Towards a Decentralized Data Marketplace for Smart Cities

Gowri Sankar Ramachandran, Rahul Radhakrishnan and Bhaskar Krishnamachari
Center for Cyber-Physical Systems and the Internet of Things
Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA 90089
Email: {gsramach, rahulrad, bkrishna}@usc.edu

Abstract—One of the ways in which a city can become smarter is to grow a local economy around the sharing of data from IoT devices and other open data that can be used in applications to improve the lives of its citizens. Prior work and ongoing projects have examined or are currently focused on the development of centralized data marketplaces for smart cities. Here we explore how a decentralized data marketplace could be created using blockchain and other distributed ledger technologies. We consider the possible benefits of such a decentralized architecture, identify different elements that such a decentralized marketplace should have, and show how they could be potentially integrated into a comprehensive solution. We also present a simple smart contract implementation of a decentralized registry where data products can be posted by data owners for retrieval by potential buyers.

Index Terms—Decentralized Data Marketplace; Blockchain; Smart Cities

I. INTRODUCTION

We posit that a smart city is one that can continually improve the lives of its citizens through a wide range of efficient data-driven services ranging from transportation and safety to health and sanitation. These services might be provided by the municipal government and also by companies harnessing both the entrepreneurial spirit and the community know-how of the local citizenry.

In order to enable a rich set of data-driven services [1], [2], a number of different types of IoT devices including various types of sensors, cameras, may be deployed in the city; other sources of data for these services may include human inputs such as surveys and mobile application-based requests or automatically crowd-sourced data from mobiles and vehicles, such as GPS location information [3], [4].

Now, there are several models under which such data sources could be owned and operated. They may a) be wholly owned and operated by a city government; b) they may be deployed and operated/managed by a single organization for each application (that organization may be a company that sells its services directly to customers or the government); or c) they may belong to a heterogeneous group of owner individuals or organizations. In the last case, which is of interest because it is intuitively the most scalable, because the interactions start to cross trust boundaries, there may not be an inherent incentive for parties to exchange data with each other. Instead, there starts to be a need for data consumers (which may be application developers or service providers) to incentivize and compensate the data owners for sharing data with them.

Given the economic value associated with IoT data from different parties in this context, it becomes essential for parties involved to sell, find and buy the data as easily as possible. Such an issue gives rise to the need for IoT data marketplace platforms in smart cities. In recent years, such an idea has been independently proposed and started to be explored by several researchers and practitioners [5]–[9].

One example of a real-time IoT data marketplace is the work being done by the I3 consortium led by the University of Southern California [7] to establish such a marketplace. I3 aims to provide a real-time data marketplace in which there is a single website that stores and retrieves the information about all sellers and their data products to be extracted and shown to buyers as needed. In its prototype form, the I3 domain controller for a single community is designed using a centralized pub-sub broker, with the authorization to data buyers (subscribers) being turned on/off based on their payments and acceptability (concerning trust) to the data sellers (publishers). A fundamental principle embodied in I3 is that data owners have the right to decide whether, whom and when to share their data.

While I3 and other efforts we have encountered to date in building a data marketplace for smart cities have all been centralized, we are motivated to explore the possibility of decentralized data marketplaces for several reasons. First, it is interesting and novel from an architectural perspective to determine what it would mean to make a data marketplace decentralized - what elements it would contain, and how they could be developed without centralized servers or trusted third parties, as there are no such systems in existence and operation today (that we are aware of). Second, a carefully designed decentralized architecture for a data marketplace could potentially avoid accusations that the operator of the marketplace is biased and recommending or prioritizing sellers that are “bribing” it, creating an undesirable “pay for play” environment [10]. Operators of such platforms have significant monopoly market power over how data and products are advertised and presented to users on the platform and the potential exists for them to abuse this power through distorted rankings. On the other hand, using decentralized trust mechanisms such as storing and retrieving meta-data and ratings from a Blockchain can potentially improve the transparency and trustworthiness of ratings. Another reason to explore the design of an open decentralized marketplace is...
research that shows that controlling which buyers can access which sellers (more likely to occur in a closed centralized marketplace) results in market inefficiencies [11]. Finally, the use of cryptocurrencies to enable seamless micro-payments for data allows for more flexible, even shorter-term interactions (though it should be admitted that this is not novel or equivalent to decentralization in itself as prior efforts have already advocated the use of cryptocurrencies for centralized data marketplaces [8], [9]).

Our goal in this work is to begin to explore how such a decentralized data marketplace could be designed and deployed using Blockchain and other distributed ledger technologies. We identify some of the key challenges inherent in achieving the big goals we lay out, show what the architecture and components of a decentralized data marketplace should be, and share implementation details of a smart contract that can provide a search-able decentralized registry of sellers and their data-products. We also put into the context of this work a recent prior work from our lab on developing a technology-agnostic streaming data payment protocol (SDPP [12]), that could be used as the peer to peer data-for-value exchange protocol in the proposed architecture.

The rest of the paper is structured as follows: Section II presents the elements of a decentralized data marketplace. The implementation of a decentralized data marketplace is discussed in Section III. Section IV concludes the paper.

II. ELEMENTS OF A DECENTRALIZED DATA MARKETPLACE

We hypothesize that the following are some of the key components needed in a decentralized data marketplace:

A. Buyers and Sellers

We assume there are a set of sellers of data in a city with one or more static or streaming data products (these could be data coming from sets of sensor devices measuring anything from air quality to building occupancy), and buyers interested in their data that are the fundamental participants in the decentralized data marketplace.

B. Posting and Discovery

One of the crucial questions for a decentralized marketplace is how to create a method for buyers and sellers to find each other, and for buyers to get to know what types of relevant data products are available, without merely posting them on a single website (the traditional centralized approach to creating a marketplace platform). While P2P networks tried to address this very issue in research and developments focused on search and querying in unstructured overlay networks [13] as well as indices based on distributed hash tables in structured overlay networks [14] — they typically do not account for incentives and potential lack of trust between the parties involved. We present in the following sections an implementation of a smart contract to allow sellers to register themselves and post data products, that buyers could search to find relevant information. In our proof of concept implementation, the seller posting is essentially a hashed pointer that is on the blockchain to a description stored on an off-chain distributed file storage system (specifically, IPFS [15]). Buyers can use this pointer to obtain metadata describing the data product (sensor type, including the address of the server to connect to, pricing information, and so forth).

C. Meta-Data Organization

A related question is whether the meta-data describing data sources provided by each data seller should be organized hierarchically or tagged in some way, or left unstructured. Example of hierarchies that are important for sensor data include: spatial hierarchies (Lat-Long, Building, Floor, and Room), sensor type (Analog - Climate - Temperature), organizational (Government - City - Agency - Division) based on ownership of data source. It is possible to use standards like SensorML [16] or OCF [17] data models to determine a standardized way to organize the sensor information. In general, in light of evolving standards, it is best for a marketplace to be able to support any standardized JSON format for metadata to simplify search and discovery.

D. Data Transfer and Payments

Once a buyer learns about a seller and how to connect to it to get the data it is interested in, it can use a designated protocol to collect that data in exchange for payment or other incentives. One such protocol that is suitable for real-time data with micropayments is the streaming data payment protocol (SDPP) [12], which uses a TCP client-server connection for data transfer, a cryptocurrency based payment channel and a distributed ledger-based channel to store key transaction records such as invoices and receipts (detailed further in section III-E). In ongoing work, we are also exploring ways to connect publish-subscribe brokers with crypto-currency based micro-payments, which would allow multiple buyers to subscribe simultaneously to data from the same seller or even multiple sellers.

E. Data Quality - Ratings

While a small data marketplace could be browsed in its entirety by a buyer to determine which seller and data products are relevant, such an approach doesn’t scale to larger marketplaces. Users need a way to assess the quality of the data product provided by a given seller. One approach to this is to allow buyers and sellers to rate each other. The benefit of storing such ratings on the blockchain would be that they cannot be falsified or tampered with. As part of the rating mechanism, sellers could give some indication on the buyer’s reliability of paying for what was delivered (if an escrow mechanism is not used to ensure that), and the buyer could indicate the quality and timeliness of the data received. This could be done in a decentralized manner using smart contracts, potentially with additional objective verification based on the transaction records that the buyer and seller in question did indeed interact with each other. It could also be based on summarized statistics of the interactions such as frequency of...
interactions, amount of data purchased, etc. The information about these ratings would then be available for future potential buyers and sellers to use to screen each other, or for automated decentralized applications to recommend products to buyers.

**F. Data Quality - Curation and Recommendations**

A different approach to the problem of helping buyers find good quality, reliable sources of data in a marketplace is to create one or more *token-curated registries* to curate a set of recommended data sources for a given city. The idea of token curated registries is somewhat recent, most prominently implemented in the context of Digital Advertising, a system called AdChain [18]. In a token curated registry (TCR) a candidate must stake some tokens to apply for membership in a list, and the set of token-holders can vote to challenge or accept that candidate. The thinking behind TCRs is that token holders have an economic incentive to maintain a high-quality list - because it would increase the value of the tokens they hold. In our context, one or more such TCR’s could be developed to help buyers find good quality and reliable sellers to meet their data needs. We believe it may also be possible to extend the idea of token-weight voting from maintaining lists to voting on recommendation engines that give useful and objective recommendations to buyers.

**G. Trust in Buyers**

Sometimes it is not the buyers that need input on which sellers are reliable. Certain owners of sensitive data may be cautious about whom they share their data with and for what use. TCR’s and ratings of buyers can also help such sellers understand whether their buyers are trustworthy in terms of who they are, what they do and how they would handle sensitive data.

**H. Access-Control and Identity Management**

It may be necessary to add an identity layer for role-based access control — sellers may want to restrict who can buy or need some agreement that is digitally signed by the buyer that is legally enforceable. Likewise, the buyers may need some guarantee that a given seller is whom he/she claims to be. We believe it should be possible to use a blockchain-based decentralized secure identity platform such as Civic [19] for this. There may also be the need for federated identity and access management (IAM) systems that allow individuals within each buyer/seller organization to be able to exercise relevant access rights.

**I. Security**

In our architecture, the product, ratings and curation systems are decentralized using blockchain and inherit its security properties in terms of lack of centralized points of failure or a single target server for DDoS attacks, distributed trust, and immutability. The data exchanged end to end might need to be encrypted with assurance about integrity, possibly using transport layer security (TLS), but this is assumed to be handled by the individual protocol chosen by the seller to provide data in exchange for payments (such as SDPP). The use of secure identity and federated identity and access management systems should be carefully structured to add additional layers of trust and security.

**J. Privacy-sensitive Data Management**

Because of new legal regulations such as the General Data Protection Regulation (GDPR) in the European Union [20], and the California Consumer Privacy Act in the USA [21], it is essential to ensure that privacy-sensitive data is not leaked on the marketplace. Most importantly for us, as we advocate the use of blockchain technologies for decentralization, it is important to ensure that privacy-sensitive data is not logged on to an immutable ledger. One way to ensure this is to keep data that is exchanged on separate, private, encrypted data channels between buyers and sellers. Another is to try and put any sensitive meta-data stored on the blockchain only in

---

**Fig. 1. Key Elements of a Decentralized Data Marketplace.**
the form of hashed pointers to actual data that is stored off-chain, which could then be more easily modified or deleted if necessary. The careful deployment of access management tools could also help with privacy by ensuring that only authorized users are allowed to access certain raw data and meta-data, while others are provided data that is anonymized, possibly using techniques such as differential privacy.

III. IMPLEMENTATION OF DECENTRALIZED DATA MARKETPLACE

We have developed a proof-of-concept implementation of such a decentralized data marketplace, focusing on the product smart contract and querying components (in addition to SDPP for data transfer and payments, which was previously implemented [12]). Figure 2 provides the functional overview of the decentralized system. The end-to-end interaction process is described below:

1) The seller is required to generate a detailed product description, including the type of data, seller identification, price, IP address, etc. in JSON format. This product information is stored in a distributed file storage (DFS) framework such as IPFS [15] or Storj [22].

2) When the product information is successfully stored in DFS, a storage identifier is returned by the storage framework. The seller has to register the product using this product string.

3) The seller has to store the product metadata including the storage identifier and the protocol type to a blockchain, as indicated by Step 3 in Figure 2.

4) All the sellers are required to post product metadata in the blockchain following steps 1 - 3. The buyer would browse the product catalog by querying the blockchain. Besides, the buyer is required to parse the catalog by retrieving the product description from DFS. This approach increases the processing overhead on the buyer-side, but it reduces the overall cost as the storage and transaction costs associated with on-chain computation is typically quite expensive.

5) Steps 5 and 6 in Figure 2 show the interaction with DFS for retrieving the product information.

6) The buyer is required to select the desired product after analyzing the information collected from the blockchain and the DFS framework. To purchase a product, the buyer utilizes the streaming data payment protocol (SDPP).

7) The data exchange between the seller and the buyer in SDPP happens off-chain with the payments happening via a cryptocurrency and key transactions being recorded in a ledger.

8) After the end of the data transaction, both the seller and the buyer can rate each other. Independently, token-curated registries may also help recommend quality data sources and trust-worthy buyers.


A. Storing product description as IPFS

IPFS, which stands for InterPlanetary File System, is a peer-to-peer file storage platform. It is a decentralized platform for storing files, including web pages, images, text documents, etc., among the peers in the IPFS network. When the file is successfully stored in IPFS, the user will receive a hash-index, which would allow the user to retrieve the file in the future. For the decentralized data marketplace, we provide a JSON template for the user to describe and store the product description in the IPFS. The product template is made available as an open-source software here: https://github.com/ANRGUSC/DDM. The process for storing the product description in IPFS is described below:

- The seller has to install the IPFS software in his/her machine.
- When the IPFS software is successfully installed, the IPFS daemon should be started using ipfs daemon command.
- The seller has to describe the product using the marketplace template provided here: https://github.com/ANRGUSC/DDM
- Once the product is described, the seller has to add the JSON file to the IPFS using ipfs add filename command. This command will return the hash-index associated with the files. The hash-index is an alphanumeric string, which is used to access the files stored in the IPFS storage.
- The buyer can retrieve the product description using client libraries provided by IPFS.
- To advertise the data product, the seller has to enter the hash-index and the protocol-type to the blockchain (see Section III-D)

Figure 3 shows the product description stored in the IPFS for a product that sells temperature sensor data.

B. Product Description Template

The buyers select a data product by browsing through the list of available product in the marketplace. To allow the buyers to decide on a suitable product, the seller has to provide sufficient information about the product. We believe that the description provided in Figure 3 would be adequate for the buyer to choose the desired product. The type of data, location tag, price, and the rating of the seller are critical for the buyer to validate the usefulness of the data. When the seller is concerned about his/her privacy, he/she may not be willing to share information such as the location or the type of the sensor. In such cases, an off-line storage mechanism with privacy control may be desired. We will address this problem in our future work.

C. Marketplace Data Format

Operations involving computation and storage operations in the blockchain network comes at a cost, as the transaction fees given as a reward for block miners. The cost associated with the transactions and the contract must be minimized to reduce the cost involved in using a decentralized data marketplace. We propose a custom short data format for storing the meta-data into the blockchain as shown in Figure 4.
As shown in Figure 4, the first 4 bits represent the type of the streaming protocol, and the remaining 256 bytes represent the hash-index for the product (see Section III-A).

### D. Marketplace Smart Contract

The marketplace smart contract allows the seller to register his/her product in the blockchain. Our prototype implementation is built on the Ethereum blockchain. We used Ethereum’s solidity programming language [24] to create our smart contract. The two key elements of smart contracts are: 1) a data structure to store the hash-index and the protocol type for each of the data products, and 2) functions to store product information to the blockchain. We release our data marketplace smart contract as an open-source software here: https://github.com/ANRGUSC/DDM.

The smart contract was deployed on an Ethereum testnet using Remix IDE, which is an Ethereum tool for distributed application (dApp) developers. Note that the deployment of a contract or any type of transactions in Ethereum blockchain incurs a transaction fee. In Ethereum blockchain, the costs are estimated using a unit called *gas*. The miners in the system typically determine the price of the gas. A transaction with a lower gas price may take very long since the miners prefer to handle transactions that offer a high incentive. The seller
TABLE I

Transaction Costs for Deploying the Contract and Storing Product Descriptions. At the time of writing this paper, 1 ETH was equivalent to 263.79 USD.

<table>
<thead>
<tr>
<th>Contract Name</th>
<th>Transaction Cost (in Gas)</th>
<th>Transaction Cost (in Ether)</th>
<th>Price in USD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DecentralizedMarketplace</td>
<td>50 GWEI</td>
<td>0.01138435</td>
<td>$3.00</td>
</tr>
</tbody>
</table>

is encouraged to assess the transaction trends using platforms such as https://ethgasstation.info/ to select the right transaction fee.

Due to the transaction fees associated with Ethereum transactions, the seller is required to own Ether. MetaMask allows the sellers to sign the transactions using a web-based Ethereum application securely. Table I shows the cost of deploying the smart contract and storing product descriptions in the blockchain.

E. Streaming Data Payment Protocol

The Streaming Data Payment Protocol [12] (SDPP) is a recently proposed application-layer protocol that allows for the bidirectional flow of real-time data and payments. As can be seen in figure 6, it consists of three parallel channels. The first is a traditional TCP socket-based connection for a client-server application, which allows the server (seller) to send a menu of data products (with pricing information) and then the data stream itself when the client places an order (request for a particular data product). The second is a payments medium, that can be implemented using any cryptocurrency allowing the flow of micropayments for every \( K \) units of data (\( K \) is a granularity parameter that can be set by the server). The third is a records medium which can be implemented using any distributed ledger technology that allows the storage of essential transaction records such as the order, invoice, and payment receipts, in an immutable, auditable fashion. While SDPP is agnostic as to which cryptocurrency and distributed ledger technology are used, our open source reference implementation (see https://github.com/ANRGUSC/SDPP) is implemented using IOTA [25]. In the context of this work, we propose to use SDPP as an example of a data transfer and payment protocol that could be used with the decentralized data marketplace.
F. Ratings Smart Contract

Another element of the decentralized marketplace that can be realized as a smart contract is the rating of buyers by sellers and vice versa. In the simplest form this smart contract logs a score along with information on who is rating whom. In a sophisticated implementation, there could be some process that allows only buyers and sellers whose interaction can be validated through the records medium in SDPP to rate each other, and also provide additional textual comments that could be stored off-chain in a distributed file storage system like IPFS. The ratings would be stored in an immutable manner on the blockchain. A query over the ratings recorded in this smart contract can provide a histogram of ratings for a given seller or data-product or buyer.

G. Proof-of-Concept Demonstration

Figure 5 shows the screen-shots and the console output of the demonstration of our proof-of-concept decentralized data marketplace. Screen A in Figure 5 shows the output of the querying application executed by a seller. The latest transactions sent to the product smart contract along with their costs in Ether are shown in Screen B. Screen C presents the graphical user interface of EtherScan IDE for registering product information to the Blockchain. Besides, the seller can register the data product using NodeJS client libraries.

IV. DISCUSSION AND CONCLUSIONS

In this work, we have presented the motivation for establishing decentralized data marketplaces for smart cities. We have discussed the architecture and some essential components for such a marketplace system including ways for sellers to post product data, buyers to find relevant data sources, for buyers and sellers to exchange data for value, as well as various trust, recommendation and identity mechanisms. We have also presented some preliminary work showing how key elements of such a decentralized marketplace could be implemented using smart contracts.

There are many challenges to be addressed in advancing this kind of a decentralized marketplace. These include the following:

- **Managing system complexity**: with many possible moving parts and components the decentralized marketplace system could get fragmented or hard to scale - how to prevent this? On the other hand, keeping too rigid a framework could also hurt scalability.

- **Economic incentives and centralization**: Without sufficient economic incentives for all concerned parties it is possible that the decentralized marketplace will not function effectively. On the other hand, it may be necessary to design the incentives carefully so that it doesn’t lead to more decision-making power and data getting concentrated into the hands of certain parties that may distort the search, rankings, and recommendations in unfair ways.

- **Applications and interfaces**: A thriving marketplace will require user-friendly interfaces and decentralized applications (dapps) that make it easy to post, browse, search, rate, recommend, curate, verify, data products, sellers, and buyers. There should be sufficient incentives for people and organizations to develop and provide valuable and user-friendly applications while keeping them as decentralized.

In ongoing work, we are trying to refine the various elements of the architecture presented in this work and to build a working prototype that assembles all the pieces into a functioning system. Our open source code related to this project will be posted online at https://github.com/ANRGUSC/DDM.

ACKNOWLEDGMENT

This work is supported by the USC Viterbi Center for Cyber-Physical Systems and the Internet of Things (CCI).

REFERENCES


