Poster: MicroFL: A Lightweight, Secure-by-Design Edge Network Fabric for Decentralized IoT Systems

Ronghua Xu
Binghamton University
rxu22@binghamton.edu

Yu Chen
Binghamton University
ychen@binghamton.edu

Jian Li
Binghamton University
lij@binghamton.edu

With the aggressive growth of smart environments, a large amount of data are generated by edge devices in distributed Internet-of-Thing (IoT) hierarchies, consisting of the cloud, the fog and edge devices. It is challenging to build secure edge networks on top of highly heterogeneous, resource-constrained IoT devices with the increasing security and privacy concern. The edge computing and IoT community has recognized the desperate need of a secure-by-design, self-adaptive, and totally decentralized network architecture to provide information assurance and device protection in the still quick-growing large scale IoT environment. Ideally, an ultimate solution shall be a protocol that functions as what TCP/IP does in the Internet.

The concept of Federated Learning (FL) [7] has recently been introduced as a promising decentralized Machine Learning (ML) approach to facilitate large-scale collaborative learning of a global ML model without transmitting raw data from distributed edge devices to a central server. This significantly reduces communication costs as only model updates rather than raw data are sent from edge devices to the center server, which further improves user privacy, and reduces computational costs by leveraging computing resources at edge devices. Meanwhile, it also raises new challenges of performance, incentive, privacy and security. For example, the stochastic gradient descent model updates may leak important user information although raw data are not transmitted. To address this issue, several approaches have been proposed by leveraging differential privacy [8] or secure aggregation [2]. However, these approaches only consider Crash Fault Tolerant (CFT) with the occurrence of random dropouts, and assume that edge devices are honest and will not submit "fake" updates nor collude with other devices to poison the learning process. In fact, FL framework should be Byzantine Fault Tolerant (BFT) [5] compatible to secure model learning and inference process under a trustless distributed network environment.

With a great potential to revolutionize information technology, blockchain provides a decentralized architecture that improves system performance and removes the single point failure issue caused by a central hub, which is for aggregation in FL. Also, blockchain brings a public distributed ledger technology to support immutability, auditability and traceability for data provenience. Furthermore, incentive mechanism in blockchain motivates more users to actively participate collaborating training and discourages participants from misbehaving. Researchers are considering adopting blockchain to support distributed learning for data provenance and incentive purposes [1], [6], [9]. Merely integrating existing cryptocurrency-oriented blockchain technologies into IoT, however, these approaches incur tremendous challenges. Due to the large volume of transactions among participants, high throughput and low latency become the keys. The computing intensive consensus protocols like Proof-of-Work (PoW) is not affordable to IoT devices. Furthermore, blockchain consensus requires frequent data transmissions on a peer-to-peer network, which brings significant communication overhead on the light edge computing network and extra energy consumption on IoT devices.

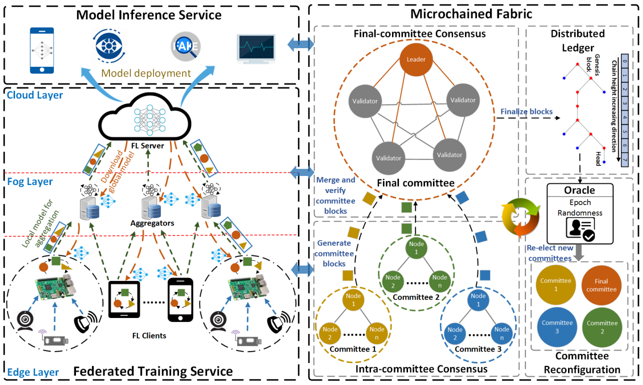


Fig. 1. The MicroFL System Architecture.

Rationale and System Design of MicroFL: Inspired by blockchain based solutions for Distributed Machine Learning, a Microchain [10] enabled Federate Learning fabric (MicroFL) is proposed. Aiming at a secure-by-design, self-adaptive, and totally decentralized network architecture, MicroFL enables an efficient, privacy-preserving and secure cooperative training under decentralized cross-device FL setting [4] in heterogeneous Mobile Edge Computing (MEC) environment. Figure 1 shows the MicroFL architecture consisting of (1) *a hierarchical FL framework* that leverages a hierarchical network to support distributed training and model inference services, and (2) *a hybrid microchained fabric* that provides decentralized security and privacy-preserving properties for the FL system. MicroFL builds an organic mutual reinforced framework in which blockchain interconnects key components and tasks in FL. The rationale behind the MicroFL is described as follows:

(i) *Microchain* provides fundamental networking and security mechanism for FL system. Leveraging a public distributed ledger to record and verify data in FL process without relying on a centralized server, microchain improves data availability by mitigating performance bottleneck and removing the single point of failure. The lightweight and hybrid consensus mechanism adopted by Microchain also brings security features like immutability, auditability and traceability. They efficiently enhance privacy and security issues of existing FL approaches.

(ii) *FL* is considered as the upper-level enabling technology on the microchained network to provide cooperative training on decentralized data and global model inference for smart applications. The fast and high quality services provided by FL, such as computation, storage and optimization algorithms based on trained global models, are also promising to improve performance and security in microchained network.

In MicroFL, the interaction between FL and microchain can be envisioned analogous to the TCP/IP protocols in the Internet. Microchain acts as the network infrastructure to provide decentralization and security for FL, like IP functions as a connection-less, best-effort service network layer protocol for the TCP. On the other hand, FL utilizes its capability of model learning and inference to enable high level intelligent applications as well as the optimization for microchained network, like TCP provides a connection-oriented, reliable, end-to-end transport layer service to tackle the open issues that IP does not address and meet high quality of service requirements for upper lever applications.

Hierarchically Federated Learning Framework: Figure 1 demonstrates the FL framework consisting of hierarchical cloud-fog-edge computing architecture. FL clients, the aggregator and the FL server are key players in the model training cycle, including global model propagation (down-stream) and local model aggregation (up-stream). For a FL task, a FL server firstly advertises task specifications including global model data and training program, among aggregators. The aggregators are deployed on fog layer as intermediates to transfer task specifications to edge computing devices. All FL clients are allocated at edge layer to locally compute a model update by executing the training program. In model aggregation process, each aggregator calculates a fog level model update based on received model updates and sends it to FL server, where the global model is aggregated. Given the analysis on trained global model, good candidates are selected and deployed on smart applications, like keyboard perdition and anomalous behavior detection based on multi-camera.

Hierarchically distributed computing architecture not only provides system scalability for large-scale deep learning tasks based on geographically distributed IoT devices, it also supports flexible management and coordinated central and local decisions among heterogeneous networks and application domains. The microchain interconnects FL server, aggregator and FL clients to protect data and model updates and supports secure multi-party computing functions in FL services.

Hybrid Microchained Fabric: The hybrid microchained fabric is based on microchain, a hybrid blockchain architecture to enable a secure, scalable and lightweight distributed ledger for resource-constrained IoT systems at the network [10]. The microchain relies on a permissioned network that provides basic security primitives, such as public key infrastructure (PKI), identity authentication and access control, etc. It is easily deployed in heterogeneous MEC networks with varying security requirements. Given a random committee election algorithm, only a small subset of participants in the network are chosen as validators to perform consensus protocol. Therefore, the message propagation complexity and block confirmation time are reduced, such that consensus performance is improved to allow a lower communication cost and a high throughput for the IoT network. In addition, microchain uses a computationally efficient virtual mining to enable a probabilistic block generation, which is affordable to resource-constrained IoT devices. Furthermore, the microchain adopts incentive compatibility based on rewarding and punishment strategies, which helps to address incentive issues in FL.

To improve the scalability of consensus and ensure privacy preservation in the hierarchical FL network, a two-level committee consensus protocol is proposed: intra-committee consensus and final-committee consensus, as shown in right of Fig. 1. In local model computing aggregation process, an aggregator and its associated FL clients form a local committee, which executes an efficient Practical BFT (PBFT) [3] style intra-committee consensus protocol. Only data and transactions within the committee are validated by committee members. Therefore, it brings higher throughput and lower communication overhead on the edge network and data privacy is protected. The final-committee performs a hybrid Proof-of-Credit (PoC)-Voting-based Chain Finality (VCF) consensus [10] among the cloud and aggregators to maintain a global distributed ledger. All intra-committee blocks and final-committee transactions are finalized on a global distributed ledger, which provide immutability, auditability and traceability for model learning, aggregation and inference processes. The committee reconfiguration utilizes a bias-resistant epoch randomness foundation to ensure an unpredictable and statistical representative committee selection.

Conclusions: We present the design of MicroFL, a lightweight and secure-by-design edge network fabric for decentralized IoT systems. The MicroFL enables efficient, secure and privacy-preserving cooperative training by leveraging advantages of FL and microchain, a lightweight blockchain customized for IoT systems at the edge. Our future work is to integrate MicroFL with ongoing smart surveillance project to enable anomalous behavior detection based on Multi-view cameras.

References

[1] S. Awan, F. Li, B. Luo, and M. Liu, "Poster: A reliable and accountable privacy-preserving federated learning framework using the blockchain," in *ACM CCS, 2019.*

[2] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *ACM CCS, 2017.*

[3] M. Castro, B. Liskov *et al.*, "Practical byzantine fault tolerance," in *OSDI*, vol. 99, no. 1999, 1999, pp. 173–186.

[4] P. Kairouz, H. B. McMahan *et al.*, "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.

[5] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Trans. Prog. Lang. Sys.*, vol. 4, no. 3, 1982.

[6] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial iot," *IEEE Transactions on Industrial Informatics*, 2019.

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, 2017.

[8] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning differentially private recurrent language models," *arXiv preprint arXiv:1710.06963*, 2017.

[9] J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive," *IEEE Trans. Depend. Sec. Comp.*, 2019.

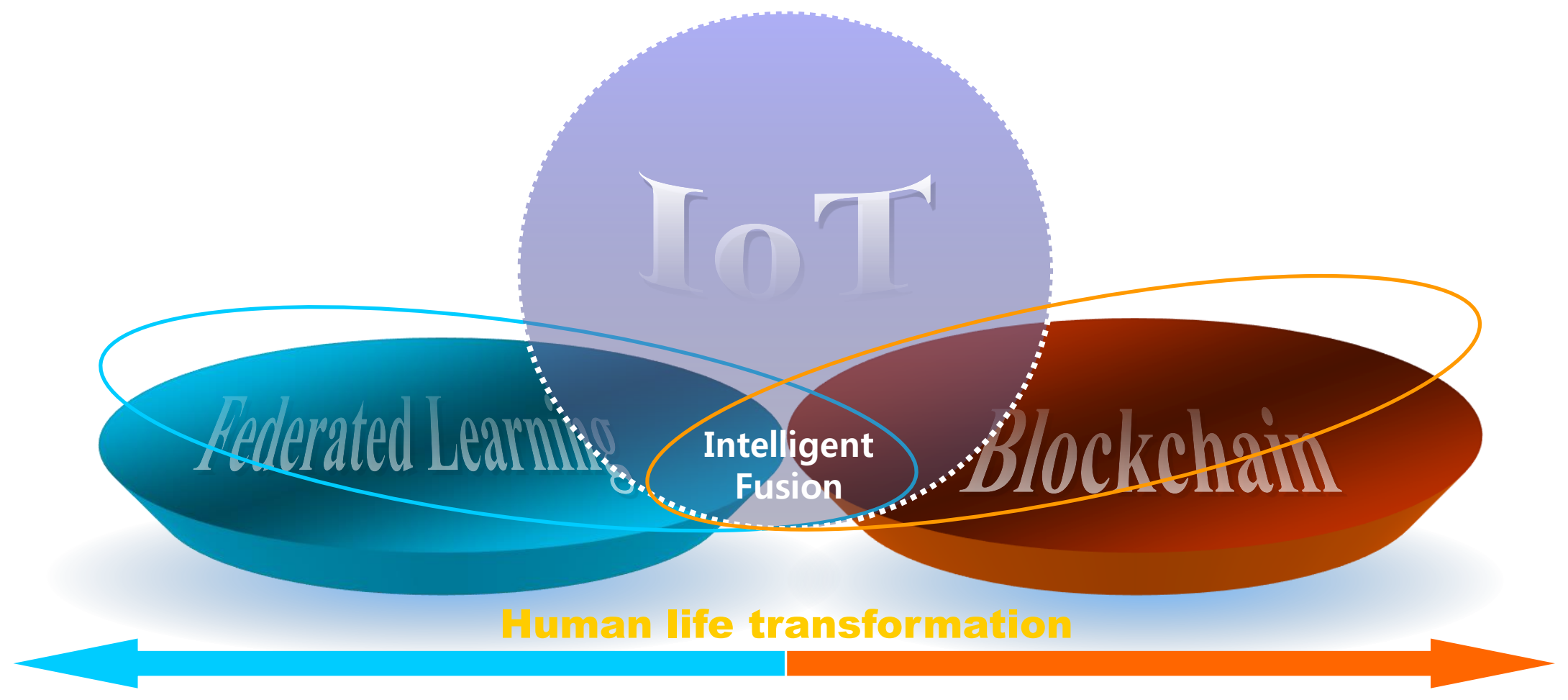
[10] R. Xu, Y. Chen, E. Blasch, and G. Chen, "Microchain: A hybrid consensus mechanism for lightweight distributed ledger for iot," *arXiv preprint arXiv:1909.10948*, 2019.

NDSS 2020

MicroFL: A Lightweight, Secure-by-Design Edge Network Fabric for Decentralized IoT Systems

Ronghua Xu, Yu Chen, Jian Li
Binghamton University, SUNY, Binghamton, NY 13902, USA

Introduction



- Distributed Internet-of-Thing (IoT) systems generate a large amount of data for Machine Learning (ML)
- Federated Learning (FL) facilitates large-scale collaborative learning of a global ML model with privacy preservation
- Blockchain brings decentralization, security and privacy enforcement into distributed cross-devices FL scenarios
- Intelligent Fusion based on IoT, FL and Blockchain needs a secure-by-design, self-adaptive, and totally decentralized network architecture

Key Challenges

IoT	• Constrained computing and storage resource • Energy efficiency and heterogeneous network • Data security and privacy concerns
FL	• Privacy-preserving in learning and aggregation • Robust to adversarial attacks and failures • Complex incentive mechanism
Blockchain	• The trade-off between scalability and efficiency • The computation cost for consensus algorithm • The ledger data storage overhead on host

Main Objectives

- Design a secure, self-adaptive, and totally decentralized network architecture
- Enable an efficient, privacy-preserving and secure cooperative training framework on top of distributed IoT hierarchies, consisting of the cloud, the fog and edge devices
- Integrate an optimized and lightweight blockchain fabric to enhance privacy, security and incentive compatible guarantees for FL
- Build an organic system in which FL and blockchain have a mutual reinforced relationship

Contact Information

Web : https://www.binghamton.edu/ece

Email : rxu22@Binghamton.edu

Phone : +1 (607) 338-8444

System Architecture

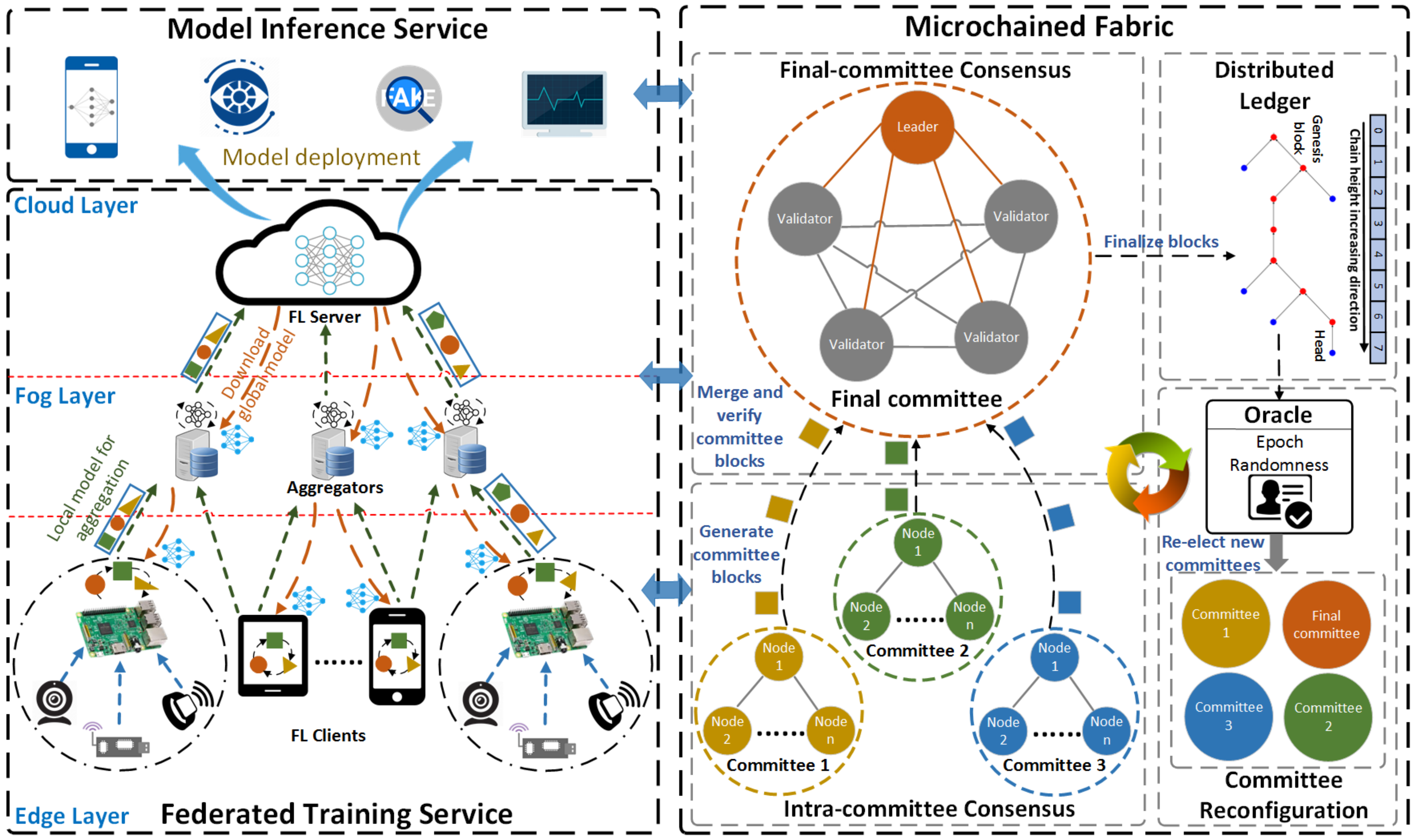


Figure 1 : System Architecture of MicroFL.

Hierarchical FL Framework

- Hierarchical cloud-fog-edge computing architecture
- Players: FL clients, aggregators and FL server
- Main workflows: Global model propagation (down-stream) and local model aggregation (up-stream)
- Support scalability, flexible management and coordinated central and local decisions for distributed FL service under heterogeneous network and IoT devices

Hybrid Microchained Fabric

- Interconnect key players in FL to enhance user data & model updates protection and multi-party computing security
- Two-level committee consensus protocols offer an efficient, scalable and privacy-preserving distributed ledgers for hierarchical FL framework
- A bias-resistant public randomness mechanism ensures unpredictable and statistical representative committee formation

Design Rationale

- **Immutability**
- **Auditability**
- **Traceability**

Federated Learning FL acts as upper-level enabling technology to provide decentralized cooperative training and global model inference

Microchain Microchain provides fundamental networking and security mechanism for FL system

- **Computation & storage**
- **Optimization algorithm**
- **Intelligent services**

Microchain Summary

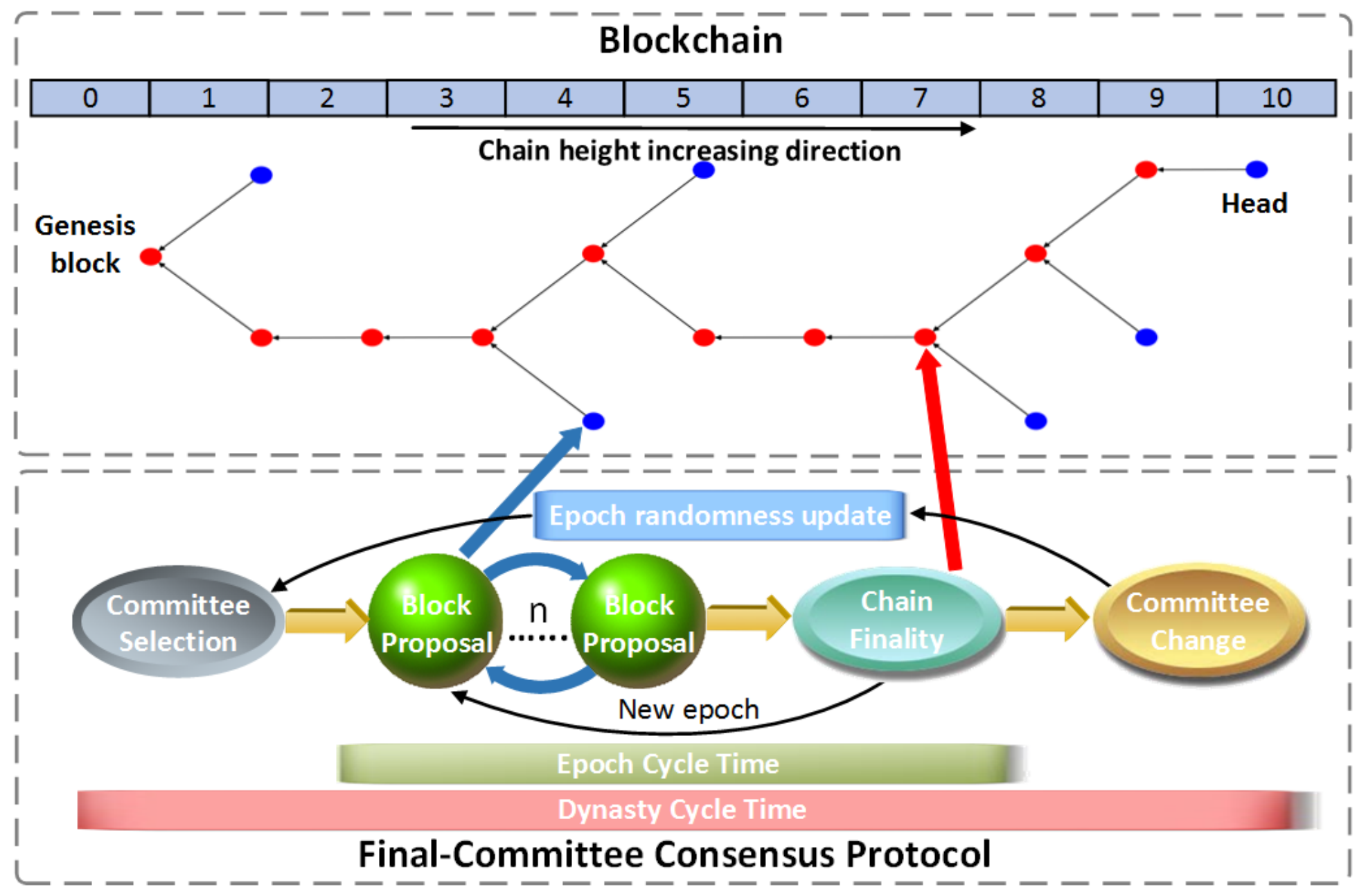


Figure 2: Key Components and Workflows in Microchain [1].

The features of Microchain

Permissioned Network	Random Committee Election
• Provide basic security primitives, such as public key infrastructure (PKI) and access control • Handle heterogeneous networks with varying security requirements	• Unpredictable committee selection based on unbiased randomness protocol • Lower communication cost and high throughput for IoT
Efficient Virtual Mining	**Incentive Capability**
• A computational efficient virtual mining manner for probabilistic block generation • Limited computation and storage overhead as executing on IoT devices	• Adopt incentive compatibility based on rewarding and punishment strategies • Helpful to address incentive issues in FL system

References:

[1] R. Xu, Y. Chen, E. Blasch, and G. Chen, “Microchain: A hybrid consensus mechanism for lightweight distributed ledger for iot,” *arXivpreprint arXiv:1909.10948*, 2019.

Ongoing Efforts and Future Directions

Table 1: Configuration of Experimental Nodes.

Device	Dell Optiplex 760	Raspberry Pi 3 Model B+
CPU	3 GHz Intel Core TM (2 cores)	Broadcom ARM Cortex A53 (ARMv8), 1.4GHz
Memory	4GB DDR3	1GB SDRAM
Storage	250G HHD	32GB (microSD card)
OS	Ubuntu 16.04	Raspbian GNU/Linux (Jessie)

Table 2: Network latency for one round of Microchain (In seconds; computed over 50 runs; Setting I:4 nodes; Setting II: 16 nodes)

Operations	Setting I	Setting II
Commit Transaction	0.16	0.25
Block Proposal	0.5	1.7
Chain Finality	1.4	21.5

Table 3: Throughput based on average transactions rate (in M/h: Mbytes per hour)

Block size	Transactions rate
512 K	202
1 M	293
2 M	405
4 M	263

Table 4: Average processing time of running each Microchain operations with 1 M block size (In milliseconds, computed over 50 runs; Setting I: Desktop; Setting II: Raspberry Pi)

Operations	Setting I	Setting II
Verify Transaction	47	179
Mining Block	27	147
Validate Block	358	492
Verify Vote	303	436

	CPU usage (%)
Setting I	10
Setting II	15

Our future work includes but not limited to:

1) Integrating MicroFL with our current smart surveillance research to enable anomalous behavior detection based on Multi-view cameras;

2) Investigating impacts of executing MicroFL in terms of detection accuracy of training algorithm, networking efficiency, data security and privacy protection.