XuperData: Secure Data Circulation On Blockchain

Bing Duan, Senior R & D Engineer@Baidu Inc.

July 31, 2019 Tokyo, Japan
Data Circulation Overview

Collecting → ETL → Publication → Joint-Compute → Utilization → Destroying

Data Circulation Network

Anchor → Authenticate

Blockchain Network

Internet Companies

Big Data exchanges

Internet Court
Challenges in Data Circulation

Europe GDPR-ready by 25 May 2018

Data privacy is one of top 10 strategic technology trends.

Privacy protection and security are the biggest barriers of using big data. From CAICT

Privacy security is becoming a large obstacle in data circulation!
The Specific Issues in Data Circulation

Multi-level Data Protection

What we do

Privacy Security
DP, K-anony...

Data Security
Classification
Encryption
Isolating
Traceability
......

Platform Security
Secure transport, Secure storage, Secure computing, Audit, Secure Infrastructures

Data mask, Separating ownership from right-of-use

Malicious tampering, data traceability and access control

Easy to publish, pricing and exchange
Some Assumptions

- Right-of-Use
- Value built on data aggregation
- Semi-Honest
- Alliance Network
How We Build Our Platform

Data User
- Need high-quality and large quantity data to make decisions.

Data Provider
- Data Privacy protection
- Maximizing the value of data

Privacy Computation

Accessing VS Ownership

Data Platform
- data protection regulations executor
- Computing Engine Provider
- Coordinator: Seed source, etc.
Data Anchoring

Data fingerprints:
- Proof of Ownership
- Proof of Authenticity

TSA: Time Stamp Authority
Data Publication

1. Register
2. Extract data feature for DP
3. Encrypt data and data feature
4. Store encrypted data in Local DW
5. Store <fingerprint, account> return TXID
6. Store Meta in MetaServer

Definition Meta
- Data fingerprint
- Data provider's account
- Data Schema (structured data)
- TXID (Transaction ID)
Access Control Model by Smart-Contract

1. Identification: Blockchain Account as user account
2. Authentication and Authority by Smart Contract
3. Process recorded in the ledger

Decentralized Role based Access Control
- resource: a Machine Learning model or a 2-Dim table
Privacy Computation: TEE + DP

TEE: Trusted Execution Environment; Baidu-MesaTEE
DP: Differential Privacy; Elastic Sensitivity
Privacy Computation: MPC

MPC: Secure Multi-Party Computation; Baidu-PrivC

\[ f(x_1, \ldots, x_n) = y \]
Combine Verifiable Computing with Smart Contract

- Trusted Hardware
  - MesaTEE: Hardware as a Trust Third-Party

- Verifiable Computing
  - Verifying costs less than Recomputing
  - Interactive Game
    - TrueBit
  - Non-interactive Proof
    - SNARK, STARK

- Combine privacy computing with VC, e.g. HE
Application: Financial Risks Evaluation (KYC)

Cross-region Financial Big Data
- Financial Institutions
- Black Lists
- Credits records from TP
- Internet User Behaviors
- Judicial/Tax records
- Transactions

Authorization/Certification
Data Publish → Data → Tech

XuperData
App

Joint-computing
- LR
- GBDT
- NN
- SparkSQL
- More...

Financial Risks Data sharing Alliance network

Multi-Dim Labels
- ID attributes
- Payment ability
- Debt history
- Behaviors attributes
- Consumption preferences

Probability of overdue load
Q&A

- Our product:  https://xchain.baidu.com/case/xuperdata
- Connect me:  duanbing@baidu.com