BlockX

A Secure, Decentralized Data Marketplace

George Chong, Joey Kaminsky, Lincoln Ma, and Su Aye

gchong@ucsd.edu, jkaminsky@ucsd.edu, lim007@ucsd.edu, suaye@ucsd.edu

Mentor: Sheffield Nolan

shieffieldnolan@franklintempleton.com

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

Overview

BlockX - A blockchain-based marketplace for data and code that addresses key data security issues by using decentralized storage (IPFS), automated validation (AWS Lambda), and secure, encrypted transactions between sellers and buyers, all facilitated by smart contracts on the Ethereum platform.

Concepts

Blockchain - A decentralized, distributed ledger technology that enables secure and transparent transactions without relying on central authorities. Built on cryptographic principles, it records data across a network of nodes, ensuring immutability. Transactions are validated through consensus mechanisms.

Ethereum - A decentralized blockchain platform that enables the creation and execution of smart contracts and decentralized applications through its Ethereum Virtual Machine (EVM). Ethereum's programmable blockchain allows developers to build trustless applications—such as financial tools, data marketplaces, and decentralized services—without intermediaries.

Smart Contracts - Self-executing agreements embedded in blockchain code, automatically enforce contractual terms when predetermined conditions are met.

InterPlanetary File System - A decentralized, peer-to-peer protocol for storing and sharing data using content addressing instead of traditional location-based URL. Each dataset is chunked and assigned a unique cryptographic hash called a Content Identifier (CID), enabling secure verification and retrieval of data from any node in the network.

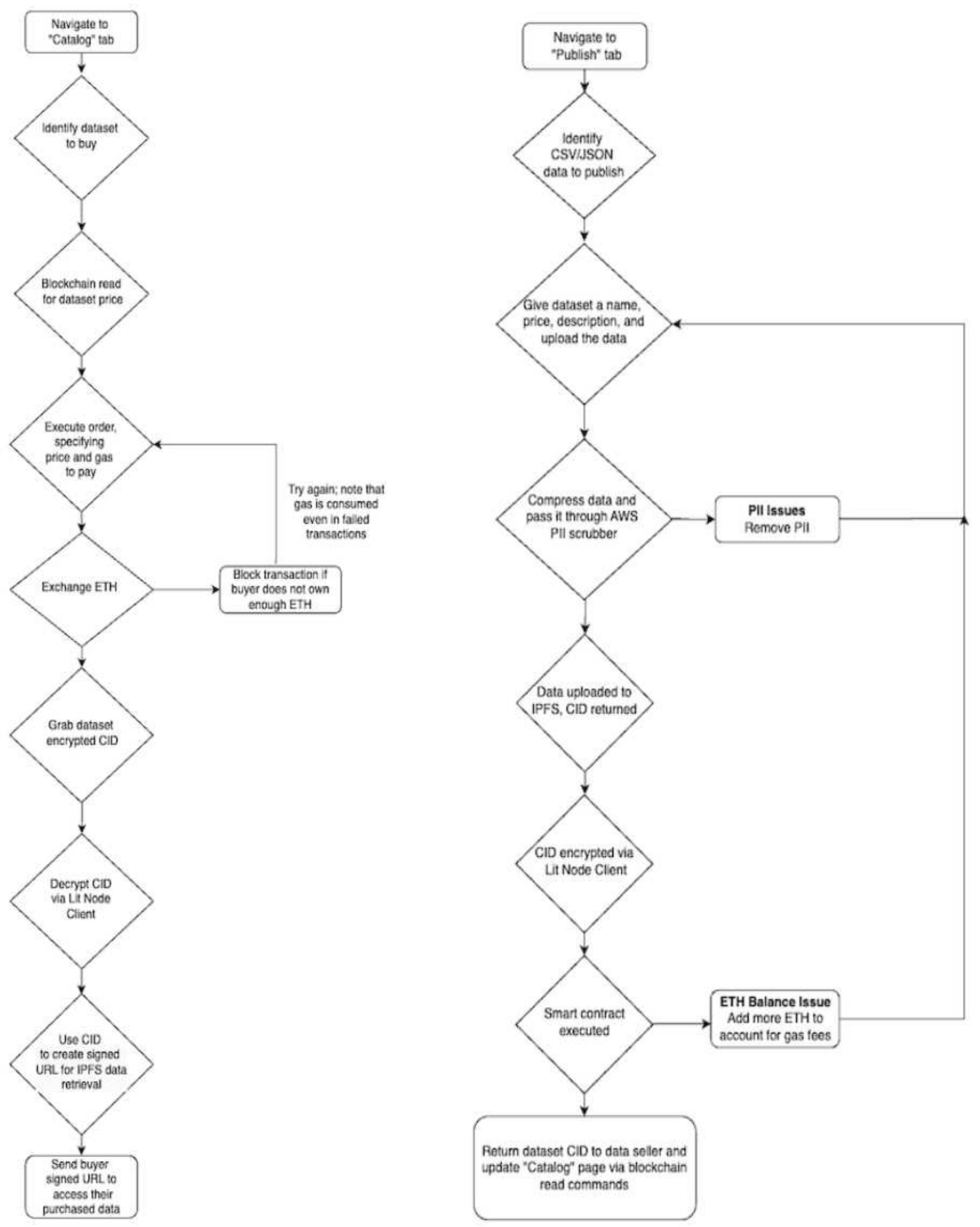
Problems with Current Marketplaces

Current data marketplaces face security and trust challenges that can lead to cyberattacks. These systems often fail to address the issues listed below.

- **Personal Identifiable Information (PII)**
  - Insecure transfers or regulations of **PII (SSN)**
  - Identity theft that can lead to significant monetary losses
- **SQL Injection Attack**
  - **Unauthorized access** of data
  - Data manipulation such as **alteration** and **deletion**
- **Centralized Storage**
  - **Single point of failure** due to server failure
  - **Lack of transparency** due to the hidden mechanism of Data Storage
- **Blind Trust of Marketplace Owner**
  - Marketplace owners may engage in malicious activities, like data theft
  - Major data security concern trusting a platform to host it

Our Solution

Buying and Publishing Data Flow Charts


Navigate to "Catalog" tab
Identify dataset to buy
Blockchain read for dataset price
Execute order, specifying price and gas to pay
Try again; note that gas is consumed even in failed transactions
Exchange ETH
Block transaction if buyer does not own enough ETH
Grab dataset encrypted CID
Decrypt CID via Lit Node Client
Use CID to create signed URL for IPFS data retrieval
Send buyer signed URL to access their purchased data
Navigate to "Publish" tab
Identify CSV/JSON data to publish
Give dataset a name, price, description, and upload the data
Compress data and pass it through AWS PII scrubber
PII Issues
Remove PII
Data uploaded to IPFS, CID returned
CID encrypted via Lit Node Client
Smart contract executed
ETH Balance Issue
Add more ETH to account for gas fees
Return dataset CID to data seller and update "Catalog" page via blockchain read commands

Dependencies

- **AWS** - Reduces server cost, and automatic data processing and validation which ensures data privacy
- **MetaMask** - Smart contract deployment and eusure the transactions has been authenticated by users.
- **Pinata** - Decentralize storage that could prevent data leakage. Data would also be encrypted before uploading to IPFS
- **Solidity** - Create self-executing contracts, allowing for the automation of processes and transactions on the blockchain

Leveraging trustworthy services ensures a secure data exchange platform while enabling future scalability

Data Scrubbing

- Scrub data during publishing for **25+ SQL injection patterns** and **PII markers** (names, emails, SSNs) using Natural Language Processing
- Rejects datasets with **3+ low-risk PII** types (name, age, etc.) or **1+ high-risk PII** (SSN)
- Charges a **0.0025 ETH penalty** after 3 failed upload attempts to deter abuse of smart contract

Benefits of Our Marketplace

1 Decentralized Trust Architecture

- Stores data on **IPFS** with AES-256 encryption, tied to wallet addresses
- Processes transactions via **EVM smart contracts** for tamper-proof logging
- Purchased data is accessible via the buyer's public address to prevent redistribution

2 Cost-Efficient Validation

- **AWS Lambda** used for serverless PII/SQL checks, reducing computation costs by 72%
- Maintains **sub-minute transaction speeds** despite encryption/validation layers

3 Transparency Mechanisms

- Open-source smart contracts and validation code
- Public blockchain records of all data publishes and purchases

Next Steps

- Develop solutions to allow AWS to handle data bigger than **10 MB**
- Add additional security measures to prevent redistribution and data upload scams
- Deploy on the **Ethereum Mainnet** and be used for real transactions among trusted parties
- Add additional functionality (rating, sorting, and searching) to improve the user experience

Acknowledgements

We would like to express our deep appreciation to our mentor, **Sheffield Nolan**, Professor Bellur, and TAs for their guidance and support throughout this project.

QR code to Website