Blockchain and Federated Learning: Protecting and Securing Information

Nii Attoh-Okine, PhD., P. E., F. ASCE

UNIVERSITY OF DELAWARE
ENGINEERING

Introduction

- Rapid advancement of digital Information
- Data explosion
- Data collaboration
- Weakness of Traditional Data Sharing and Storage

- Using blockchain to create an immutable audit trail for federated models

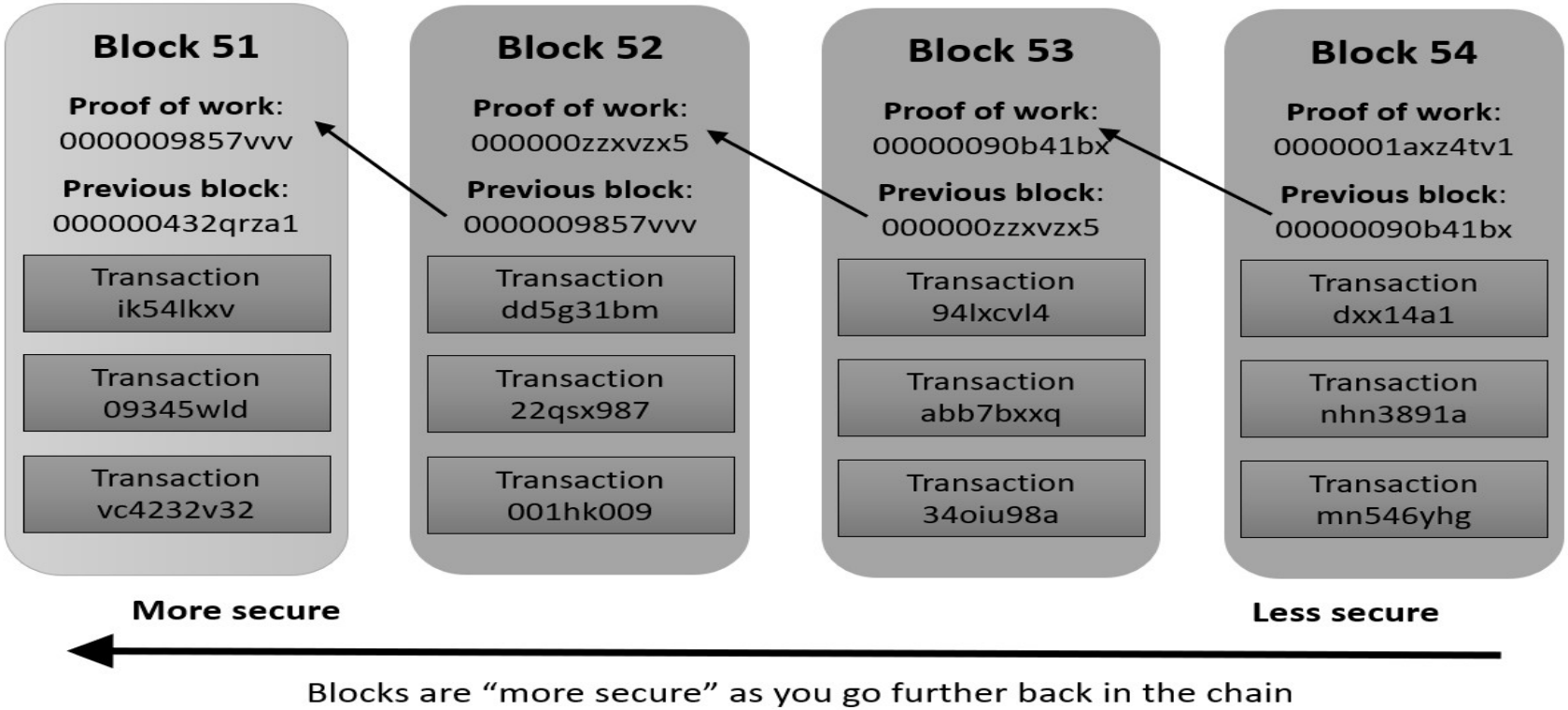
- Enhancing encryption between nodes and federated model to maintain better privacy-preservation

Why Blockchain

- Blockchain solves a challenging problem in Data Science of exchanging reliable information over unreliable network on which some of the participants cannot be trusted.

Blockchain

- A Database encompassing a physical chain of fixed length blocks that can include 1 to N transactions
- Each transaction added to a new block is validated
- When block is completed it is added to the end of existing chain of blocks

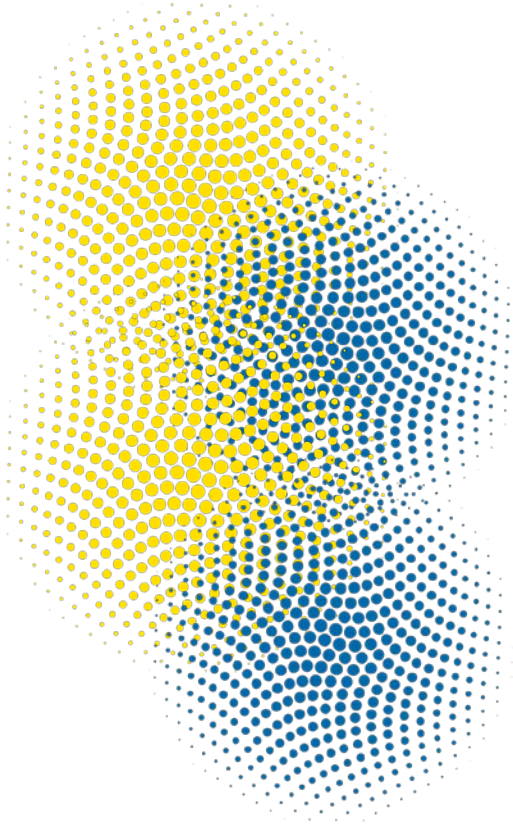

Block 51
Proof of work:
0000009857vvv
Previous block:
000000432qrza1
Transaction
ik54lkxv
Transaction
09345wld
Transaction
vc4232v32
Block 52
Proof of work:
000000zzxvzx5
Previous block:
0000009857vvv
Transaction
dd5g31bm
Transaction
22qsx987
Transaction
001hk009
Block 53
Proof of work:
00000090b41bx
Previous block:
000000zzxvzx5
Transaction
94lxcvl4
Transaction
abb7bxxq
Transaction
34oiu98a
Block 54
Proof of work:
0000001axz4tv1
Previous block:
00000090b41bx
Transaction
dxx14a1
Transaction
nhn3891a
Transaction
mn546yhg
More secure
Less secure
Blocks are “more secure” as you go further back in the chain

Steps

- And new and undeletable transactions and organize them into blocks
- Cryptographically verify each transaction in the block
- Append the new block to the end of the existing immutable blockchain

Examples: Smart Contracts

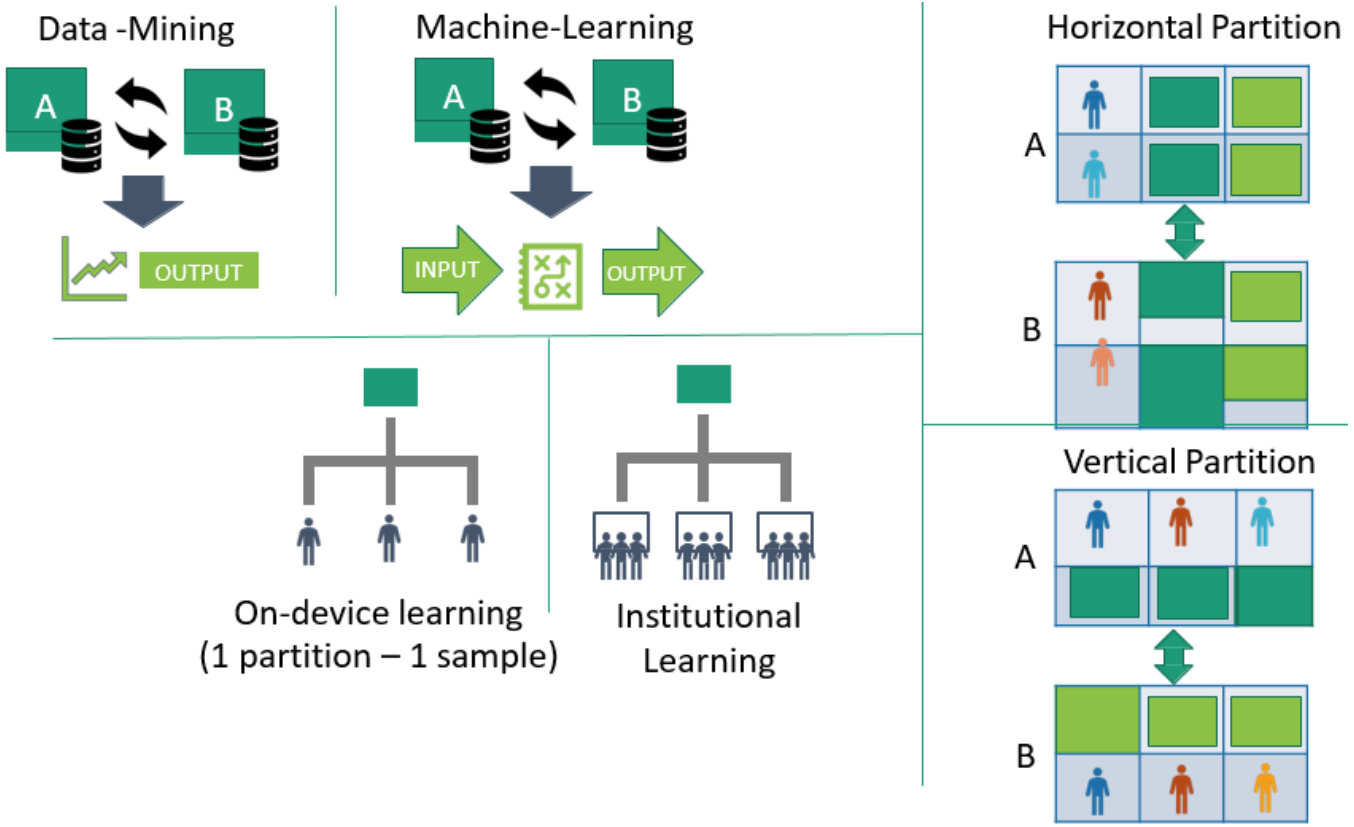
- Finance
- Agriculture
- Health
- Supply Chain
- Smart Cities



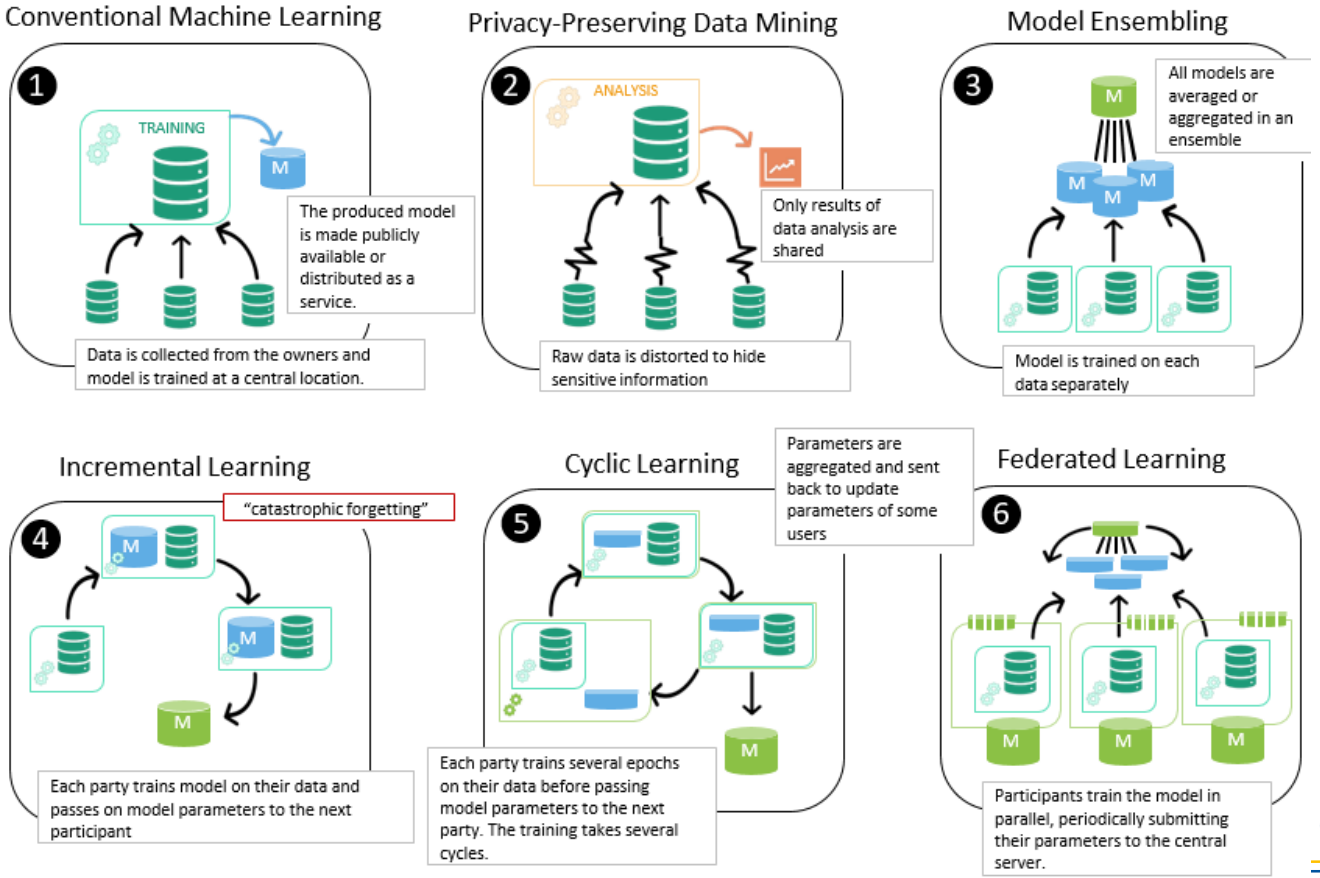
Data Collaboration Definition

Data Collaboration– utilizing data across several institutions for the purposes of creating knowledge or providing services without sharing the data.

Data Collaboration Taxonomy



Data Collaboration Architectures



Privacy and Security of Federated Learning

Privacy Protection by design:
- Ephemeral
- Focused
- Aggregate

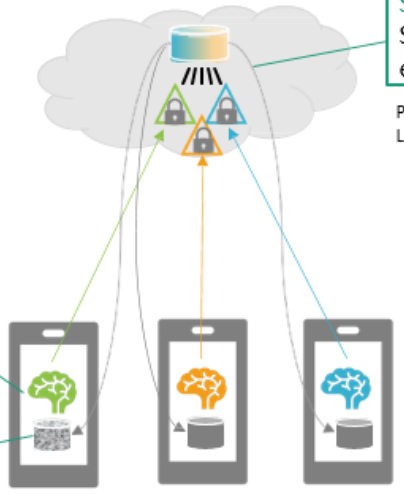
Data leakage to the central server
Individual model updates can be sensitive depending on the model
(e.g. keyboard prediction model)

Inference attacks on the model
If the resulting model can be inspected or queried by the end user membership inference attacks are possible

Membership Inference Attacks Against Machine Learning Models. (Shokri et al. *2017)*

Secure Aggregation:
SMC, homomorphic encryption

Practical Secure Aggregation for Privacy-Preserving Machine Learning. (Bonawitz et al. 2017)



Model Poisoning
Vulnerable by design against a client-side adversary, because the central server is blind to user data

How To Backdoor Federated Learning. (Bagdasaryan et al. 2018)

Privacy Accountant

Deep Learning with Differential Privacy((Abadi et al. 2016)

Blockchain Technology

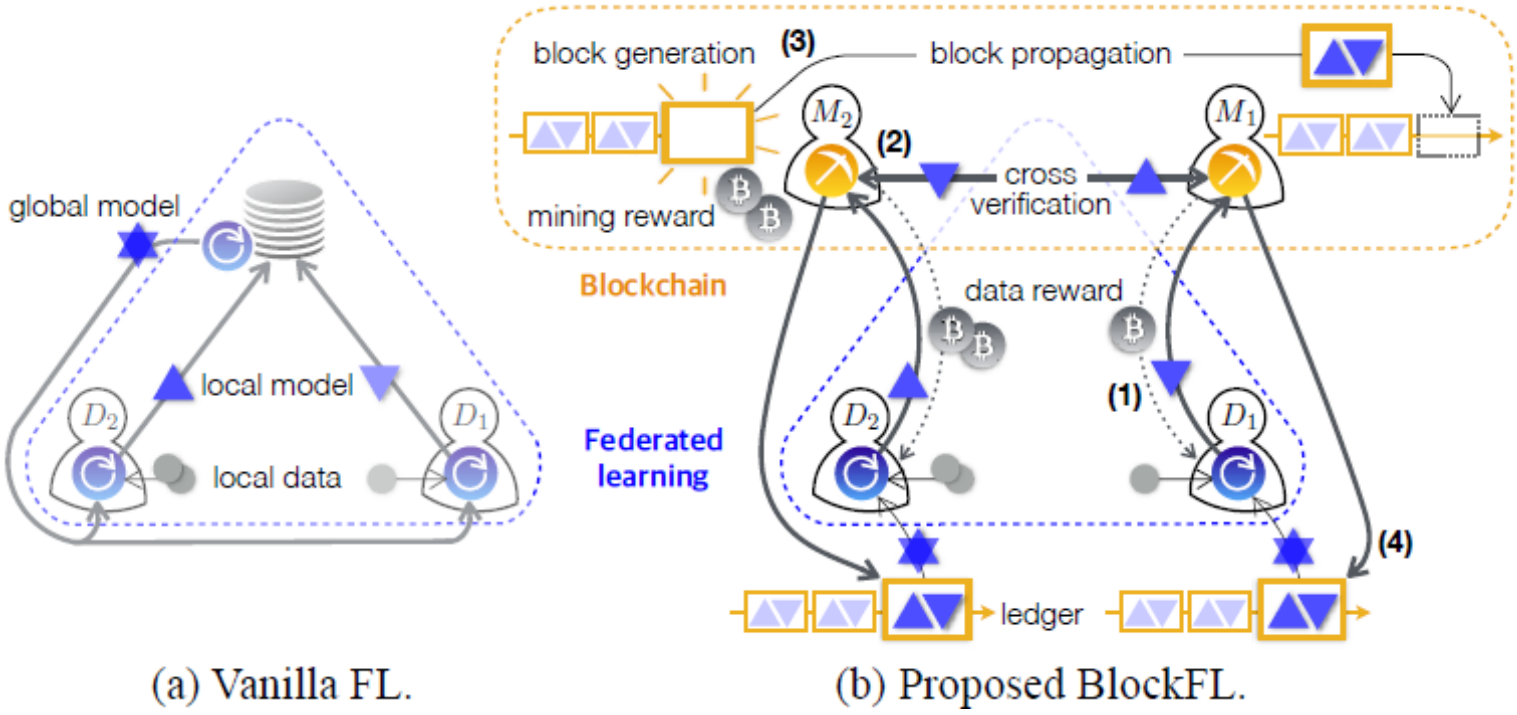
On-Device Federated Learning via Blockchain and its Latency Analysis. (Kim et al. 2018)

Differential Privacy

Differentially Private Federated Learning: A Client Level Perspective. (Geyer et al. 2017).

* None of the Privacy-Preserving Federated Learning Frameworks had been tested so far on real-life medical data

BlockFL (Hyesung Kim et al.)



(a) Vanilla FL.

(b) Proposed BlockFL.

BlockFL

- Each node and uploads the local model update to its associated miner in the blockchain network
- Miners exchange and verify all the local model updates
- Miners generates a block where the verified local model updates are recorded
- The generated block storing the aggregate local model updates is added to a blockchain.
- Each device computes the global model update from the new block.

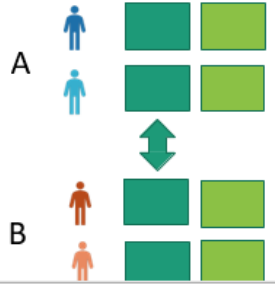
Medical Data Collaboration Use Cases

Clinical Decision Support System

- Computer-aided Interpretation of medical images
- COVID
- Design treatment plans
- Survival prediction

Multi-institutional

Horizontal



References

1. Blockchained On-Device Federated Learning Hyesung Kim et al.
2. Blockchain : Bambara, JJ and Allen, PR. McGraw Hill 2018.
3. Risk and Advantages of Federated Learning for Health Care Data Collaboration. ASCE Journal of Risk and Uncertaity Mangement. Bogdanova, A; Attoh-Okine, N; and Sakurai, T.

UNIVERSITY OF DELAWARE

ENGINEERING

CIVIL & ENVIRONMENTAL ENGINEERING ce.udel.edu