NFTGDP: Non-fungible token and Geo Data Processing

By: Zhaoming Yin
stplaydog@gmail.com
Zhaoming Yin had a joyful undergraduate study in Hunan University learning Software Engineering. After that he spent 3 years in Peking University under the supervision of Shiwen Yu for NLP research. Then he headed for Atlanta at GaTech and earned his PhD degree in the area of combinatorial research for bioinformatics and graph analysis advised by Prof David A. Bader. Zhaoming got his first job at Oracle in the Bay area working on Golden Gate, a database replication software. Two and half years later, he moved back to China and lives in Hangzhou, and worked in Alibaba Cloud on ODPS – a big data platform. After a short stint at TRIAS lab as Chief Algorithm Architect and Senior Software Expert in PayTM labs, he currently works for Google on Android Automotive Embedded systems.
Data Privacy is becoming more and more important these days

➔ Euro GDPR
  summary of [GDPR fines](#)

➔ China DiDi Crackdown
didi [crackdown](#).

➔ US Privacy laws
  privacy [laws](#)
Data privacy is not just politics

➔ Should you driving data belong to the government?  
NO!

➔ Should you driving data belong to the Autonomous driving company?  
NO!

➔ Should your data belong to yourself?  
YES!
Limitations of centralized tech stack

➔ Tech companies promise they will treat your data nicely with their privacy ‘technologies’, but it turns out these technologies are more about self regulation, but lack of real crypto technology.

➔ Ask yourself: do you control any private keys on your data from: Apple, Amazon, Facebook etc?

➔ If no, then you don’t have control over your data
How data privacy is enforced in centralized systems

Admin

Ask for permission

grant permission

KV DB metadata

Consumer

Access data

Check permission

Distributed storage

User

No Access

No Access

No Access
Buyer

Install transferred model, and drive with seller’s experience.

Data company

Buy
Train
Model

Autonomous Driving company

Seller

List driving data and ask for a price
Solution

Insurance company

Charge insurance based on driver's record data.

Pay as you drive!!

Data company

Seller

List driving data and ask for a price
How data privacy is enforced in decentralized systems

Admin
- Ask for permission
- Grant permission

Consumer
- Access data

KV DB metadata
- Purchase data
- Check permission

Distributed storage
- Import data

User
- Generate data
- Data for sale

IPFS
OEMs

Platform

Apps

Data Companies

How?
System Architecture (Automotive)

Apps
- APP Backend
- DataBackend

Service
- generate data
- purchase data

Data Companies

Platforms
- Applications
- Platform + Hardware abstraction
- SOC
- Engine
- HVAC
- Sensor
- Steering

OEMs

APPs

Data
- APPs
- DataBackend

Platforms
- Data
- APPs

Companies
Data companies are just Smart contracts trading NFTs!
Block-chain

Fully Distributed Database like BTC

Properties:

• Highly Secure
• Transparent
• Immutable
- **Smart contract**

- Executable code
- Turing Complete (this is what smart contract different from BTC)
- A contract can hold fund and make the decision based on the contract logic (Yes, it's like a company!)
Smart contract

• Every node contains a virtual machine (similar to Java)
  ▪ Called the Ethereum Virtual Machine (EVM)
  ▪ **Compiles** code from high-level language to bytecode
  ▪ Executes smart contract code and broadcasts state

• *Every full-node on the blockchain processes every transaction and stores the entire state, so no centralized party decide the code execution result.*
Fungible token (ERC20)

- ERC20 is a token standard for create ethereum token contracts
  - ERC: Ethereum request for comments
- Any token contract adhere to specification follows ERC20, and most of ICOs are done on ERC20 (UNI, LINK, SUSHI, SHIB etc...)
Non-Fungible token (ERC721)

- Non-Fungible Tokens are not interchangeable and not divisible. These tokens are unique and cannot be replaced by another item.
- The ERC-721 Standard is the backbone for the creation and trading of non-fungible assets.
- Michael took Uber from Brooklyn to Long Island on 07/16/2021 at 7:00AM, this event will only happen once in the history, it's Non-Fungible
How?

**IPFS** is a public decentralized data storage system.

User encrypt data and put it into IPFS:

```
encrypted encrypt(public_key, trace_data, camera_data, sensor_data, operation_data, gps_data, .......) {.....}

ipfs_address send_to_ipfs(encrypted)
```
How?

user create an NFT and put the decryption key in the NFT and ask for a price.

```solidity
erc721.mint(ipfs_address, owner_address, ){.....}
```

Ethereum
How?

➔ Data consumer buys the NFT and get the decryption key

```python
erc20.transferfrom(buyer_address, seller_address, price) {...}
erc721.transferfrom(seller_address, buyer_address, NFT_id) {...}
```
How?

Data consumer use the decryption key to decrypt data and use it for model training.

```python
ipfs_address = eth21.tokenURI(NFT_id)
```
Challenges

- **complexity to implement**
  The complexity to implement is low, because Ethereum is in charge of every complexity.

- **potential failures**
  Every movement is irreversible, for example, if there is data leaks, no one can change back.

- **critical aspects**
  The data flow speed is the key, currently the public block-chain has the high latency and low bandwidth issues.

- **NFTs management**
  No human beings will be involved in the NFT management, everything will be controlled by smart contract.
Geo data representation in the public blockchain

Example: NYC taxi data, This data is about the taxi trip information in New York City. Now there are 4 different type of operators:

Green, Yellow, FHV trip, high volume FHV (uber, lyft)
Geo data representation in the public blockchain

Example: NYC taxi data format

<table>
<thead>
<tr>
<th>Vendor</th>
<th>dep_pickup_dat</th>
<th>dep_dropoff_dat</th>
<th>passenger_count</th>
<th>trip_distance</th>
<th>RatecodeID</th>
<th>store_and_fwd</th>
<th>PULocationID</th>
<th>DOLocationID</th>
<th>payment_type</th>
<th>fare_amount</th>
<th>extra</th>
<th>mta_tax</th>
<th>tip_amount</th>
<th>tolls_amount</th>
<th>improvement_su_total_amount</th>
<th>congestion_surcharge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2020-01-01 02:28</td>
<td>2020-01-01 03:30</td>
<td>1</td>
<td>1.2</td>
<td>1</td>
<td>N</td>
<td>239</td>
<td>239</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>0.5</td>
<td>1.47</td>
<td>0</td>
<td>0.3</td>
<td>11.27</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 03:30</td>
<td>2020-01-01 04:43</td>
<td>1</td>
<td>1.2</td>
<td>1</td>
<td>N</td>
<td>239</td>
<td>239</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>0.5</td>
<td>1.3</td>
<td>0</td>
<td>0.3</td>
<td>12.3</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 04:43</td>
<td>2020-01-01 05:53</td>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>N</td>
<td>238</td>
<td>238</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0.3</td>
<td>10.8</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 05:53</td>
<td>2020-01-01 1:00</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>N</td>
<td>235</td>
<td>1251</td>
<td>1</td>
<td>5.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.36</td>
<td>0</td>
<td>0.3</td>
<td>5.16</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 1:00</td>
<td>2020-01-01 1:04</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>2</td>
<td>3.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>4.8</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 1:04</td>
<td>2020-01-01 1:10</td>
<td>1</td>
<td>0.03</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>2</td>
<td>2.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>3.8</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 1:10</td>
<td>2020-01-01 1:16</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>1</td>
<td>2.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.01</td>
<td>0</td>
<td>0.3</td>
<td>3.81</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 1:16</td>
<td>2020-01-15:2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>1</td>
<td>2.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>6.3</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-15:2</td>
<td>2020-01-15:3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>1</td>
<td>2.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>6.3</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-15:3</td>
<td>2020-01-01 0:40</td>
<td>2</td>
<td>0.7</td>
<td>1</td>
<td>N</td>
<td>246</td>
<td>48</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>0.5</td>
<td>2.35</td>
<td>0</td>
<td>0.3</td>
<td>14.15</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 0:40</td>
<td>2020-01-01 1:12</td>
<td>2</td>
<td>0.7</td>
<td>1</td>
<td>N</td>
<td>246</td>
<td>48</td>
<td>1</td>
<td>12</td>
<td>3</td>
<td>0.5</td>
<td>1.75</td>
<td>0</td>
<td>0.3</td>
<td>17.05</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 1:12</td>
<td>2020-01-01 1:16</td>
<td>2</td>
<td>0.8</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>2</td>
<td>9.5</td>
<td>3</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>13.3</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 1:16</td>
<td>2020-01-01 1:21</td>
<td>2</td>
<td>0.8</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>2</td>
<td>9.5</td>
<td>3</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>13.3</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 1:21</td>
<td>2020-01-01 2:07</td>
<td>2</td>
<td>0.8</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>2</td>
<td>9.5</td>
<td>3</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>13.3</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 2:07</td>
<td>2020-01-01 3:00</td>
<td>2</td>
<td>0.8</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>23</td>
<td>24.15</td>
<td>0.5</td>
<td>0.5</td>
<td>4.15</td>
<td>0</td>
<td>0.3</td>
<td>24.15</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 3:00</td>
<td>2020-01-01 2:07</td>
<td>2</td>
<td>0.8</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>23</td>
<td>24.15</td>
<td>0.5</td>
<td>0.5</td>
<td>4.15</td>
<td>0</td>
<td>0.3</td>
<td>24.15</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 2:07</td>
<td>2020-01-01 2:30</td>
<td>2</td>
<td>0.8</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>23</td>
<td>24.15</td>
<td>0.5</td>
<td>0.5</td>
<td>4.15</td>
<td>0</td>
<td>0.3</td>
<td>24.15</td>
</tr>
<tr>
<td>1</td>
<td>2020-01-01 2:30</td>
<td>2020-01-01 2:30</td>
<td>2</td>
<td>0.8</td>
<td>1</td>
<td>N</td>
<td>193</td>
<td>193</td>
<td>23</td>
<td>24.15</td>
<td>0.5</td>
<td>0.5</td>
<td>4.15</td>
<td>0</td>
<td>0.3</td>
<td>24.15</td>
</tr>
</tbody>
</table>
User encrypt taxi request put it into IPFS

Ethereum user create an NFT and put the decryption key in the NFT and ask for a price.

Purchase data

Provide service

User encrypt taxi request put it into IPFS
IPFS data Storage

Add files to IPFS

```
ipfs add taxi_request -r
added QmdNMBVqQJ9baQDAJMUmwi8LzV2a52cTLEHoinaKFEdFY
taxi_request
```
Data privacy and data sharing in smart contracts are all decentralized, they provide protection of different granularity

- only owner of contract can access

abstract contract Ownable is Context {
    address private _owner;
    event OwnershipTransferred(address indexed previousOwner, address indexed newOwner);
    function owner() public view returns (address);
    modifier onlyOwner();
    function renounceOwnership() public onlyOwner;
    function transferOwnership(address newOwner) public onlyOwner;
    function _transferOwnership(address newOwner) internal;
}
Data privacy and data sharing in smart contracts are all decentralized, they provide protection of different granularity

- only owner of contract can access
- role based access control

```
abstract contract AccessControl is Context {
    function hasRole(bytes32 role, address account) public view returns (bool) {
    function getRoleMemberCount(bytes32 role) public view returns (uint256) {
    function getRoleMember(bytes32 role, uint256 index) public view returns (address)
    function getRoleAdmin(bytes32 role) public view returns (bytes32) {
    function grantRole(bytes32 role, address account) public virtual {
    function revokeRole(bytes32 role, address account) public virtual {
    function renounceRole(bytes32 role, address account) public virtual {
    function _setupRole(bytes32 role, address account) internal virtual {
    function _setRoleAdmin(bytes32 role, bytes32 adminRole) internal virtual {
    function _grantRole(bytes32 role, address account) private {
    function _revokeRole(bytes32 role, address account) private {
```
Data Privacy and data Sharing

Data privacy and data sharing in smart contracts are all decentralized, they provide protection of different granularity

- only owner of contract can access
- role based access control
- discretionary access control (access control list)

function tokenURI(uint256 tokenId) external view returns (string memory) {
    require(_msgSender() == ownerOf(_tokenId), "Only NFT owner can update");
    // return token URI address in IPFS
}
contract IGeoDataMarket {

  // meta data contains the IPFS info
  function create_token_NFT (  
      address currency_token,  
      string calldata meta,  
      uint256 price,  
      uint256 unit_size,  
      uint256 num_units  
  ) external ;

  function buy_token_NFT (  
      uint nft_id,  
      uint token_id  
  ) external ;

  function sell_token_NFT (  
      uint nft_id,  
      uint token_id,  
      uint256 price  
  ) external override;

  function redeem_token_NFT (  
      uint nft_id,  
      uint token_id  
  ) external override ;

}
Data scam detection and prevention

We can use Oracle to judge the correctness of data, such as Augur.
Data auction theory and implementation

Bounce protocol can be used for data auction
crypto kittens arbitrage