation, even in large-scale blockchain networks. However, the complexity of the underlying blockchain consensus mechanism can introduce additional computational overhead.

Nguyen et al. [47] proposed a federated learning-based crypto-fraud detection system that utilizes a blockchain infrastructure. They noted that the computational complexity of the blockchain consensus protocol, such as Proof-of-Work (PoW) or Proof-of-Stake (PoS), can significantly impact the overall system performance. For example, the computational complexity of the PoW consensus algorithm used in Bitcoin is $O(2^n)$, where n is the number of leading zeros required in the hash value [48]. This exponential complexity can become a bottleneck for the scalability of blockchain-federated learning systems, as the model aggregation and consensus processes need to be tightly coupled.

To address this challenge, some studies have explored alternative consensus mechanisms with lower computational complexity. Li et al. [49] proposed a federated learning-based anti-money laundering system on the Ethereum blockchain, which utilizes the Proof-of-Authority (PoA) consensus algorithm. The PoA algorithm has a computational complexity of $O(N)$, where N is the number of authorized validators, which is typically much lower than the complexity of PoW or PoS.

Another research work presented in [50] proposed a deep learning model using a unique metaheuristic optimization strategy for Ethereum fraud detection. The proposed model combined a Genetic Algorithm (GA) with the Cuckoo Search (CS) technique. The suggested technique and Support Vector Classification (SVC) models outperformed the rest with the highest accuracy. In contrast, deep learning with the proposed optimization strategy outperformed the RF model, with a slightly higher performance of 99.71% versus 98.33%.

Furthermore, using techniques like sharding and off-chain computations can potentially reduce the algorithmic complexity of blockchain-federated learning systems. Sharding allows the blockchain network to be divided into smaller, more manageable segments, reducing the computational burden on individual nodes [51]. Off-chain computations, such as those performed in side chains or state channels, can also offload some model training and aggregation tasks, improving the overall system efficiency.

In terms of real-world applicability, the algorithmic complexity of BCFL algorithms plays a crucial role in real-world applicability and in determining the system's scalability and performance. Systems with lower computational complexity, such as those utilizing efficient consensus mechanisms and

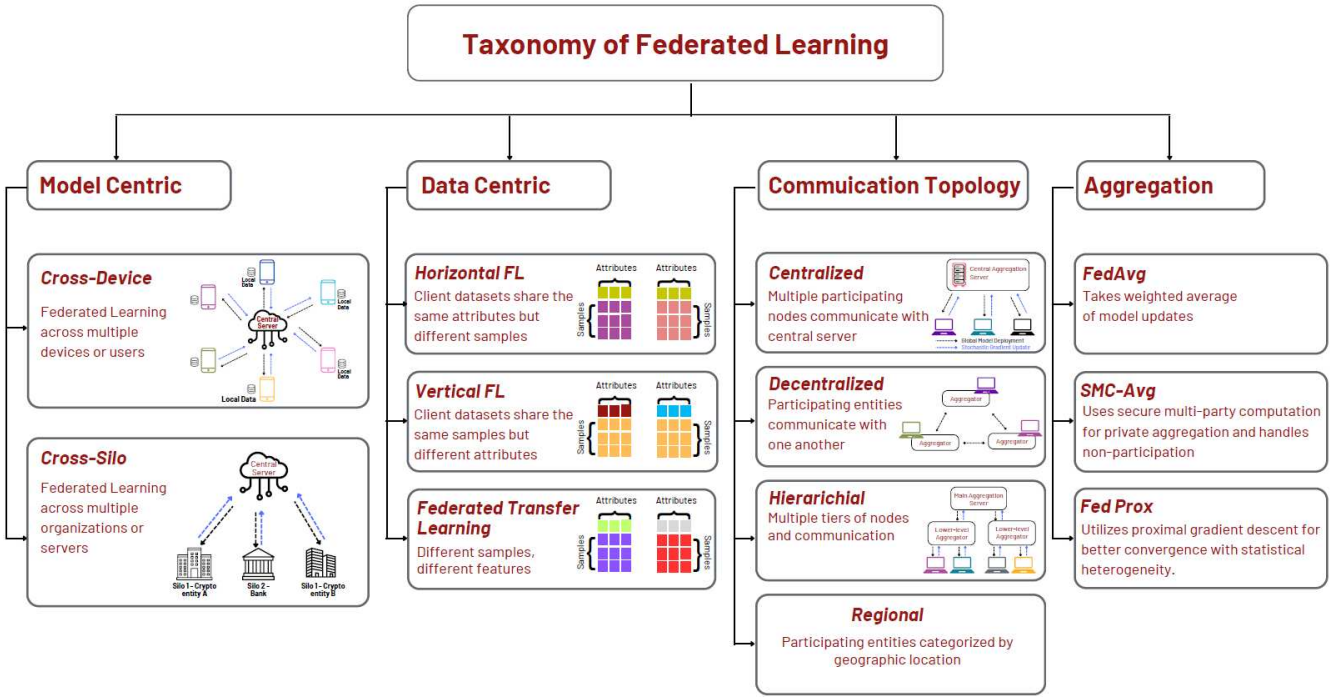

FIGURE 2. Federated learning taxonomy.

off-chain computations, are more likely to be able to handle large-scale blockchain networks and crypto fraud detection scenarios with hundreds or thousands of participating entities across the world.

D. CYBERSECURITY CONSIDERATIONS OF BCFL

BCFL paradigms have unique security considerations that must be carefully evaluated. For instance, blockchain networks are often applauded for their security features, such as cryptographic techniques, distributed consensus mechanisms, and immutable transaction records. The decentralized nature of blockchains helps mitigate single points of failure, making the system more resilient to attacks [48]. However, blockchain networks are not without vulnerabilities. The consensus protocol used, such as Proof-of-Work (PoW) or Proof-of-Stake (PoS), can be subject to various attacks, including 51% attacks, where a malicious actor(s) controls the majority of the network's computing power or stake [14]. These attacks can compromise the integrity of the blockchain and enable fraudulent activities. Also, vulnerabilities in the smart contract code or the blockchain's underlying infrastructure can lead to security breaches, such as the DAO hack on the Ethereum blockchain [50]. Attackers can exploit these vulnerabilities to gain unauthorized access, steal funds, or disrupt the system's operations.

Federated learning systems address data privacy and security concerns by training machine learning models on decentralized data sources, such as crypto wallets and exchanges, without aggregating raw data [5]. This approach helps mitigate the risk of data breaches and preserves user privacy. However, federated learning systems are not immune to security threats. Malicious participants or outside attackers can attempt to manipulate the model training process, known as model poisoning attacks [52]. These attacks can introduce biases or vulnerabilities into the global model, compromising its effectiveness and reliability.

In addition, the communication channels between the federated learning nodes and the central coordinator can be targeted by network-based attacks, such as man-in-the-middle attacks or denial-of-service (DoS) attacks [53]. These attacks can disrupt the model aggregation process or intercept sensitive information exchanged during the federated learning workflow.

Enhancing the security of blockchain networks by implementing more robust consensus protocols or additional cryptographic safeguards may come at the cost of increased computational complexity and reduced transaction throughput [4]. This can impact the real-time performance and scalability of the overall system, which is crucial for effective crypto fraud detection. Similarly, implementing advanced security measures in federated learning, such as secure multi-party computation or differential privacy techniques, may introduce computational overhead and communication latency, potentially affecting the system's ability to rapidly detect and respond to evolving fraud patterns [50].

E. FL MODEL CENTRIC APPROACH

The model-centric category can be classified as cross-device and cross-silo approaches. The cross-device FL approach involves many computing devices containing sensitive

TABLE 2. FL model centric approaches.

Metric	Cross-Device FL	Cross-Silo FL
Model Training Contributors	An enormous quantity of mobile and IoT devices	Diverse large entities such as banks and healthcare facilities
Constituency Size	Up to 1010 entities	Between 2-100 entities
Participation in Model Training	A subset of the contributing constituency	All invited contributors
Contribution Reliability	Inconsistent and unreliable	Somehow reliable with robust fault tolerance
Data Partition Axis	Horizontal and fixed partition	Predefined and can be configured either horizontally or vertically

datasets from various contributing entities. These devices, –which are predominantly mobile and IoT devices [54]–, participate in the FL model training process. However, only a subset of these devices is available for training at any given time [53]. In contrast, the cross-silo FL approach involves the use of siloed datasets to train the global FL model. The contributors in this setup are not individual devices, but larger entities such as crypto exchanges, banks, schools, and government institutions [45].

Table 2 compares the characteristics of cross-device and cross-silo approaches. The cross-silo category offers excellent stability when used for an FL setup with multiple crypto exchanges participating in the global model training. All contributors continuously provide updates for the model refinement process [45].

F. FL DATA CENTRIC APPROACH

The data-centric category can be classified as Horizontal Federated Learning (HFL), Vertical Federated Learning (VFL), and Federated Transfer Learning (FTL) approaches [46]. As shown in Figure 3, HFL converges multiple participating crypto entities with heterogeneous data samples and shared feature space. This setup enables collaborative model training where substantial overlap exists across decentralized datasets [29]. HFL-based approaches utilize expansive feature sets with considerable overlap hosted on distributed crypto entities with high commonality and a central aggregation server [55].

As depicted in Figure 4, VFL involves participating entities with distinct feature data corresponding to an identical set of samples with large datasets overlap [56]. VFL is prevalent in applications where data is vertically partitioned due to privacy regulations or ownership restrictions, such as in collaborative healthcare analysis or financial risk assessment. VFL is particularly suited for situations characterized by a substantial overlap within the sample space of the dataset [29].

HFL and VFL approaches are viable when dealing with datasets that share standard features or samples within the federation. However, in practical scenarios, the existence of such common entities may be restricted, thereby potentially diminishing the federation's appeal [57].

FTL combines the distributed privacy-preserving federated learning paradigm with transfer learning techniques, enabling knowledgeable model initialization and optimization for data-constrained client entities. It is used when the participating entities have different feature spaces and diverse datasets. In other words, the data samples are located in different domains with shared latent space. FTL approaches are typically constrained to a two-party structure, bearing significant resemblance to the protocols of VFL [26]. Furthermore, the performance of FTL systems remains contingent on cross-domain similarity, rendering this paradigm optimal for collaborative modeling between organizations in similar industries with higher feature space congruity.

G. FL COMMUNICATION TOPOLOGY ARCHITECTURE

The selection of FL communication topology architecture is crucial and contingent upon the unique domain needs of the FL system (see Figure 5). Here are the most commonly used communication topologies for FL:

- **Centralized – Client-Server Architecture**: The participating parties communicate with a central server in the federation, which manages the FL model training and aggregation process. This architecture is not scalable to a large number of participating parties.
- ***Decentralized – Peer-to-Peer Architecture***: The interactions between the participating entities are decentralized, eliminating reliance on a central server. This architecture is scalable; however, there are some complexities in managing distributed communication between the contributing constituency;
- **Hierarchical Architecture**: This hierarchical structure incorporates multiple tiers of participating entities, which facilitates the communication between the entities in the same layer for exchanging the model updates before disseminating them up to the hierarchy. While this architecture can potentially diminish communication costs and enhance system robustness, it may concurrently introduce heightened complexity in system management and coordination.
- **Regional Architecture**: The participating entities in this structure are categorized based on geographic locations, where each location possesses a regional server responsible for the communication between the participating entities within its designated region. While this architecture holds promise for enhancing communication efficiency and preserving data privacy, it may necessitate sophisticated mechanisms for group management and coordination [27].

H. FL AGGREGATION TECHNIQUES

This section presents the most popularly used aggregation techniques in FL during the federated learning process, as follows:

- **FedAvg**: It is a centralized aggregation method based on Stochastic Gradient Descent (SGD), which is an

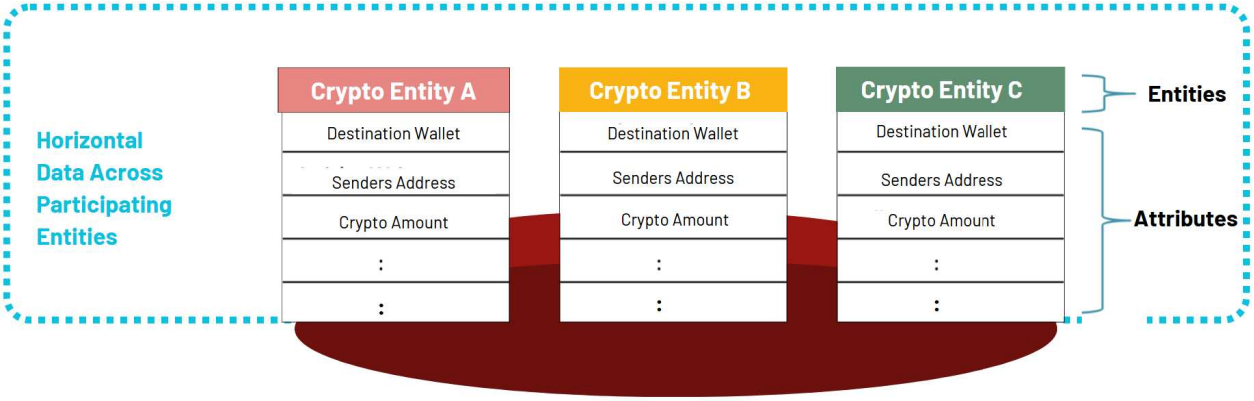

FIGURE 3. Horizontal Federated Learning (HFL).

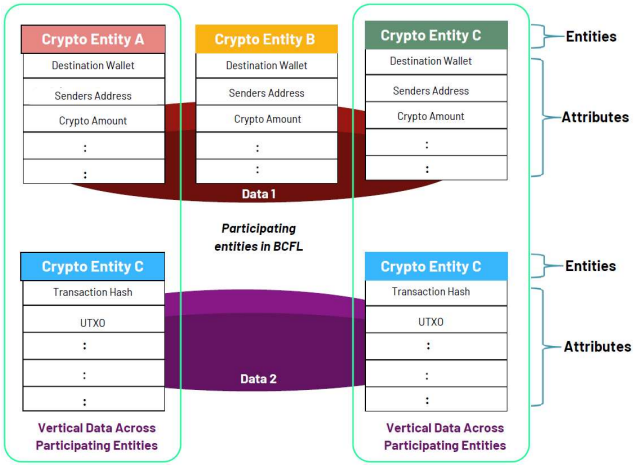

FIGURE 4. Vertical Federated Learning (VFL).

optimization algorithm for approximating the model parameters that correspond to the best fit between predicted and actual outputs. In particular, FedAvg calculates the average of updates received from the participating entities in federation [33].

- **SMC-Avg**: It is a centralized aggregation method that uses secure multi-party computations and facilitates a privacy aggregation of FL model updates from the participating entities. It notably exhibits a remarkable performance even with a low 33% non-participation rate [9].
- **FedProx**: A centralized aggregation method provides aggregation stability and supports participating device heterogeneity. FedProx is used in highly heterogeneous settings using a proximal gradient descent algorithm to stabilize aggregation and improve convergence. It provides better convergence guarantees than FedAvg when learning over data from non-identical distributions (i.e., statistical heterogeneity) [58].

I. BLOCKCHAIN TECHNOLOGIES

Blockchain is a revolutionary cryptocurrency technology that records financial transactions, leveraging a decentralized and secure digital ledger [31]. Blockchain can permanently document transactions in a transparent and tamper-proof manner. This technology utilizes a distributed network of nodes to create an immutable chain of blocks, one of which is securely linked to another block. This ensures verifiable, transparent, and tamper-proof transaction documentation, transforming how transactions are recorded and verified [59].

Network block can record and maintain financial transactions, offering advantages such as decentralization, immutability, transparency, and anonymity [59]. Transactions are shared among a network's nodes and cryptography links data blocks. Blockchain can facilitate the exchange of cryptocurrencies without an intermediary agent. Its importance lies in simplifying business operations, preventing corruption, and eliminating institutional interference. The universally accessible and verifiable permanent record of payments reduces the risk of human error and exploitation [60].

Access control in blockchain technologies can be classified into three groups: private, consortium, and public. Public blockchains offer unfettered access to any network node, while private ones restrict participation through some permissions mechanisms. Consortium blockchains occupy the middle ground, granting control over network membership and block creation to a designated group, resulting in a semi-decentralized architecture [37].

Figure 6 illustrates the blockchain system architecture, which consists of five layers, each with distinct responsibilities to enable decentralized, secure data storage and perform various transactions. The innate modular design allows different components to be further improved in terms of efficiency and performance. These layers facilitate cryptography, networking, consensus protocols, and other functions needed for transparent ledgers without centralized control. The functionalities of the five layers are described as follows:

- **Model Consumption Interfaces for Application**: It is also called the application layer, which enables user interaction with the applications deployed on the blockchain network.

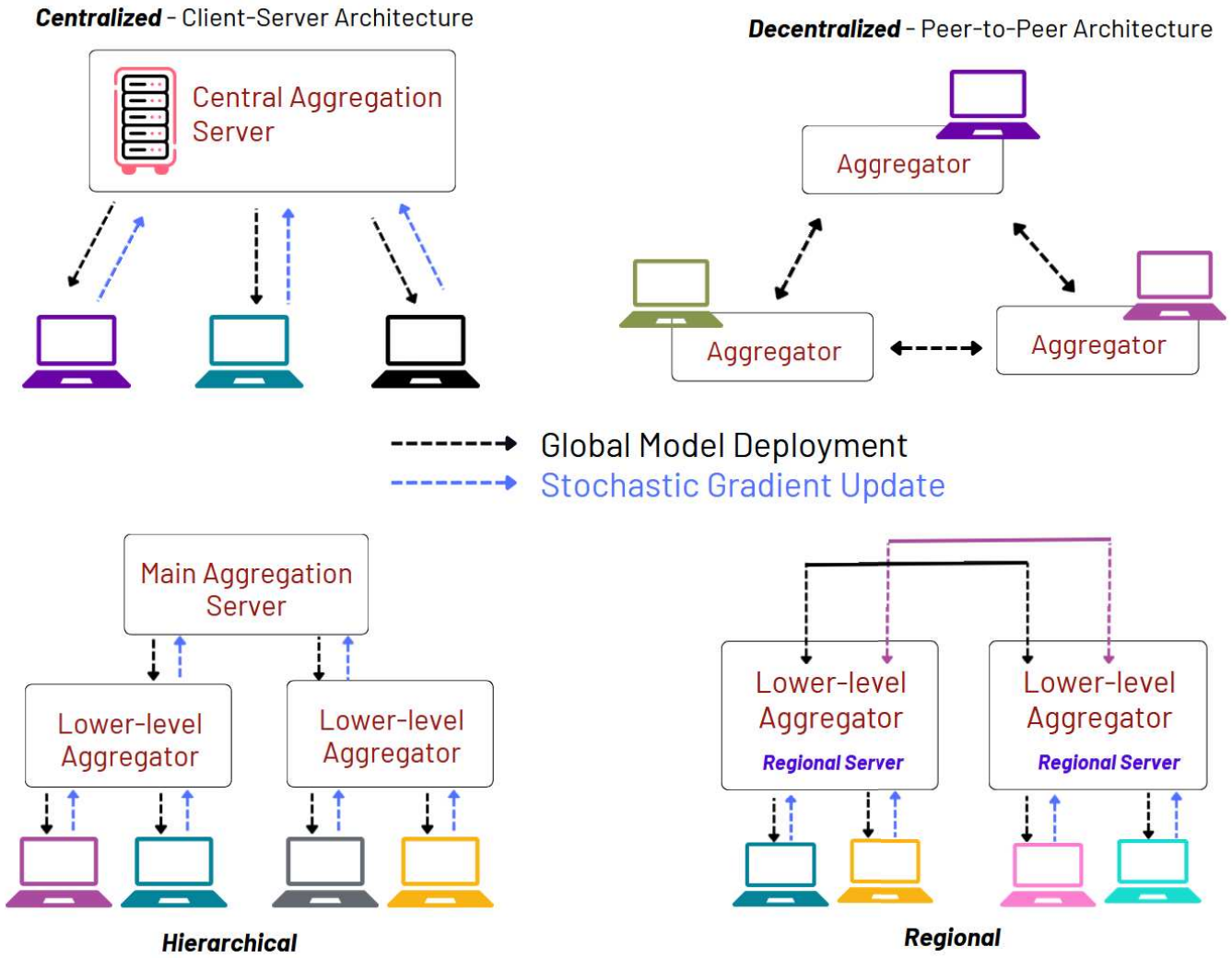

FIGURE 5. FL communication topology architecture.

- **Network Layer**: It facilitates node communication and data transmission.
- **Consensus Layer**: It maintains that all network nodes agree on the current blockchain state using consensus algorithms.
- **Value Distribution Mechanism Layer**: It is also called the incentive layer, which rewards nodes for contributing to the network's performance and security.
- **Data Layer**: It maintains data security, integrity, and immutability on the blockchain.

Blockchain applications are classified into three categories: permissioned, permissionless, and federated blockchain. Table 3 compares these categories regarding publicity, authority, security, transaction speed, and cost metrics [61].

IV. BCFL-BASED SOLUTIONS FOR CRYPTO FRAUD DETECTION

Integrating blockchain's decentralization approach with FL (i.e., BCFL) enhances the reliability of such systems.

TABLE 3. Blockchain application classification.

Metric	Permissioned	Permissionless	Federated
Publicity	Public	Private	Private
Authority	Decentralized	Centralized	Decentralized
Security	Less Secure	Most Secure	Secure
Speed and Cost	High	Lower	Lower

It eliminates the risks associated with a centralized having a point of failure in the aggregation process. Various BCFL architectures have been presented in literature [22], [62], classified broadly into three categories: (1) Fully Integrated BCFL; (2) Flexibly Integrated BCFL; and (3) Loosely Integrated BCFL.

Figure 7 illustrates the BCFL system architecture, which consists of three layers: Network Participants, Interplanetary File System (IPFS), and Decentralized Ledger layers. The system architecture is engineered as a Decentralized Identifier (DID)-oriented authentication system, which supports secure entry to the federated learning platform via self-governed DIDs documented on the blockchain [63].

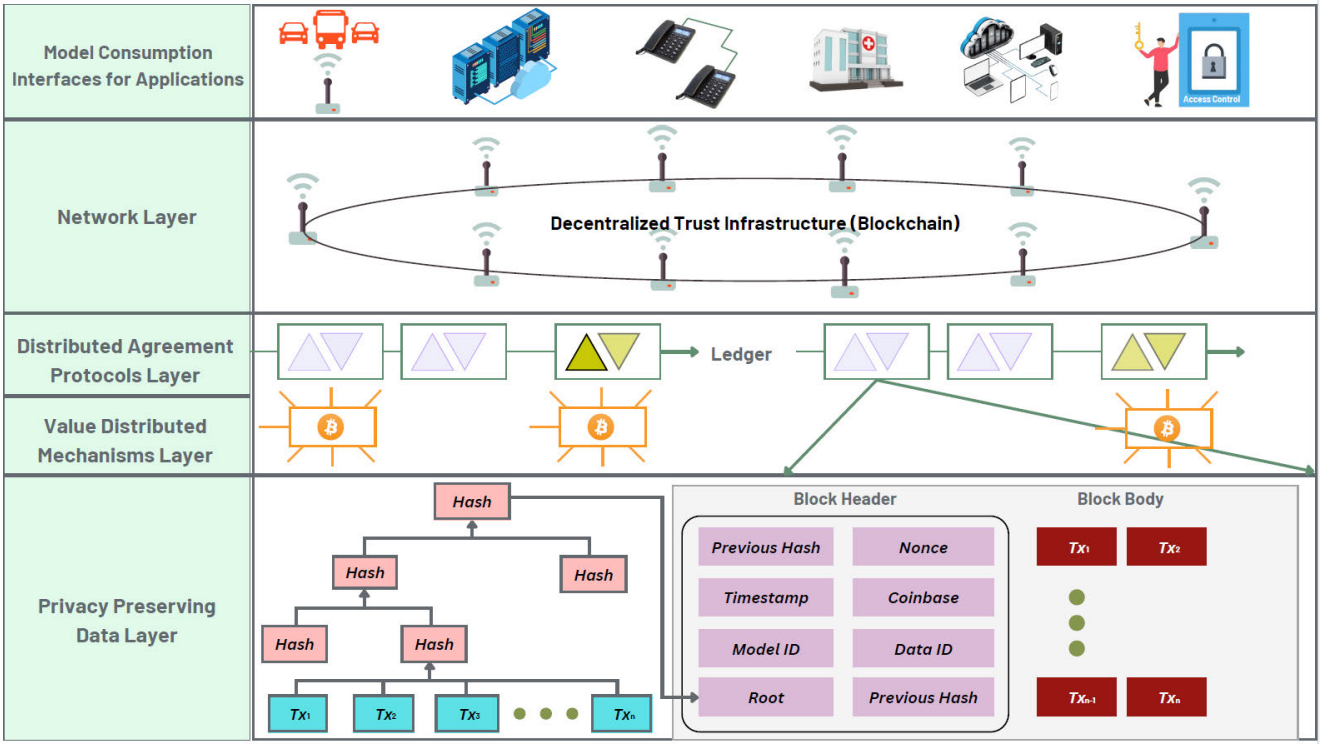

FIGURE 6. Blockchain system architecture.

Smart contracts are used in BCFL to provide safe and reliable crypto exchange operations.

The main components of the BCFL system architecture are described below:

- **The Client System** represents the participating entities interacting with the BCFL server. Its primary responsibility is to execute the training tasks received from the server.
- **The Job Creator**, also called Job Initiator, initiates the FL model training tasks and sends them to the BCFL server. These tasks include instructions for training part of the FL model.
- **The Training Configuration System** configures the training tasks by setting up parameters and conditions for the FL training process.
- **The Container** is a workspace storing the training data and FL model. It is a placeholder where model training takes place.
- **The Evaluator**, also called Validator, assesses and evaluates the effectiveness and efficiency of the learning process of the FL model.
- **The Aggregator** component collects and combines the FL model updates from the client system. It enables the sharing of these updates across all nodes.
- **The Job Distributor** assigns the training tasks to clients based on their training dataset.
- **The Token Contract** refers to a virtual contract that handles the transfer of tokens between clients and the BCFL server. These tokens can serve as rewards or incentives for participating in the federated learning process.
- **The Blockchain Contract** records transactions at the BCFL server, ensuring system integrity and security.

Figure 8 shows the workflow of the BCFL-based solutions for crypto fraud Detection, which consists of six steps divided into two stages, as follows:

- **Step 1**: In the first stage, the crypto fraud job is initiated by the Job Creator. A BCFL architecture is structured for detecting fraudulent transaction patterns, which can be learned from diverse data streams, such as transaction data, user data, and device data.
- **Step 2**: Once the BCFL architecture is developed, a diverse pool of participating entities is recruited to contribute to the crypto fraud detection job. These entities can access training data relevant to the specific crypto fraud detected. Furthermore, independent researchers and other financial institutions with access to such data can also be recruited to participate in the crypto fraud detection job.
- **Step 3**: This is the first step in the iterative refinement stage, where local FL models are independently trained using the participant's datasets, thus preserving data privacy. For instance, consider a cybersecurity firm might use a dataset from past investigations of crypto fraud over a time-bound period to train an FL model.

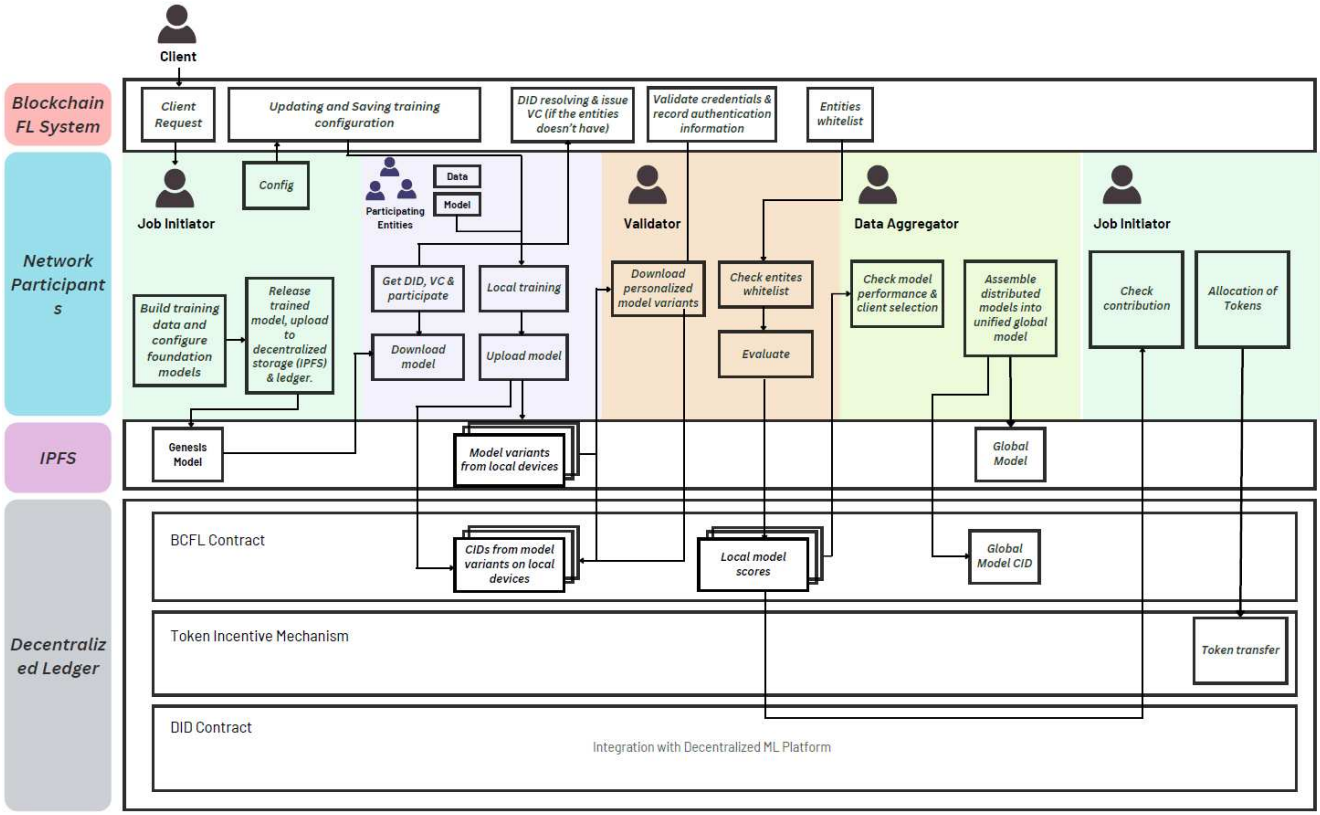

FIGURE 7. BCFL system architecture.

- **Step 4**: This step evaluates the performance of the local models of the participating entities in detecting fraudulent transactions using a validation dataset.
- **Step 5**: Based on the assessment in Step 4, a number of participating entities are selected to participate in the aggregation process to construct the global model. For instance, local models demonstrating a high detection rate are given a greater significance in the aggregation process.
- **Step 6**: In the global model deployment step, the BCFL network deploys the updated global model to all participating entities to use in the crypto fraud detection process.

This iterative phase in this BCFL workflow requires frequent collection of training data from the participating entities regularly, which is used to retrain the global model and improve its ability to detect crypto fraud transactions.

V. PERFORMANCE EVALUATION OF EXISTING BCFL SOLUTIONS

This section presents the grand challenges in developing BCFL solutions for crypto fraud detection regarding privacy, cybersecurity, scalability, and performance. It also presents the existing approaches to address these challenges and open gaps in the literature.

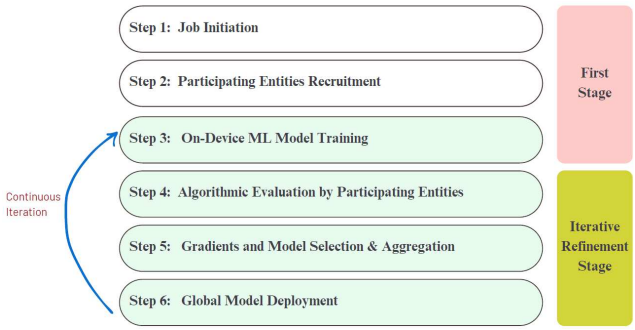

FIGURE 8. The workflow of the BCFL-based solutions for crypto fraud detection.

A. PRIVACY AND CYBERSECURITY CHALLENGES OF BCFL FOR CRYPTO FRAUD DETECTION

Table 4 presents an overview of the existing work on addressing the privacy challenges in BCFL for fraud detection. As shown in the table, one of the primary privacy and cybersecurity concerns in BCFL arises from the potential existence of malicious miners in the network. Given the crucial role miners play in consolidating the global model and achieving consensus, there exists a possible threat of malevolent miners taking advantage of weaknesses in the reward allocation system. This could cause a decline in the

TABLE 4. An overview of existing work on addressing the privacy challenges in BCFL for fraud detection.

	Privacy Approach	Privacy Challenges
Wang et al. [67]	BPFL– Fusion of homomorphic encryption and Multi-Krum	Optimizing the model's fusion and effectively managing homomorphic encryption techniques.
Zhao et al. [21]	Differential Privacy (DP) leveraging Laplace Noise.	Determining the level of noise and choosing privacy settings.
Miao et al. [22]	PBFL (Privacy-preserving Blockchain FL) leverages full homomorphic encryption & cosine similarity for secure, efficient federated learning on-chain.	The framework assumes a relatively balanced data distribution across client devices rather than situations with highly skewed or non-independent and identically distributed (non-IID) local datasets.
Sun et al. [68]	Homomorphic encryption (BCP-based) for gradient	Blockchain-based auditing adds a layer of security to encrypted gradients, but the computational overhead and latency temper its potential for large-scale deployment it introduces.
Li et al. [69]	BFLC – Blockchain-driven Federated Learning framework with Committee consensus	While the CCM approach offers benefits such as security and robustness, it comes at the cost of increased energy consumption due to the extensive communication overhead during model updates.
Lu et al. [23]	PBFL – Privacy-preserving data sharing the mechanism for distributed multiple parties	Enhancing the effectiveness of data models derived from unprocessed data is essential.
Qu et al. [70]	BFL – A distributed approach for big data-driven cognitive computing (D2C)	Enhances the Markov decision process (MDP) instead of tackling the privacy concerns associated with a blockchain that is presumed to be resistant to tampering.
Wan et al. [20]	BFL – Wasserstein generative adversarial network (WGAN)	The necessity for effective communication and computational techniques, the issues of privacy and security in federated learning, and the challenge of non-identically independently distributed (non-IID) data distribution in edge computing environments.
Shayan et al. [71]	Biscotti: a fully decentralized peer-to-peer (P2P) approach to multi-party ML	The need, for reliable data samples in multi Krum, the challenges in scaling up deep learning models and the susceptibility, to privacy attacks.
Qi et al. [18]	BFL – Prediction of Traffic Flow (TFP)	The convergence rate is slower due to the decentralized learning approach and model aggregation phase, in contrast to the SAE model's centralized learning method.
Rehman et al. [72]	BFL – Reputation-aware fine-grained	Prioritizing data privacy and integrity, this reputation-aware federated learning system safeguards sensitive information through secure and private communication channels.
Arachchige et al. [73]	PriModChain – Federated ML, Ethereum blockchain, Differential Privacy (DP) and Smart contracts.	Identify vulnerabilities and offer suggestions for security enhancements.
Wang et al. [74]	BEMA – The initialization of a system involving multiple parties and multiple classes, with off-chain sample extraction and on-chain mining processes.	Absence of assured resilience in the face of Byzantine attacks.
Alzubi et al. [75]	Utilizing Deep Learning and Blockchain Methods for Preserving Privacy in Electronic Health Records	Classifying users, integrating with cryptographic methods, and verifying client contributions before aggregating the model.
Salim et al. [76]	DP-BFL – Differential Privacy blockchain-driven explainable FL	Securing participants' privacy, preserving the global model's performance, and reducing the effects of harmful local updates.
Chen et al. [77]	ESB-FL – Blockchain-driven Federated learning system technique using Cryptography	Preserving the privacy of FL participants, upholding the accuracy of the global model with minimal communication expenses.
Liu et al. [19]	A permissioned Blockchain that supports Federated Learning, enabled with Privacy-Preservation, Multi-Party Computation, and Fully Homomorphic Encryption.	Safeguarding participant privacy, ensuring anonymity, and securely updating models using multi-party computation and comprehensive homomorphic encryption.
Qu et al. [28]	BFL – Digital Twin Networks (DTN)	Issues faced in digital twin networks encompass the concentration of processing tasks, the potential for data manipulation, breaches of privacy, and the absence of motivation systems.

earnings of genuine miners and exert a severe influence on the mining pool, eventually culminating in attacks on the pool mining system.

The decentralized functioning of blockchain enables FL to be fault-tolerant and can help to avoid attacks effectively. While verifying transaction data is crucial for consensus, it creates a gaping loophole for privacy violations [64]. With the data laid bare to all nodes, malicious actors could potentially exploit it for nefarious purposes [65].

1) DATA PRIVACY

BCFL aims to keep each node's data private during learning. However, achieving this is difficult as each node must share information with the other during the learning process.

Current research has suggested that differential privacy and homomorphic encryption can be alternatives to enable computation without decryption. However, these methods are computationally expensive and challenging to implement in a scalable way [77].

2) DE-ANONYMIZATION RISK

The risk of de-anonymization is heightened with small leaks and is particularly concerning for crypto transactions, which may contain sensitive information such as wallet addresses and transaction amounts. An approach to mitigating this risk is implementing differential privacy. However, there's a cost trade-off between accuracy and privacy budget [78].

Differential Privacy (DP) protects the privacy of individual data points by adding noise to the data. However, this noise can also reduce the accuracy of the learning model. The amount of noise that is added depends on the privacy budget, which is a parameter that controls the trade-off between privacy and accuracy. When the noise was added to the data with differential privacy, it became more difficult for the model to learn the true relationship between the data and the labels [79]. DP can reduce the accuracy of the learning model, especially when the privacy budget is high. The impact of differential privacy on the quality of the learning model depends on the specific task in the job creation workflow of the BCFL with the participating crypto entities and the amount of available data [24].

3) HOMOMORPHIC ENCRYPTION (FHE) DIFFICULTIES

Keeping user-related data private from participating crypto entities while using it for fraud detection can be challenging. While homomorphic encryption can protect data privacy during training, it can be computationally expensive and difficult [34]. Here are some of such difficulties:

- *Computational Complexity*: High computational expense in Homomorphic Encryption (HE) algorithms makes them impractical for large-scale BCFL applications, significantly slowing down fraud detection model training and evaluation and hindering real-time performance crucial for prompt detection and prevention of fraudulent transactions [80].
- *Limited Expressiveness*: Current HE schemes support only basic operations like addition and multiplication, limiting the feasibility of training more complex models involving non-linear operations or requiring access to raw data features [25].
- *Key Management Overhead*: HE relies heavily on cryptographic keys to secure encryption and decryption. Securely managing these keys is a significant challenge in BCFL settings, with multiple nodes involved. Key distribution, revocation, and rotation must be carefully handled to prevent unauthorized access to sensitive data [81].

4) USER PRIVACY

Protecting user privacy within the context of data utilization for fraud detection is a fundamental challenge. It is essential to ensure the exclusive use of data from participating entities for fraud detection with strict provisions against unauthorized access and dissemination to third parties [13]. Technically, the potential for data privacy exploitation exists if entities within the federated system participate with the sole intent of acquiring data from other crypto entities to predict market trends or analyze the user behavior of competitors, among other things. In such a scenario, the efficacy of token-based incentivization may no longer serve as an effective incentive once this exploitation is detected [13].

5) HANDLING SENSITIVE DATA

To ensure data and user privacy are not compromised, it is essential to handle crypto transactions containing sensitive information effectively in a federated environment [35].

6) INTEGRITY OF THE LEARNING PROCESS

The integrity and reliability of the FL process can be compromised if privacy preserving mechanisms such as differential privacy that adds noise to the data influences the accuracy of the learning model [33].

7) POLICY AND REGULATION COMPLIANCE

Navigating the intricate complexities of varying data protection regulations and standards is a significant challenge in BCFL. This is because the evolving jurisdictional standards and regulations demand that all participating entities in the federated system comply with the latest mandates [15].

8) INFRASTRUCTURE VULNERABILITIES

Vulnerabilities in the physical and digital underpinnings of the BCFL system could be exploited by adversaries, posing privacy risks. For instance, attackers might compromise a node to steal data or disrupt the network, impeding communication between nodes [82].

B. ADDRESSING THE PRIVACY AND CYBERSECURITY CHALLENGES OF BCFL FOR CRYPTO FRAUD DETECTION

1) DATA PRIVACY

Several existing solutions presented in the literature address the challenges of data privacy, including Secure Multi-Party Computation (SMPC) [83], incorporating advanced encryption standards approach [34], and Federated Learning of Cohorts (FLoC) [84].

- *Secure Multi-Party Computation (SMPC)*: This cryptography method allows various nodes to collaboratively contribute to the learning process in total privacy, offering security during data exchange. In SMPC, the concept *Collaboration without Trust* allows each node to perform its computation on its data, revealing the final result afterward without exposing individual data. This is particularly useful in the crypto fraud detection

ecosystem, where trust between nodes may be limited. In addition, SMPC allows for the correlation of activities from different nodes to identify anomalies that can help signal potential fraud, as anomalous transactions in crypto environments typically indicate crypto fraud [83].

- *Incorporating Advanced Encryption Standards*: The deployment of sophisticated encryption standards, such as Elliptic Curve Cryptography (ECC), stands out for its ability to deliver a superior level of security with reduced computational demands, using more minor keys compared to alternative cryptographic methods. This efficiency is critical when dealing with large volumes of data and transactions in crypto fraud detection. The lightweight nature of ECC makes it better adapted for systems with many nodes participating in the model training process requiring privacy. It offers participating entities the assurance against unauthorized sharing or manipulation of sensitive data [34].
- *Federated Learning of Cohorts (FLoC)* This methodology clusters nodes that share similar attributes. Data analysis from these clustered entities ensures the safeguarding of personal information and promotes the model's training. The cohort-based training enhances speed and efficiency compared to individual node operations, facilitating rapid fraud detection. Also, it allows for the easy identification of fraudulent patterns and behaviors within cohorts of nodes with similar features. This proves instrumental in detecting subtle and complex crypto fraud patterns in BCFL [84], [85].

2) DE-ANONYMIZATION RISK

- *Privacy Budget Management*: Managing the privacy budget is a crucial aspect of DP. The privacy budget refers to the cumulative privacy loss that a user is willing to tolerate. Spending the privacy budget wisely can help balance privacy and utility. The careful management of the privacy budget is necessary to balance privacy and accuracy optimally. A lower privacy budget minimizes the noise added to data but compromises privacy and vice versa. Showing a proactive effort to manage privacy budgets can build user trust in the system, making participating entities more willing to contribute their data for learning, resulting in better fraud detection outcomes [86].
- *Introduction of Multiplicative Noise in DP*: Instead of additive noise, multiplicative noise can sometimes decrease the risk of information leakage while ensuring privacy. Although this is not commonly used, with adequate interpretation capabilities, models can detect anomalies even in data modified by multiplicative noise. Such anomalies are an indicator of fraudulent transactions. DP typically involves the addition of noise, often from a Laplace or Gaussian distribution, to the results of a query. The amount of noise added is proportional to the sensitivity of the function (i.e., the maximum amount that any single individual could change the function's output) and inversely proportional to the desired privacy [23].
- *Enhanced Differential Privacy Techniques*: Advanced versions of differential privacy like Local Differential Privacy (LDP) [24] or Renyi Differential Privacy (RDP) [79] can provide better data privacy while preserving accuracy. LDP adds noise to individual data points rather than the output of a function over the dataset, while RDP provides tighter privacy guarantees for multiple computations. RDP provides a method that is both operationally efficient and quantitatively precise for monitoring the cumulative loss of privacy during the implementation of an individual differentially private mechanism, as well as across multiple such mechanisms [79].
- *Anomaly Detection Models*: Models for detecting anomalous transactions can be developed. For example, an unsupervised learning Expectation Maximization (EM) algorithm can cluster the dataset, and anomaly detection can be performed using Random Forest [87].

3) HOMOMORPHIC ENCRYPTION DIFFICULTIES

Here, we present some solutions that can help better crypto fraud detection. They enable secure and efficient utilization of distributed data in BCFL and provide a way to manage trade-offs between privacy, computational efficiency, and scalability in fraud detection systems [88].

- *Hardware Acceleration Techniques*: Leverage high-performance computing or parallel processing techniques to mitigate the high processing needs of HE and improve system performance [89].
- *Hybrid Solutions*: Combine non-HE local model training with HE-protected model aggregation. This approach leverages the privacy-preserving aspect of HE where needed while limiting its computational burden [90].
- *Layered Approach*: Implementing a layered security architecture where HE is used only for susceptible data, thus reducing its computation needs [86].
- *Function Secret Sharing*: FSS is a cryptographic primitive that evaluates specific functions without disclosing raw data, which could be an alternative to HE for complex operations [23].
- *HE-based Approximation Techniques*: Utilize HE-based approximation techniques, such as Chebyshev polynomial approximation of piecewise linear approximation [91], to approximate complex functions and enable their evaluation within the HE environment.

C. SCALABILITY AND PERFORMANCE CHALLENGES OF BCFL FOR CRYPTO FRAUD DETECTION

This section presents the major scalability and performance challenges of BCFL for crypto fraud detection.

1) COMPUTATIONAL POWER AND ENERGY EXPENDITURE

Blockchain technologies, specifically those utilizing a proof-of-work consensus mechanism such as Bitcoin, necessitate

substantial computational resources and energy consumption, potentially posing challenges to scalability and performance. As the number of participating entities in the federated system grows, this would lead to increased costs and environmental concerns [92].

2) DATA STORAGE

The need to store all transaction data on each node of a blockchain system can present significant storage challenges as the volume of transactions scales. It becomes impractical for resource-constrained entities participating in the model training [93].

3) MODEL LEARNING BOTTLENECK

When the model's updates propagate across the nodes for aggregation, it could lead to notable delays and inefficiencies; and consequently, a bottleneck limits the scalability of the system [94]. This can limit the system's ability to respond to emerging threats in real-time in the following ways.

- *Varied Computation Times*: Nodes may have different computational capabilities, leading to variations in the time it takes to generate model updates [81].
- *Communication Overhead*: The need to exchange large amounts of model update data between nodes can lead to network congestion and delays [27].

4) DATA HETEROGENEITY

Merging different types of datasets from diverse sources with varying quality and formats could pose challenges for FL models. This data heterogeneity may result in less accurate fraud detection models. It could limit the system's ability to recognize new fraud patterns, leading to inaccurate fraud detection models [27]. Some features of data heterogeneity that affect the federated process include:

- *Temporal Variations*: The inherent characteristics of data can change over time due to factors such as device upgrades and user behavior. This can make it challenging to train and update federated learning models consistently [3].
- *Data Quality*: Data collected from various sources may contain noise, missing values, or inconsistencies. This can lead to inaccurate model predictions and hinder fraud detection [69].

5) NON-UNIFORM BLOCK STRUCTURES

The absence of standardized block structures across different BCFL implementations poses interoperability challenges, complicating the exchange and processing of data among participating entities in the BCFL system. Standardization refers to establishing standard protocols and formats for data exchange and communication between different BCFL systems. Each system may have its own proprietary block structure without standardization, preventing seamless integration and interoperability [61]. This fragmentation could impede broader acceptance by crypto exchanges and financial institutions in two ways:

- *Compatibility*: Incompatibility among different block structures makes it difficult for BCFL systems to exchange data and model updates [95].
- *Integration Complexity*: Complexity in integrating different BCFL systems as each system has its own unique data formats and communication protocols [96].

6) SUSCEPTIBILITY TO ADVERSARIAL THREATS

As the number of nodes in the BCFL network increases, the system becomes increasingly vulnerable to adversarial attacks that aim to manipulate or disrupt the consensus process. Such attacks can potentially compromise the fraud detection system's integrity, reliability, and scalability [53].

7) NETWORK LATENCY

Communication delays are expected in a growing network with the addition of more nodes to a blockchain. This has the potential to slow down communication and is particularly challenging for real-time fraud detection [47].

8) CHALLENGES IN ACHIEVING CONSENSUS

With the growing number of nodes, verifying and validating transaction updates and model parameters from every other node in the network in reaching consensus becomes computationally demanding and complex. This increased complexity leads to potential disruptions and delays in the fraud detection process [68]. Some factors contribute to the difficulty in achieving consensus in large-scale BCFL networks include:

- *Byzantine Fault Tolerance:* BCFL systems must tolerate Byzantine faults, where nodes fail or behave maliciously. This requires sophisticated consensus mechanisms that maintain system integrity even in the presence of faults [22].
- *Increased Communication Overhead:* As the number of nodes grows, the communication overhead between nodes increases due to the need to exchange transactions and model update data. This can lead to network congestion and delays in message propagation [92].
- *Limited Node Capacity:* Each node has limited computational and storage resources, which can constrain its ability to process and verify large amounts of data. This can lead to bottlenecks and slow down the consensus process [93].

9) MANAGING PROOFS

Verifying the authenticity of transactions and preventing double-spending in a scalable manner is technically challenging with BCFL systems. Double-spending occurs when a user spends the same cryptocurrency multiple times. This is a critical issue in crypto fraud detection as it can be used to manipulate the system for fraudulent gains. For scalable systems to function optimally, addressing these

challenges is crucial to avoid vulnerabilities and fraud within the system [97].

10) CHALLENGES WITH REAL-WORLD IMPLEMENTATION

Real-world implementation of blockchain-based federated learning systems, particularly in detecting cryptocurrency fraud, faces several practical challenges [63], [89], [98]. These can be broadly classified into network-related issues [99], hardware constraints [89], and participant computational resource discrepancies [92]. Addressing these issues is pivotal for the system's success in real-world applications [6].

11) NETWORK CONNECTIVITY CHALLENGES

Achieving stable and consistent network connectivity is crucial for BFL, as intermittent connections can hinder the training process and impede the synchronization of model updates [100]. Potential strategies to address these challenges include:

- Incorporating resilience mechanisms, such as robust peer-to-peer networking protocols and dynamic node discovery [2].
- Network redundancy should be employed to ensure alternative pathways for data transmission and enhance resilience [101].
- Edge computing is utilized for local data processing, decreasing reliance on continuous network connectivity [102].
- Implement local caching of updates to safeguard against data loss during connectivity disruptions and synchronize updates upon connection restoration or timed cache deletion [103].
- Using decentralized storage solutions, like distributed file systems or content delivery networks, can help mitigate the impact of network failures on the availability of model updates and training data [10].

12) HARDWARE CONSTRAINTS

The variability in hardware capabilities among participants can result in bottlenecks that slow down the federated learning process [98]. Strategies to mitigate hardware limitations may include:

- Establishing minimum hardware requirements for participation in the federated learning network [89].
- Adaptive learning approaches that tailor model training tasks to the hardware capabilities of individual participants [81].
- Cloud-based virtualization solutions allow participants to access standardized virtual hardware, ensuring consistent performance across the network [74].

13) VARIATION IN PARTICIPANT COMPUTATION CAPABILITIES

Differences in computational power among participants may create imbalances in contributions to the federated model, which could impact the learning process [100]. Addressing this challenge may involve:

- Deploying algorithms that account for variations in computing power, ensuring equitable weighting of updates regardless of the source's computational strength [55].
- Employing resource-aware task assignments to align model training tasks with participants' computational capacities [42].
- Promoting asynchronous updates, allowing participants to update the model at their own pace and preventing more capable participants from overwhelming the training process [60].

D. ADDRESSING THE SCALABILITY AND PERFORMANCE CHALLENGES OF BCFL FOR CRYPTO FRAUD DETECTION

While several research works have been conducted in this domain [2], [9], [101], [104], it is essential to highlight how combining some of these proposed solutions and techniques gives a decent solution to the scalability and performance challenges experienced in the BCFL system.

1) DATA STORAGE CHALLENGES

- *Data Pruning*: Data pruning techniques can reduce the storage requirements of nodes in a blockchain. These techniques encompass the removal of obsolete or redundant data from the blockchain. One such method involves eliminating old transactions from participating crypto entities once they have been finalized and integrated into the blockchain; their details can be deleted from individual nodes, thereby reducing storage overhead. Another approach is the elimination of redundant data. By identifying and eliminating identical or duplicate data, the overall size of the blockchain can be reduced. Also, purging inactive accounts by archiving or removing data linked with dormant or inactive accounts contributes to the conservation of storage space [105].
- *Sharding*: Due to the partitioning that occurs during sharding, nodes will only store relevant portions of the data, thereby reducing the overall storage requirements for each node [106].
- *Distributed Storage*: Distributed storage solutions such as Interplanetary File System (IPFS) and Swarm can share data across multiple nodes in the federated process such that storage requirements for individual crypto entities participating are reduced [93].
- *State Channels*: State channels are off-chain mechanisms allowing direct and private transactions between parties without broadcasting every transaction to the main blockchain. This can significantly reduce the amount of data stored on the blockchain, as only the final state of the channel is recorded on-chain. However, this approach has concerns about security constraints [99].

2) LEARNING BOTTLENECKS ON MULTIPLE DEVICES

Delay issues that arise during the aggregation process in BCFL can be minimized by implementing more efficient

aggregation algorithms or protocols that aggregate model updates and parameters faster [9]. Also, implementing the Stochastic Gradient Descent (SGD) technique [12], which processes smaller subsets of data independently, can help speed up model updates and their subsequent aggregations. A combination of data pruning and SGD allows for efficient utilization of resources in the learning process so available nodes efficiently execute ongoing computations [107].

Strategies for workload allocation could be developed to address the issue of varying computational times across nodes [108]. These strategies would distribute tasks in accordance with the computational capacity of each node. Additionally, implementing machine learning algorithms that accommodate the heterogeneity among devices or nodes could prove advantageous [65], [102].

To mitigate the issues of communication overhead, techniques such as gradient compression can be employed [12]. This decreases the volume of data exchanged between nodes during the updating of models. Additionally, using an asynchronous communication protocol, where model updates are transmitted and received independently of synchronization with all nodes in the network, can be beneficial. This approach can aid in reducing network congestion and delays, thereby enhancing the efficiency of the learning process [100].

The application of incremental learning can enhance the efficiency of the learning process. Incremental learning in BCFL is a strategy that facilitates continuous adaptation and improvement of the learning model as new data becomes available. It is a beneficial approach to the federated process with the decentralization and constant update of data [30], [109].

A recent study proposed a blockchain-based differential optimization federated incremental learning algorithm, which integrates incremental learning into the federated learning framework [109]. This approach was demonstrated to not only improve the accuracy of the model but also reduce the risk of data leakage stemming from gradient updates. Incremental learning also reduces resource consumption by eliminating the need for complete model retraining with each new data instance, a critical feature in crypto transactions where new data is constantly available. It enables real-time fraud detection and mitigates the challenge of catastrophic forgetting, a phenomenon where a model loses previously learned information when new data is encountered [103].

3) NON-UNIFORM BLOCK STRUCTURE

- *Data Governance Frameworks*: Establishing data governance frameworks can facilitate the integration of diverse BCFL systems. These frameworks can provide data security, privacy, and ethical use guidelines, thereby fostering broader adoption by a larger pool of participating crypto entities [110].
- *Standardization*: Establishing standard protocols and formats for data exchange and communication between different BCFL systems can ensure the interoperability of participating crypto entities. This would involve creating a standardized block structure that all BCFL systems adhere to, facilitating seamless integration and data exchange [63].
- *Integration Tools & Interoperability Protocols*: Implementing integration tools and interoperability protocols or standards can significantly reduce integration complexity associated with incorporating different data formats and BCFL systems. This could involve settling on a standard data format and communication protocol, enabling effective communication and data exchange between systems [61].
- *Compatibility Measures*: Creating middleware or interface systems capable of translating and interpreting diverse block structures can be a viable solution for compatibility challenges. These systems would serve as intermediaries that allow different BCFL systems to communicate with each other. Translating the data and model updates into a standard format enables an effective exchange of information despite the variations in block structures across different systems. This can mitigate compatibility issues, facilitate system integration, and enhance overall scalability [111].

4) SUSCEPTIBILITY TO ADVERSARIAL THREATS

- *Enhanced Security Measures*: Incorporating robust security measures like cryptographic techniques and implementing intrusion detection tools can offer protection against adversarial threats early [112].
- *Byzantine Fault Tolerance Mechanisms*: Byzantine fault tolerance mechanisms can enhance system resilience against adversarial attacks by ensuring system integrity even if some nodes behave maliciously [22].
- *Reputation Based Systems*: The deployment of a system that assigns credibility ratings to nodes, grounded in their historical conduct, can aid in pinpointing and isolating harmful nodes before they inflict damage [9].

5) NETWORK LATENCY

Efficient and optimized communication protocols are needed to minimize the network latency [47]. BCFL latency optimization demands a holistic framework that jointly optimizes data offloading, node transmit power, channel bandwidth allocation, node computation, and hash power allocation [47]. It is essential also to understand the specific needs in the BCFL decentralized network to select the protocols that best fit, such as asynchronous updates or innovative edge computing techniques [113]. Some protocols, such as QUIC, have reduced latency compared to TCP, making it useful for BCFL systems with real-time efficiency requirements [81].

6) CHALLENGES IN ACHIEVING CONSENSUS

- *Adoption of Scalable Consensus Protocols*: Implementing protocols such as the Practical Byzantine Fault

TABLE 5. A comparison of the BCFL architectures for crypto fraud detection.

Architecture	Description	Pros	Cons	Best Scenario
Fully Integrated BCFL	Tightly integrated architecture where training and mining tasks are combined on the client side.	Seamless data sharing workflows	Limited flexibility	Best for data sharing workflows and handling the computational load combining client-side training and mining tasks [99].
Flexible Integrated BCFL	Adjustable integration between blockchain and federated learning components based on requirements.	Adaptability to specific needs	Increased complexity	Best used if the system needs to adjust the degree of integration between the blockchain and FL components based on particular requirements [99].
Loosely Integrated BCFL	Blockchain and federated learning components communicate through an intermediary messaging engine or user interface.	Suitable for systems with inherent challenges	Potential communication overhead and security concerns.	Preferred in systems with inherent data sharing challenges and practical when minimizing client computational load by separating training and mining tasks is necessary [99].

Tolerance, Delegated Proof of Stake, or Sharding can be beneficial as these protocols are designed to ensure quick consensus despite the growing number of nodes [92].

- *Limited Node Capacity*: Solutions like greedy algorithms for node selection and discarding nodes with poor communication conditions can equally help improve overall efficiency [101].
- Dealing with Byzantine Fault Tolerance Mechanisms with protocols like PBFT Protocol or Federated Byzantine Agreement can ensure the system continues functioning correctly even if some nodes fail or behave maliciously [114].

7) MANAGING PROOFS

- *Implementation of Transaction Validating Protocols*: Protocols such as Two-Phase Commit Protocol (2PC), Atomic Commit Protocol (ACP), and Three-Phase Commit(3PC) [115] ensure the authenticity and consistency of operations between participating crypto entities in the BCFL system. 2PC involves a coordinator node sending a commit or abort request to participating nodes. Nodes respond with agreement or disagreement; if all agree, the transaction is committed; otherwise, it is aborted. Conversely, ACP is designed for distributed networks and guarantees that a series of changes is dedicated to all nodes successfully, or not at all. 3PC is an improvement of 2PC and adds a 'prepared commit' phase before the commit/abort phase to reduce the risk of data inconsistency in the event of network failure [112].
- *Implementation of Zero-Knowledge Proofs*: The application of Zero-Knowledge Proof-based Federated Learning (ZKP-FL) strategies [97] allows for the verification of computational processes without needing raw data [114]. This significantly reduces the computational load associated with proof management. Specific strategies, such as the Practical ZKP-FL (PZKP-FL), support fractional and non-linear operations that might occur in the nodes of participating entities. Performance analyses of the PZKP-FL scheme indicate that the total runtime is less than a minute when executed in parallel.

VI. DISCUSSION AND CONCLUSION

This paper effectively highlights the potential of leveraging BCFL to detect and mitigate cryptocurrency fraud. It underscores its pivotal role in enhancing data privacy, bolstering reliability, and ensuring scalability within the realm of crypto fraud detection. By analyzing the most recent advancements and existing gaps in this domain, it is clear that the decentralization, privacy preservation, and scalability attributes of the BCFL approach make it a promising strategy for tackling the mounting issues of cryptocurrency fraud. Most critical contributions from the studies reviewed point to the importance of ensuring data privacy, dealing with scalability issues, reducing the risk of inference and malicious attacks, and managing decentralized communication.

In summary, various BCFL architectures for crypto fraud detection have been presented in literature [37], falling broadly into three categories: (1) fully integrated BCFL, (2) flexibly integrated BCFL and (3) loosely integrated BCFL. Table 5 compares these architectures in terms of pros, cons, and the best scenarios when they should be employed in the industry. Each BCFL architecture is designed to incorporate a learning model that leverages diverse data sources, including transaction data, user data, and device data. This integration facilitates the identification of specific fraudulent transaction patterns.

The type of crypto fraud targeted during the job creation phase of the BCFL workflow, coupled with the resources available to the entities participating in the federated system, largely dictates the most appropriate architecture to adopt. This ensures that computational resources undergo thorough analysis before deployment. In certain scenarios, customization specific to the system may be required to optimize performance [115].

The BCFL workflow phases of crypto fraud detection extend over multiple stages, each of which contributes to developing and implementing an effective predictive model capable of identifying crypto fraud [42]. From establishing a task to deploying the updated global model, every step is instrumental in ensuring a robust, scalable, and secure crypto fraud detection system [52].

Further elaborations on the existing issues in BCFL, mainly stemming from privacy and scalability, are needed for further studies. In particular, the scalability challenge is displayed in several forms, such as computational power requirements, energy expenditure, data storage challenges, and learning bottlenecks, all of which can constrain the efficiency and effectiveness of the BCFL system. Also, privacy threats like inference attacks remain a concern due to the exposure of sensitive data during the consensus process. Additionally, the BCFL system is critically vulnerable to adversarial threats aiming to manipulate or disrupt the consensus process, which can compromise the fraud detection system's integrity, reliability, and scalability.

Generally, a mash-up of the different techniques and approaches presented in this study could offer scalable and privacy solutions and potentially significantly impact fraud detection in the BCFL system. Additionally, ML approaches can work hand-by-hand with BCFL to mitigate crypto fraud. However, the scarcity of genuine financial transaction data impedes the development of modern ML models for detecting fraud in crypto transactions, which remains a grand challenge to achieve this objective.

In summary, this paper presented a comprehensive analysis of various research studies and methodologies that underscore the substantial potential of BCFL in mitigating fraudulent activities in cryptocurrency transactions. Despite the challenges posed by privacy concerns, scalability constraints, and threats from adversarial attacks, the inherent benefits of decentralization, security, and privacy preservation that BCFL offers make it a promising solution in dealing with crypto fraud.

Incorporating blockchain not only decentralizes the learning system but also provides transparency, immutability, and a high degree of security, which are critical in preventing crypto fraud. To achieve scalability and privacy in BCFL, a combination of various proposed solutions has the potential to enhance crypto fraud detection significantly. However, the initial step involves the establishment of a data governance framework and standardization, which will serve as a guide for interoperability and integration for multiple participating entities interested in model training in the federated system. This sets the tone and comforts crypto companies in participating in the training. While several solutions have been proposed in this study, a geographically distributed cloud computing model that utilizes SMPC [83] and lightweight consensus algorithms and protocols will be a beneficial and recommended approach. This model effectively manages computational overheads and privacy, enabling the federated system to scale.

There is still a need to develop and implement protocols for managing sensitive data in compliance with global privacy standards, such as the General Data Protection Regulation (GDPR) and local regulatory requirements for contributing entities. The integration of advanced cryptographic techniques, including homomorphic encryption, enables secure computation on encrypted data, ensuring the confidentiality and integrity of information throughout the federated learning process.

The success of BCFL for crypto fraud detection hinges on the effective coordination and integration of multiple participating entities, each with its own data, infrastructure, and objectives. Establishing a robust data governance framework and standardization is crucial to facilitate seamless interoperability and integration. This data governance framework could be collaboratively developed by the participating crypto entities, regulators, and other stakeholders to ensure alignment with industry best practices and regulatory requirements. The proposed framework should consider the following aspects:

- *Data Sharing Agreements*: Development of standardized agreements that define the terms of data sharing, including permissions, responsibilities, and mechanisms for dispute resolution.
- *Data Ownership and Sovereignty*: It is essential to establish explicit ownership and control protocols for the data provided by each participating entity to maintain data sovereignty throughout the federated learning process.
- *Data Security and Privacy*: Given the sensitive nature of cryptocurrency transactions, the framework should incorporate stringent security measures to protect data from unauthorized access and breaches. This may include encryption and anonymization techniques to protect user privacy and ensure the confidentiality of transactional data.
- *Compliance and Audit Trails*: Establish mechanisms for monitoring data origin and ensuring compliance with regulatory standards. This includes using blockchain technology to maintain immutable audit trails, thereby facilitating transparency and accountability throughout the data lifecycle.
- *Data Quality and Standardization*: Defining data quality standards, formats, and validation procedures to ensure the reliability and consistency of the data used for model training across the federated system across participating entities.
- *Governance Structure and Decision-Making*: Defining the governance structure, roles, and decision-making processes for managing the overall federated learning system, including model updates, conflict resolution, and system maintenance.
- *Standardization for Interoperability*: Beyond the data governance framework, implementing industry-wide standardization is essential to facilitate smooth interoperability and integration among the varied participants in the blockchain-federated learning ecosystem globally. Financial institutions, such as banks, can join as participants in the model training process, making it imperative to establish standards that enable efficient interoperability.
- *Protocol and Interface Standards*: Defining common communication and data exchange protocols that ensure

seamless interactions between different blockchain platforms and federated learning systems is essential. One such protocol could be obtained from the IEEE Standards for Blockchain-based Federated Machine Learning. The IEEE has proposed several standards for blockchain-based federated machine learning. For example, P3127 is a guide for a blockchain-based federated machine learning architectural framework. Other standards include P3227 for a reference framework of data security circulation system based on blockchain and federated computation and P3229 for industrial software applications based on blockchain.

- *Model Exchange Formats*: Standards can be established for the representation and exchange of ML models, facilitating the sharing of models among entities without compatibility challenges.
- *Blockchain Integration and Consensus Mechanisms*: Defining standard approaches for integrating the federated learning system with the underlying blockchain infrastructure, including the consensus protocol, transaction formats, and smart contract specifications.
- *Auditing and Compliance*: Developing standardized auditing and compliance frameworks to ensure the overall blockchain-federated learning system's transparency, accountability, and regulatory adherence. For instance, the VFChain is a verifiable and auditable federated learning framework based on the blockchain system that can be implemented. It addresses the problem of designing a secure federated learning framework to ensure the correctness of the training procedure.
- *Standardization*: Standardization and interoperability are critical elements in ensuring the seamless integration of systems across various entities, especially within the financial industry, which includes cryptocurrency entities. By aligning with existent frameworks such as the Financial Information eXchange protocol(FIX), which is widely used for the communication of financial information, blockchain federated learning systems can ensure a common language and set of processes for information sharing and transactions. With the FIX protocol, the integration complexity between disparate blockchain and federated learning systems is significantly reduced, as they can utilize a standardized method of data transmission. The advantage of the reduced learning curve is that entities do not need to invest as heavily in training staff or developing custom interfaces for their systems, as the FIX protocol provides a universally recognized standard for financial communication.
- *Interoperability*: By establishing robust data governance frameworks and industry-wide standardization, the blockchain-federated learning system for crypto fraud detection can achieve high interoperability, integration, and trust among the participating entities. This, in turn, will enable the effective and coordinated deployment of the system, ultimately enhancing its scalability, security, and real-world impact in combating financial crimes in the cryptocurrency and DeFi ecosystem.
- *Geographically Distributed Cloud Computing*: A geographically distributed cloud computing infrastructure can efficiently scale to accommodate the growing amount of data and computational demand from various entities. This will potentially support more significant, more complex federated learning models and larger numbers of transactions. By distributing the computational resources and storage across multiple cloud data centers regionally, the system can better handle the increasing volume of data and model updates from the participating cryptocurrencies or participating entities.
- *Performance Enhancement*: By strategically locating cloud resources closer to the data source or end-users, latency can be reduced, leading to faster processing times for model updates and improved overall system performance. Also, the geographic distribution of the cloud nodes can enhance the resilience and fault tolerance of the system, as the failure or compromise of a single node will have a limited impact on the overall operation.
- *Secure Multi-Party Computation (SMPC)*: SMPC is specifically designed to allow entities to engage in computation without exposing their private data. In the context of federated learning, this means that cryptocurrency entities can contribute to a shared model without risking sensitive transactional data.
- *Off-chain Computation Techniques*: The leading blockchain network minimizes the computational burden by utilizing off-chain mechanisms, such as sidechains or state channels. This improves transaction throughput and reduces the costs associated with on-chain operations.

In conclusion, integrating geographically distributed cloud computing and secure multi-party computation within a blockchain-federated learning system presents a promising approach to enhance scalability, performance, and privacy. This combination can create a more resilient and efficient infrastructure for handling the computational and data demands of blockchain-federated learning systems, which are crucial for the secure and collaborative model training processes desired by cryptocurrency entities.

ACKNOWLEDGMENT

Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect NSF's views.

REFERENCES

[1] A. Mashatan, M. S. Sangari, and M. Dehghani, "How perceptions of information privacy and security impact consumer trust in crypto-payment: An empirical study," *IEEE Access*, vol. 10, pp. 69441–69454, 2022.

[2] A. F. M. S. Shah, M. A. Karabulut, A. F. M. S. Akhter, N. Mustari, A. K. Pathan, K. M. Rabie, and T. Shongwe, "On the vital aspects and characteristics of cryptocurrency—A survey," *IEEE Access*, vol. 11, pp. 9451–9468, 2023.

[3] P. Durgapal, P. Kataria, G. Garg, and A. S. Anand, "A comprehensive distributed framework for cross-silo federated learning using blockchain," in *Proc. 5th Int. Conf. Blockchain Comput. Appl. (BCCA)*, Oct. 2023, pp. 538–545.

[4] A. P. Kalapaaking, I. Khalil, M. S. Rahman, M. Atiquzzaman, X. Yi, and M. Almashor, "Blockchain-based federated learning with secure aggregation in trusted execution environment for Internet-of-Things," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1703–1714, Feb. 2023.

[5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist. (PMLR)*, Fort Lauderdale, FL, USA, 2017, pp. 1273–1282.

[6] A. A. Ahmed and N. Okoroafor, "An ML-powered risk assessment system for predicting prospective mass shooting," *Computers*, vol. 12, no. 2, p. 42, Feb. 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/2/42

[7] J. P. Albrecht, "How the GDPR will change the world," *Eur. Data Protection Law Rev.*, vol. 2, no. 3, pp. 287–289, 2016.

[8] B. Mustain, M. Sk Tanzir, A. Adnan, R. Ziaur, and I. Rafiqul, "A systematic literature review on blockchain enabled federated learning framework for Internet of Vehicles," 2022, *arXiv:2203.05192*.

[9] J. Qi, F. Lin, Z. Chen, C. Tang, R. Jia, and M. Li, "High-quality model aggregation for blockchain-based federated learning via reputation-motivated task participation," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18378–18391, Oct. 2022.

[10] Y. Qu, L. Gao, T. H. Luan, Y. Xiang, S. Yu, B. Li, and G. Zheng, "Decentralized privacy using blockchain-enabled federated learning in fog computing," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5171–5183, Jun. 2020.

[11] A. M. A. Mo'men, H. S. Hamza, and I. A. Saroit, "A survey on security enhanced multicast routing protocols in mobile ad hoc networks," in *Proc. 7th Int. Symp. High-Capacity Opt. Netw. Enabling Technol.*, Dec. 2010, pp. 262–268.

[12] H. A. Madni, R. M. Umer, and G. L. Foresti, "Blockchain-based swarm learning for the mitigation of gradient leakage in federated learning," *IEEE Access*, vol. 11, pp. 16549–16556, 2023.

[13] R. Xu, C. Li, and J. Joshi, "Blockchain-based transparency framework for privacy preserving third-party services," *IEEE Trans. Depend. Secure Comput.*, vol. 20, no. 3, pp. 2302–2313, Jun. 2023.

[14] A. M. A. Mo'men, H. S. Hamza, and I. A. Saroit, "New attacks and efficient countermeasures for multicast AODV," in *Proc. 7th Int. Symp. High-Capacity Opt. Netw. Enabling Technol.*, Dec. 2010, pp. 51–57.

[15] Z. Wang, Q. Hu, R. Li, M. Xu, and Z. Xiong, "Incentive mechanism design for joint resource allocation in blockchain-based federated learning," *IEEE Trans. Parallel Distrib. Syst.*, vol. 34, no. 5, pp. 1536–1547, May 2023.

[16] Q. Hu, Z. Wang, M. Xu, and X. Cheng, "Blockchain and federated edge learning for privacy-preserving mobile crowdsensing," *IEEE Internet Things J.*, vol. 10, no. 14, pp. 12000–12011, Sep. 2023.

[17] Z. Lian, Q. Zeng, W. Wang, T. R. Gadekallu, and C. Su, "Blockchain-based two-stage federated learning with non-IID data in IoMT system," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 4, pp. 1701–1710, Nov. 2023.

[18] Y. Qi, M. S. Hossain, J. Nie, and X. Li, "Privacy-preserving blockchain-based federated learning for traffic flow prediction," *Future Gener. Comput. Syst.*, vol. 117, pp. 328–337, Apr. 2021.

[19] J. Liu, X. He, R. Sun, X. Du, and M. Guizani, "Privacy-preserving data sharing scheme with FL via MPC in financial permissioned blockchain," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.

[20] Y. Wan, Y. Qu, L. Gao, and Y. Xiang, "Privacy-preserving blockchain-enabled federated learning for B5G-driven edge computing," *Comput. Netw.*, vol. 204, Feb. 2022, Art. no. 108671.

[21] Y. Zhao, J. Zhao, L. Jiang, R. Tan, D. Niyato, Z. Li, L. Lyu, and Y. Liu, "Privacy-preserving blockchain-based federated learning for IoT devices," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1817–1829, Feb. 2021.

[22] Y. Miao, Z. Liu, H. Li, K. R. Choo, and R. H. Deng, "Privacy-preserving Byzantine-robust federated learning via blockchain systems," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2848–2861, 2022.

[23] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4177–4186, Jun. 2020.

[24] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3454–3469, 2020.

[25] N. Baracaldo and H. Shaul, "Federated learning meets homomorphic encryption," IBM Res. Blog, New York, NY, USA, 2022. [Online]. Available: https://research.ibm.com/blog/federated-learning-meets-homomorphic-encryption

[26] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[27] J. Kone ný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.

[28] Y. Qu, L. Gao, Y. Xiang, S. Shen, and S. Yu, "FedTwin: Blockchain-enabled adaptive asynchronous federated learning for digital twin networks," *IEEE Netw.*, vol. 36, no. 6, pp. 183–190, Nov. 2022.

[29] S. Saha and T. Ahmad, "Federated transfer learning: Concept and applications," *Intelligenza Artificiale*, vol. 15, no. 1, pp. 35–44, Jul. 2021.

[30] J. Dong, L. Wang, Z. Fang, G. Sun, S. Xu, X. Wang, and Q. Zhu, "Federated class-incremental learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10154–10163.

[31] A. Brecko, E. Kajati, J. Koziorek, and I. Zolotova, "Federated learning for edge computing: A survey," *Appl. Sci.*, vol. 12, no. 18, p. 9124, Sep. 2022.

[32] A. M. A. Moamen and H. S. Hamza, "On securing atomic operations in multicast AODV," *Ad-Hoc Sensor Wireless Netw.*, vol. 28, no. 8, pp. 137–159, 2015.

[33] A. Qammar, A. Karim, H. Ning, and J. Ding, "Securing federated learning with blockchain: A systematic literature review," *Artif. Intell. Rev.*, vol. 56, no. 5, pp. 3951–3985, May 2023.

[34] A. Madi, O. Stan, A. Mayoue, A. Grivet-Sébert, C. Gouy-Pailler, and R. Sirdey, "A secure federated learning framework using homomorphic encryption and verifiable computing," in *Proc. Reconciling Data Anal., Autom., Privacy, Secur., Big Data Challenge (RDAAPS)*, May 2021, pp. 1–8.

[35] Q. Lin, K. Xu, Y. Huang, F. Yu, and X. Wang, "Privacy-enhanced data fusion for federated learning empowered Internet of Things," *Mobile Inf. Syst.*, vol. 2022, pp. 1–8, Oct. 2022.

[36] Q. Li, Z. Wen, and B. He, "Practical federated gradient boosting decision trees," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, 2020, pp. 4642–4649.

[37] Z. Wang, Q. Hu, M. Xu, Y. Zhuang, Y. Wang, and X. Cheng, "A systematic survey of blockchained federated learning," 2021, *arXiv:2110.02182*.

[38] I. Ullah, X. Deng, X. Pei, P. Jiang, and H. Mushtaq, "A verifiable and privacy-preserving blockchain-based federated learning approach," *Peer Peer Netw. Appl.*, vol. 16, no. 5, pp. 2256–2270, Sep. 2023.

[39] S. Yuan, B. Cao, Y. Sun, Z. Wan, and M. Peng, "Secure and efficient federated learning through layering and sharding blockchain," 2021, *arXiv:2104.13130*.

[40] H. Guo, C. Meese, W. Li, C.-C. Shen, and M. Nejad, "B2SFL: A bi-level blockchained architecture for secure federated learning-based traffic prediction," *IEEE Trans. Services Comput.*, vol. 16, no. 6, pp. 4360–4374, Nov. 2023.

[41] N. F. Aurna, M. D. Hossain, Y. Taenaka, and Y. Kadobayashi, "Federated learning-based credit card fraud detection: Performance analysis with sampling methods and deep learning algorithms," in *Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR)*, Jul. 2023, pp. 180–186.

[42] U. D. of Justice. *Crypto Enforcement*. Accessed: Jan. 5, 2024. [Online]. Available: https://www.justice.gov/criminal/criminal-fraud/mimf/crypto-enforcement

[43] B. Liu, N. Lv, Y. Guo, and Y. Li, "Recent advances on federated learning: A systematic survey," 2023, *arXiv:2301.01299*.

[44] A. A. Moamen and N. Jamali, "An actor-based middleware for crowd-sourced services," *ICST Trans. Mobile Commun. Appl.*, vol. 3, no. 8, Sep. 2017, Art. no. 153070.

[45] X. Liang, Y. Liu, J. Luo, Y. He, T. Chen, and Q. Yang, "Self-supervised cross-silo federated neural architecture search," 2021, *arXiv:2101.11896*.

[46] Y. Cheng, Y. Liu, T. Chen, and Q. Yang, "Federated learning for privacy-preserving AI," *Commun. ACM*, vol. 63, no. 12, pp. 33–36, Nov. 2020, doi: 10.1145/3387107.

[47] D. C. Nguyen, S. Hosseinalipour, D. J. Love, P. N. Pathirana, and C. G. Brinton, "Latency optimization for blockchain-empowered federated learning in multi-server edge computing," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 12, pp. 3373–3390, Dec. 2022.

[48] S. Nakamoto. (Mar. 2019). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Cryptography Mailing List. [Online]. Available: https://metzdowd.com
[49] Z. Li, W. Wang, V. Ceikute, V. Kalogeraki, and D. Gunopulos, "Federated learning for decentralized anti-money laundering," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, 2020, pp. 1–3.
[50] R. M. Aziz, R. Mahto, K. Goel, A. Das, P. Kumar, and A. Saxena, "Modified genetic algorithm with deep learning for fraud transactions of Ethereum smart contract," *Appl. Sci.*, vol. 13, no. 2, p. 697, Jan. 2023.
[51] M. Zamani, M. Movahedi, and M. Raykova, "RapidChain: Scaling blockchain via full sharding," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 931–948.
[52] W. Li, B. Yang, and Y. Song, "Secure multi-party computing for financial sector based on blockchain," in *Proc. IEEE 14th Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Oct. 2023, pp. 145–151.
[53] P. Liu, X. Xu, and W. Wang, "Threats, attacks and defenses to federated learning: Issues, taxonomy and perspectives," *Cybersecurity*, vol. 5, no. 1, pp. 1–19, Dec. 2022.
[54] A. A. Ahmed, "A model and middleware for composable IoT services," in *Proc. Int. Conf. Internet Comput. (IoT)*, 2019, pp. 108–114.
[55] W. Huang, T. Li, D. Wang, S. Du, J. Zhang, and T. Huang, "Fairness and accuracy in horizontal federated learning," *Inf. Sci.*, vol. 589, pp. 170–185, Apr. 2022.
[56] V. Mugunthan, P. Goyal, and L. Kagal, "Multi-VFL: A vertical federated learning system for multiple data and label owners," 2021, *arXiv:2106.05468*.
[57] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, "A secure federated transfer learning framework," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 70–82, Jul. 2020.
[58] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. 3rd Mach. Learn. Syst. Conf.*, Aug. 2020, pp. 429–450.
[59] L. Ma, Q. Pei, Y. Qu, K. Fan, and X. Lai, "Decentralized privacy-preserving reputation management for mobile crowdsensing," in *Proc. Int. Conf. Secur. Privacy Commun. Syst.*, Orlando, FL, USA. Cham, Switzerland: Springer, Oct. 2019, pp. 532–548.
[60] G. Wang, S. Zhang, T. Yu, and Y. Ning, "A systematic overview of blockchain research," *J. Syst. Sci. Inf.*, vol. 9, no. 3, pp. 205–238, Jul. 2021.
[61] M. Al Asqah and T. Moulahi, "Federated learning and blockchain integration for privacy protection in the Internet of Things: Challenges and solutions," *Future Internet*, vol. 15, no. 6, p. 203, May 2023.
[62] H. Kim, J. Park, M. Bennis, and S.-L. Kim, "Blockchained on-device federated learning," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1279–1283, Jun. 2020.
[63] E. Goh, D.-Y. Kim, K. Lee, S. Oh, J.-E. Chae, and D.-Y. Kim, "Blockchain-enabled federated learning: A reference architecture design, implementation, and verification," 2023, *arXiv:2306.10841*.
[64] A. A. Ahmed, "A privacy-preserving mobile location-based advertising system for small businesses," *Eng. Rep.*, vol. 3, no. 11, pp. 1–15, Nov. 2021.
[65] D. Li, D. Han, T.-H. Weng, Z. Zheng, H. Li, H. Liu, A. Castiglione, and K.-C. Li, "Blockchain for federated learning toward secure distributed machine learning systems: A systemic survey," *Soft Comput.*, vol. 26, no. 9, pp. 4423–4440, May 2022.
[66] N. Wang, W. Yang, X. Wang, L. Wu, Z. Guan, X. Du, and M. Guizani, "A blockchain based privacy-preserving federated learning scheme for Internet of Vehicles," *Digit. Commun. Netw.*, vol. 10, no. 1, pp. 126–134, Feb. 2024.
[67] Z. Sun, J. Wan, L. Yin, Z. Cao, T. Luo, and B. Wang, "A blockchain-based audit approach for encrypted data in federated learning," *Digit. Commun. Netw.*, vol. 8, no. 5, pp. 614–624, Oct. 2022.
[68] Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng, and Q. Yan, "A blockchain-based decentralized federated learning framework with committee consensus," *IEEE Netw.*, vol. 35, no. 1, pp. 234–241, Jan. 2021.
[69] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A blockchained federated learning framework for cognitive computing in Industry 4.0 networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2964–2973, Apr. 2021.
[70] M. Shayan, C. Fung, C. J. Yoon, and I. Beschastnikh, "Biscotti: A blockchain system for private and secure federated learning," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 7, pp. 1513–1525, Jul. 2020.
[71] M. H. ur Rehman, K. Salah, E. Damiani, and D. Svetinovic, "Towards blockchain-based reputation-aware federated learning," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 183–188.
[72] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "A trustworthy privacy preserving framework for machine learning in industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6092–6102, Sep. 2020.
[73] Q. Wang, Y. Guo, X. Wang, T. Ji, L. Yu, and P. Li, "AI at the edge: Blockchain-empowered secure multiparty learning with heterogeneous models," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9600–9610, Oct. 2020.
[74] J. A. Alzubi, O. A. Alzubi, A. Singh, and M. Ramachandran, "Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 1080–1087, Jan. 2023.
[75] S. Salim, B. Turnbull, and N. Moustafa, "A blockchain-enabled explainable federated learning for securing Internet-of-Things-based social media 3.0 networks," *IEEE Trans. Computat. Social Syst.*, early access, Dec. 28, 2021, doi: 10.1109/TCSS.2021.3134463.
[76] B. Chen, H. Zeng, T. Xiang, S. Guo, T. Zhang, and Y. Liu, "ESB-FL: Efficient and secure blockchain-based federated learning with fair payment," *IEEE Trans. Big Data*, early access, May 23, 2022, doi: 10.1109/TBDATA.2022.3177170.
[77] B. Chhetri, S. Gopali, R. Olapojoye, S. Dehbashi, and A. S. Namin, "A survey on blockchain-based federated learning and data privacy," in *Proc. IEEE 47th Annu. Comput., Softw., Appl. Conf. (COMPSAC)*, Jun. 2023, pp. 1311–1318.
[78] J. Zhou, C. Hu, J. Chi, J. Wu, M. Shen, and Q. Xuan, "Behavior-aware account de-anonymization on Ethereum interaction graph," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 3433–3448, 2022.
[79] I. Mironov, "Rényi differential privacy," in *Proc. IEEE 30th Comput. Secur. Found. Symp. (CSF)*, Aug. 2017, pp. 263–275. [Online]. Available: https://ieeexplore.ieee.org/document/8049725
[80] J. Bonneau, I. Meckler, V. Rao, and E. Shapiro, "Coda: Decentralized cryptocurrency at scale," *Cryptol. ePrint Arch.*, Jan. 2020.
[81] G. Wang, J. Li, X. Wang, J. Li, Y. Yuan, and F.-Y. Wang, "Blockchain-based crypto management for reliable real-time decision-making," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 6, pp. 3333–3342, Dec. 2023.
[82] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
[83] W. Hu, X. Xia, X. Ding, X. Zhang, K. Zhong, and H.-F. Zhang, "SMPC-ranking: A privacy-preserving method on identifying influential nodes in multiple private networks," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 53, no. 5, pp. 2971–2982, May 2023.
[84] H. S. Sikandar, S. U. R. Malik, A. Anjum, A. Khan, and G. Jeon, "Cohort-based kernel principal component analysis with multi-path service routing in federated learning," *Future Gener. Comput. Syst.*, vol. 149, pp. 518–530, Dec. 2023.
[85] P. Chatzigiannis, F. Baldimtsi, and K. Chalkias, "SoK: Blockchain light clients," in *Proc. Int. Conf. Financial Cryptogr. Data Secur.* Cham, Switzerland: Springer, 2022, pp. 615–641.
[86] A. Bkakria, A. Tasidou, N. Cuppens-Boulahia, F. Cuppens, F. Bouattour, and F. Ben Fredj, "Optimal distribution of privacy budget in differential privacy," in *Proc. Int. Conf. Risks Secur. Internet Syst.*, Arcachon, France. Cham, Switzerland: Springer, Oct. 2018, pp. 222–236.
[87] M. Bhowmik, T. Sai Siri Chandana, and B. Rudra, "Comparative study of machine learning algorithms for fraud detection in blockchain," in *Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Apr. 2021, pp. 539–541.
[88] A. A. Ahmed and T. Eze, "An actor-based runtime environment for heterogeneous distributed computing," in *Proc. Int. Conf. Parallel Distrib. Process.*, 2019, pp. 37–43.
[89] W. Yu, X. Yang, and Y. Liu, "A novel hardware acceleration technique for high performance parallel FDTD method," in *Proc. IEEE iWEM*, Aug. 2011, pp. 149–153.
[90] J. Weng, J. Weng, G. Tang, A. Yang, M. Li, and J.-N. Liu, "PvCNN: Privacy-preserving and verifiable convolutional neural network testing," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2218–2233, 2023.
[91] Z. Zhang and J. Zhang, "Chebyshev functional link spline neural filter for nonlinear dynamic system identification," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1907–1911, Mar. 2022.
[92] L. M. Bach, B. Mihaljevic, and M. Zagar, "Comparative analysis of blockchain consensus algorithms," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 1545–1550.
[93] J. Sun, X. Yao, S. Wang, and Y. Wu, "Blockchain-based secure storage and access scheme for electronic medical records in IPFS," *IEEE Access*, vol. 8, pp. 59389–59401, 2020.

[94] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1613–1622.
[95] H. Ratnayake, L. Chen, and X. Ding, "A review of federated learning: Taxonomy, privacy and future directions," *J. Intell. Inf. Syst.*, vol. 61, no. 3, pp. 923–949, Dec. 2023.
[96] OpenMined. (Dec. 2023). *Understanding the Types of Federated Learning*. [Online]. Available: https://blog.openmined.org/federated-learning-types/
[97] M. Dieye, P. Valiorgue, J.-P. Gelas, E.-H. Diallo, P. Ghodous, F. Biennier, and É. Peyrol, "A self-sovereign identity based on zero-knowledge proof and blockchain," *IEEE Access*, vol. 11, pp. 49445–49455, 2023.
[98] E. Goh, D.-Y. Kim, K. Lee, S. Oh, J.-E. Chae, and D.-Y. Kim, "Blockchain-enabled federated learning: A reference architecture design, implementation, and verification," *IEEE Access*, vol. 11, pp. 145747–145762, 2023.
[99] R. Xie, Q. Tang, C. Liang, F. R. Yu, and T. Huang, "Dynamic computation offloading in IoT fog systems with imperfect channel-state information: A POMDP approach," *IEEE Internet Things J.*, vol. 8, no. 1, pp. 345–356, Jan. 2021.
[100] F. Falahatraftar, S. Pierre, and S. Chamberland, "An intelligent congestion avoidance mechanism based on generalized regression neural network for heterogeneous vehicular networks," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 4, pp. 3106–3118, Sep. 2023.
[101] Z. Zhou, F. Sun, X. Chen, D. Zhang, T. Han, and P. Lan, "A decentralized federated learning based on node selection and knowledge distillation," *Mathematics*, vol. 11, no. 14, p. 3162, Jul. 2023.
[102] A. A. Ahmed and G. Agunsoye, "A real-time network traffic classifier for online applications using machine learning," *Algorithms*, vol. 14, no. 8, p. 250, Aug. 2021. [Online]. Available: https://www.mdpi.com/1999-4893/14/8/250
[103] J. Peng, B. Tang, H. Jiang, Z. Li, Y. Lei, T. Lin, and H. Li, "Overcoming long-term catastrophic forgetting through adversarial neural pruning and synaptic consolidation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4243–4256, Sep. 2022.
[104] X. Wang, Y. Zhao, C. Qiu, Z. Liu, J. Nie, and V. C. M. Leung, "InFEDge: A blockchain-based incentive mechanism in hierarchical federated learning for end-edge-cloud communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 12, pp. 3325–3342, Dec. 2022.
[105] C. Gamanayake, L. Jayasinghe, B. K. K. Ng, and C. Yuen, "Cluster pruning: An efficient filter pruning method for edge AI vision applications," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 4, pp. 802–816, May 2020.
[106] Y. Lin, Z. Gao, H. Du, J. Kang, D. Niyato, Q. Wang, J. Ruan, and S. Wan, "DRL-based adaptive sharding for blockchain-based federated learning," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5992–6004, Jun. 2023.
[107] Y. Jin, S. Wei, J. Yuan, and X. Zhang, "Information-bottleneck-based behavior representation learning for multi-agent reinforcement learning," in *Proc. IEEE Int. Conf. Auto. Syst. (ICAS)*, Aug. 2021, pp. 1–5.
[108] Q. Fan and N. Ansari, "Workload allocation in hierarchical cloudlet networks," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 820–823, Apr. 2018.
[109] X. Chen, C. Luo, W. Wei, J. Xu, and S. Zhang, "Differential optimization federated incremental learning algorithm based on blockchain," *Electronics*, vol. 11, no. 22, p. 3814, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/22/3814
[110] V. Malik, R. Mittal, D. Mavaluru, B. R. Narapureddy, S. B. Goyal, R. J. Martin, K. Srinivasan, and A. Mittal, "Building a secure platform for digital governance interoperability and data exchange using blockchain and deep learning-based frameworks," *IEEE Access*, vol. 11, pp. 70110–70131, 2023.
[111] K. Eze, A. A. Ahmed, and C. Akujuobi, "SecioTComm: An actor-based model and framework for secure IoT communication," *Sensors*, vol. 22, no. 19, p. 7313, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/19/7313
[112] A. A. Moamen, H. S. Hamza, and I. A. Saroit, "Secure multicast routing protocols in mobile ad-hoc networks," *Int. J. Commun. Syst.*, vol. 27, no. 11, pp. 2808–2831, 2014, doi: 10.1002/dac.2508.
[113] A. A. Ahmed and M. Echi, "Hawk-eye: An AI-powered threat detector for intelligent surveillance cameras," *IEEE Access*, vol. 9, pp. 63283–63293, 2021.
[114] Z. Xing, Z. Zhang, M. Li, J. Liu, L. Zhu, G. Russello, and M. R. Asghar, "Zero-knowledge proof-based practical federated learning on blockchain," *arXiv:2304.05590*, 2023.
[115] A. Liu, Y. Liu, Q. Wu, B. Zhao, D. Li, Y. Lu, R. Lu, and W. Susilo, "CHERUBIM: A secure and highly parallel cross-shard consensus using quadruple pipelined two-phase commit for sharding blockchains," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 3178–3193, 2024.

AHMED ABDELMOAMEN AHMED (Member, IEEE) received the B.Sc. and M.S. degrees in computer science from Cairo University, Egypt, and the Ph.D. degree in computer science from the University of Saskatchewan, in August 2017. In Fall 2017, he joined the Department of Computer Science, Prairie View A&M University (PVAMU), Prairie View, TX, USA, where he is currently an Associate Professor. Since joining PVAMU, he has been actively involved in several research projects, including IoT and machine learning applications. He is teaching undergraduate and graduate courses and is eager to transfer his passion and experience to underrepresented students at PVAMU by involving them in research. He has extensive experience in mentoring undergraduate and graduate students to do research. His research interests include parallel and distributed systems with applications in various domains, such as the Internet of Things (IoT) and artificial intelligence (AI). He serves as a reviewer for several international conferences and journals.

OLUWAYEMISI O. ALABI received the first B.Sc. degree in electronics and electrical engineering and the second B.Sc. degree in HCI with a major in 3D animation. She is currently pursuing the M.S. degree in computer information systems, specializing in cybersecurity and focusing on fraud detection. Her research interests include payment solutions and using machine learning techniques to combat fraud.