Enabling the Dissemination of High-Value IoT Data Through Decentralized Auctions on the Blockchain

3/21/2018
Austin Hwang, Alex Iansiti, Hari Kothapalli, Nathan Lee
Harvard College
Cambridge, MA 02138
austinhwang@college.harvard.edu, aiansiti@college.harvard.edu, hkothapalli@college.harvard.edu, nlee01@college.harvard.edu

***Abstract*—We aim to target high-value IoT device data generators and consumers by implementing two blockchains: one for maintaining device metadata records and another for running auctions for month-long access keys to a device's data. The datachain involves the participation of individual IoT data generators, IoT data consumers, and blockchain miners; each party is economically incentivized by monetary/cryptocurrency compensation. We plan on implementing an API specification for checking a given IoT device into the blockchain and creating a metadata ledger by using Python, Ethereum blockchain, and Solidity respectively. A React web application will handle the frontend process of filtering data and make batch auction requests.**

I. Introduction

While over 75 billion connected internet devices are projected to create 163 zetabytes of data by 2025,, most of the real time data produced by these devices today is either closed off, wasted, or exploited by technology conglomerates—data ranging from analytics and sensor data to personal and computational data. For example, Amazon has pioneered the connected home through its Alexa platform in the past few years, giving the corporation sole access and control to every piece of data generated by the smart devices' users. Today, these types of smart devices know the identity of each user and the personal data they generate through their behaviors and actions—but without giving these users direct compensation for the increasingly valuable data they freely provide. This scenario is not just unique to Amazon; present-day technology corporations often mine and exploit mass amounts of free user data.. To combat this growing trend, Mozilla introduced a proposal for a Web of Things open protocol for IoT devices to communicate with each other. Besides breaking out of these closed ecosystems, an open IoT protocol opens many doors in the way of data democratization.

Previous, similar work that has been done in this field includes the IOTA and Enigma data marketplace. The IOTA data marketplace serves to solve the problem of closed data ecosystems by offering secure, live data streams of different sensors, such as temperature and energy sensors, local weather data sensors, and humidity sensors to a compensating third party, such as data scientists and researchers. Along a similar vein is the Enigma data marketplace, which provides open data sets for trading and development activity for the various cryptocurrencies, including technologies such as Slack, Github, and Telegram activity and market cap values. The IOTA and Enigma data marketplaces are two examples of an alternative approach to accessing data beyond the closed ecosystems of technology corporations.

But while IOTA and Enigma focus on industry information and research-centric sensor data, we hope to target more personal user devices like smart fridges and smart TVs, expanding on Mozilla's vision with Web of Things to help small developers and startups gain access to previously-restricted data while anonymizing and compensating the users generating the data itself. With IoT devices projecting to have an estimated impact of 11 trillion dollars a year, open source data will become more and more valuable and should be secured through a decentralized data marketplace that protects and compensates its users' data.

II. Proposed Approach

To combat the existing practice of data centralization, we propose a method of interacting with individual IoT devices to exchange data. Using this per-device approach allots each data scientist ultimate freedom in choosing a custom dataset that best suits his/her needs. Other solutions would rely on per location or per SKU aggregation, which may not fit the needs of all data scientists who require finer control of their samples.

Then, to maintain this ecosystem of device-level exchanges, we require the implementation of two blockchains, one for maintaining device metadata records and another for running auctions on month-long access keys to a device's data. Our approach requires blockchains, as opposed to a central authority, because blockchain technology allows for an immediate standard of trust not possible in a centralized marketplace. This trust is maintained because a data scientist must possess the funds necessary to acquire a dataset in order to participate in an auction, and the deducted fees are fair and open.

In handling commerce between as many as three parties—the data scientist, the IoT device user, and a miner—each member must have his/her own economic incentives to participate in the open ecosystem. The data scientist is incentivized to participate because the type of complex data available from smart appliances is nearly

impossible to acquire elsewhere outside of the existing closed ecosystems of technology corporations. Incentivizing the user of the IoT device provides a distinct challenge because some data prove too inexpensive to incentivize participation at all, e.g. temperature sensor data, which may be only valued at a few dollars for relatively large quantities of data. In order to incentivize the users of the IoT devices to opt into this service, we plan on targeting "smarter" devices with higher valued data, such as the behavioral data generated by aforementioned smart appliances, which may be valued at several dozen dollars per user per year. Choosing to target IoT devices that generate high-valued data allows the blockchain to reach critical mass through specific use cases yet remain general enough to allow for more devices to check in and distribute data. Once a data consumer wins a bid for data, they must send the required amount of cryptocurrency to fulfill the smart contract. If successful, the key to access the IoT users' data is automatically sent back. If not, then the smart contract is fulfilled, no transaction is made, and the data goes to the next highest bidder. Lastly, we require third-party miners to maintain the ledgers themselves, incentivized by compensation through attaching transaction fees onto the auctions.

III. Experiments to be Performed

We propose a novel framework to allow data scientists to collect data from these IoT devices using an API specification and a blockchain. The API guidelines serve as a common ground on which a client can interact with any IoT device, in accordance with Mozilla's Web of Things open protocol. For this implementation, though, the guidelines pertain specifically to starting an auction and the resulting data exchange. Given an open API, any company can build their own frontend for data scientists to acquire datasets according to their own needs.

As such, it would be quite inefficient to interact with IoT devices individually, so we propose the creation of per-use-case front-ends which batch auction requests and bids to many devices at once. Such front ends would read in a ledger of device metadata and allow data scientists the freedom to apply whatever filters they wish, set the max bid price of a data point, then spit out requests to the desired devices.

To collect and maintain this device metadata, we will employ a blockchain as an open ledger containing device metadata as mandated by an OEM. A user would need to manually enable a device to check into this blockchain, enabling it to accept auction requests.

Lastly, this ecosystem requires a blockchain to manage and maintain the auctions between the data scientists and the devices.

A successful implementation would allow a consumer to purchase an key to access a device's data for one month. This should be achieved by interacting with our open API through an open-source frontend that we will build out. Inherent in this is the ability to disseminate IoT data, from a device that follows Mozilla's proposed protocol, through the blockchain.

IV. Implementation Plan

To demonstrate the necessary frontend for interacting with the IoT devices we will in the first couple weeks create a React web application which allows data scientists to view and filter checked-in IoT devices and then make batch auction requests to them.

To build the necessary blockchain to maintain device metadata, we will next create an API spec for checking the device into the blockchain and implement it on a faux IoT device using Python. The metadata ledger will be created using the Ethereum blockchain and Solidity. As this piece of the framework relies the most on trusted third parties, the intricacies of the blockchain will be easy to implement.

Finally, the blockchain which runs and manages the auctions will be written in Solidity in the last few weeks of work. The auctions will need to manage pricing of a device's single-use API key and its exchange. In addition, this blockchain will have smart contracts tied to the the data consumer's money and the IoT devices' data during an auction cycle to ensure they have the funds to pay out to a device upon winning. This piece of the framework is completed by another API specification for the sending of the API key and the resulting data exchange. This aspect of the spec will also be built into our Python reference implementation.

V. Anticipated Results

Anticipated results are twofold: first, proving the ability to use the described API spec to create and process a data transaction, with a user interacting with the frontend to request batch datasets with the auction blockchain to acquire access to the device metadata blockchain as described in the proposal; second, creating a business model for the data marketplace that gives evidence for the viability of economic incentives of all three parties to participate and maintain the blockchain. The conclusion of the project seeks to demonstrate the viability of this type of IoT data marketplace that specifically targets high-value smart device data transactions.

Related Research Papers

Tangle is different from blockchain in that it uses a Directed Acyclic Graph to maintain a ledger for transactions. Each edge set of the tangle must verify and approve two previous transactions so that in order to issue a transaction, users much help maintain the integrity and security of the Tangle system. https://iota.org/IOTA_Whitepaper.pdf

A peer-to-peer network, enabling different parties to jointly store and run computations on data while keeping the data completely private. Enigma's computational model is based on a highly optimized version of secure multi-party computation, guaranteed by a verifiable secret-sharing scheme. https://enigma.co/enigma_full.pdf