

Université de Mons

2024-2025

Report on scientific Internship

Delmas Steve

Traineeship within the University of Mons

Summary sheet

* + Type of internship: scientific internship
  + Academic year: 2024-2025
  + Autor: Delmas Steve
  + Training: Engineering student, fourth year ESIEE Paris
  + Title of the report: Federated Learning and Blockchain for Edge Computing
  + Host Organization: University of Mons
  + Host country: Belgium
  + Trainee teacher: Carlos Alberto Valderrama Sakuyama
  + Trainee tutor: Jeremie Biringanine Ruvunangiza
  + Keywords: Federated Learning, Blockchain, Decentralized AI, Non-IID Data.
  + Theme School: Artificial Intelligence, Distributed Systems and Cybersecurity in Edge Computing.

1. Introduction  
  
In a world where connected devices are generating increasingly large volumes of data, traditional centralized data processing models are reaching their limits. Edge computing has emerged as a solution, allowing data to be processed closer to its source, whether on sensors, mobile devices, or embedded systems. This approach reduces latency, lowers bandwidth consumption, and enables real-time decision-making. However, it also introduces new challenges in terms of security, privacy, and resource management.  
  
To address these challenges, this project explores the integration of two advanced technologies: federated learning and blockchain. Federated learning is a decentralized machine learning approach that enables multiple devices or nodes to collaboratively train a global model without sharing their local data. This preserves data privacy while leveraging distributed datasets. Blockchain, on the other hand, provides a transparent, immutable, and decentralized ledger that ensures data integrity and trust among participants in the system.  
  
This internship is part of a research project at the University of Mons (UMons), with the primary goal of studying the synergy between federated learning and blockchain within an edge computing context. By combining these technologies, the project aims to design a secure, efficient, and privacy-preserving architecture for distributed data processing.

**2. Federated Learning**

**2.1 Definition and Principles**

Federated Learning (FL) is a decentralized machine learning paradigm where multiple clients (such as smartphones, edge devices, or local servers) collaboratively train a shared model without transferring their local data. Instead of uploading raw data to a centralized server, clients perform local training and only send model updates—such as gradients or weights—which are then aggregated into a global model.  
  
This approach, introduced by Google in 2016, aims to address privacy, scalability, and communication challenges in distributed environments, especially for mobile devices and edge computing [1][7].  
  
FL can be implemented in two main forms:  
- Centralized FL, where a central server coordinates the aggregation of updates;  
- Decentralized (peer-to-peer) FL, where clients synchronize directly, improving robustness and scalability [2].

**2.2 Key Advantages**

- Privacy preservation: local data never leaves the device.  
- Communication efficiency: only model parameters are exchanged, reducing bandwidth usage.  
- Scalability: FL supports large-scale systems with millions of heterogeneous devices.  
- Regulatory compliance: it aligns with data protection laws like GDPR and HIPAA [1,2].

**2.3 Technical Challenges**

Despite its advantages, FL faces several challenges:  
- Non-IID data: local data distributions vary significantly, making convergence more complex [1][7].  
- Unbalanced datasets: some clients have much more data than others.  
- System heterogeneity: varying device capabilities affect training speed and reliability.  
- Privacy risks: model updates may still leak information if not properly protected.  
- Lack of incentives: motivating fair participation remains a concern [3][8].

**2.4 Optimization Techniques and Architectures**

Several strategies have been proposed to improve FL:  
- FedAvg: a baseline algorithm where clients locally train and send averaged updates [1].  
- FedProx, FedNova: to handle client heterogeneity and improve convergence.  
- Communication reduction: via gradient sparsification, compression, and selective client updates [2].  
- Privacy-enhancing methods: including differential privacy, secure multiparty computation, and homomorphic encryption[3][6].

**2.5 Applications**

Federated Learning is being applied in:  
- Healthcare: collaborative diagnosis models without sharing patient data.  
- Finance: fraud detection without exposing customer records.  
- Edge computing and IoT: enabling real-time local learning on smart sensors, wearables, and autonomous vehicles [2, 4].  
  
FL is thus a foundational technology for building privacy-aware, scalable, and decentralized AI systems—especially in the context of edge computing.

**3. Blockchain**

**3.1 Definition and Principles**

Blockchain is a distributed, immutable ledger technology in which all participants maintain a complete copy of all validated transactions. Transactions are grouped into blocks, each cryptographically linked to the previous one, forming an unalterable chain of data [4].  
  
The technology relies on consensus mechanisms (e.g., Proof-of-Work, Proof-of-Stake) to validate transactions without a centralized authority. This ensures security, transparency, and traceability in decentralized environments.

**3.2 Key Advantages**

- Decentralization: no single point of failure or control.  
- Immutability: once recorded, transactions cannot be altered.  
- Traceability: every operation is verifiable across the network.  
- Transparency: in public blockchains, all activities are openly visible and auditable [4] [5].

**3.3 Blockchain Types**

Blockchain systems can be classified into:  
- Public blockchains (e.g., Bitcoin, Ethereum): open to everyone with no permission required.  
- Private blockchains: access is restricted and controlled by a single entity.  
- Consortium blockchains: governed by a group of organizations with shared permissions [3].

**3.4 Applications in Edge Computing and AI**

In edge computing environments, blockchain enables:  
- Secure validation and logging of model updates and data exchanges across distributed nodes.  
- Verification of model integrity through cryptographic hashes and timestamps.  
- Serverless coordination among devices by enabling peer-to-peer trust and record keeping.  
  
Blockchain thus becomes a natural complement to Federated Learning, especially in settings where trust between participants is limited or where security and traceability of model updates are critical.

**4. Integration of Federated Learning and Blockchain**

**4.1 Motivation and Context**

Although Federated Learning (FL) preserves privacy by keeping data local, it typically relies on a central server to aggregate model updates. This architecture introduces several limitations:  
- A single point of failure,  
- Susceptibility to model falsification or manipulation,  
- A lack of incentive mechanisms to encourage fair client participation.  
  
To address these issues, blockchain technology has been proposed as a complementary infrastructure layer for FL, forming a new paradigm known as Blockchain-enabled Federated Learning, or FLchain [1, 4].

**4.2 Benefits of Integration**

- Full decentralization: blockchain removes the need for a central aggregation server.  
- Contribution traceability: all model updates are timestamped and immutably recorded on the blockchain.  
- Incentive mechanisms: clients can be rewarded using tokens or smart contracts for participating or behaving honestly.  
- Enhanced security: blockchain helps prevent model poisoning and tampering attacks.  
- Consensus-driven updates: participants collectively validate and approve model updates [4] [5].

**4.3 General Workflow**

A typical FLchain system follows these steps:  
1. Each client trains a local model on its private data.  
2. The resulting model update is packaged into a blockchain transaction.  
3. A group of nodes (miners or validators) verifies the transaction using a consensus mechanism.  
4. Once validated, the block containing the update is added to the blockchain.  
5. All participants download the block and use the aggregated updates to refine the global model.  
  
This cycle is repeated until the model converges or reaches a desired accuracy.

**5. Use Cases and Practical Applications**

**5.1 Healthcare**

Medical data is highly sensitive and often protected by strict privacy regulations.  
FLchain solution: Hospitals and clinics can collaboratively train diagnostic models without sharing raw patient data, while blockchain ensures traceable and tamper-proof contributions.  
Example: Multiple hospitals jointly train a cancer detection model where every update is recorded on-chain, enabling full transparency and auditability. [5] [7]

**5.2 Finance**

FLchain enables decentralized training for fraud detection, credit scoring, and risk assessment, while blockchain maintains an immutable audit trail of all model updates.  
Each update is cryptographically signed and transparently stored, reducing the risk of manipulation. [4] [8]

**5.3 Internet of Things (IoT) and Edge Computing**

FLchain allows devices to train and share models locally, avoiding central data collection.  
It uses lightweight blockchain frameworks to validate and synchronize model updates securely.  
Example: A factory’s edge devices collaboratively learn to detect machine failures, while each contribution is logged on a private blockchain [2] [4] [6].

**5.4 Smart Cities and Transportation**

FLchain enables real-time coordination of autonomous vehicles, privacy-preserving urban monitoring, and traceability of algorithmic decisions—critical for transparency and accountability in public systems [2] [5].

**6. Bibliography**

[1] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. AISTATS.

[2] Ruvunangiza, J., & Valderrama, C. (2022). Applying Federated Learning: A Parallel and Distributed Computing Perspective. Topics in Parallel and Distributed Computing.

[3] Blockchain-Enabled Federated Learning: A Survey. Qu et al., ACM Computing Surveys, 2022.

[4] Nguyen, D. C., Ding, M., Pham, Q.-V., Pathirana, P. N., Le, L. B., et al. (2021). Federated Learning Meets Blockchain in Edge Computing: Opportunities and Challenges. IEEE Internet of Things Journal.

[5] Zeng, M. et al., "A Survey on Blockchain for Federated Learning," IEEE Access, 2021.

[6] Peng, Z. et al., "BGFL: Blockchain-Enabled Group Federated Learning for Industrial Wireless Edge," Journal of Cloud Computing, 2024.

[7] Ning, Z. et al., "A Comprehensive Survey of Blockchain-Based Federated Learning: Principles, Challenges, and Opportunities," Applied Sciences, 2024.

[8] Chhetri, S. R. et al., "Privacy-Preserving Federated Learning with Blockchain: Architectures and Research Challenges," arXiv, 2023.