Decentralized marketplaces for IoT data

Author: Hans Sluijter[1] , **Supervisor: Miray Ayşen**[1] , **Responsible Professor: Zekeriya Erkin**[1]
Cyber Security Group
Department of Intelligent Systems
[1]TU Delft

Abstract

An increasingly large amount of data is generated on the IoT. Widespread use of this data may help provide valuable insights and actionable triggers for new innovative services. To tap the potential of this increasing supply of data, a platform is needed. However in creating marketplaces for this data concerns regarding privacy and misuse of data are raised. This paper analyses several technologies that leverage the benefits of blockchain to attempt to mitigate these concerns. Mechanisms regarding data transfer, registration contracts and the responsibilities of gateway nodes are considered. The benefits and drawbacks of these technologies is discussed, as well as potential mitigations for these drawbacks. Finally the paper proposes a set of guidelines for future iterations of such platforms.

1 Introduction

The "Internet of Things" (IoT) is a quickly growing network of devices that can communicate without the need of human involvement [1]. These devices tend to contain sensors, and generate an enormous amount of data every day. A lot of this data gets used for the single purpose the device manufacturer intended, then discarded or archived. Much of this data is useful for far more people, a fitness app could purchase weather data for instance to suggest pollution free running routes. In connecting the buyers and sellers of this data a platform is required. However the idea of data as a marketable asset is quite new, and as outlined in [2], there are ethical concerns regarding privacy and misuse of data in such platforms. Therefore this paper decides to focus on the potential of decentralized mechanisms to address these issues and studies existing implementations to determine to what extent they achieve this.

In order to tackle the research question of 'What are the drawbacks and benefits of currently existing blockchain technologies for IoT data marketplaces?incentivize malicious behaviour from data producers.', the following questions will be addressed:

1. What challenges are there when implementing blockchain on IoT devices?
2. How to sell IoT data, and how can blockchain play a role in this?
3. What methods exist to monetize IoT data using blockchain?
4. How can these methods be improved?

An overview of the technologies will be given in the background section. Several frameworks have been proposed for doing this, which I will review in the related works section. In the analysis section I will discuss the potential benefits and drawbacks of such solutions. The findings of this will be summarized in the results section. Then I will reflect on the ethical aspects of my research in the responsible research section. The discussion will then debate the reproducibility of these results and how they compare to those of others. And to close it off I will make suggestions for improvements to these technologies in the conclusions and future work section.

2 Related Works

The idea of using blockchains to create a platform for the monetization of IoT data isn't new, and has been outlined in a number of papers in the past.

In 2016 Opher et al. [3] published a paper addressing the need for companies in the data industry to reevaluate their positions within it. Specifically highlighting the growth of data being produced by IoT and the increasing amount of platforms to facilitate the movement of this data throughout the data economy.

Mišura et al. [4] outlines the differences IoT has from big data, and the need for a tailored marketplace. It then proposes a web interface for both data producers and consumers to interact with the marketplace.

Suliman et al. [5] then propose a solution for monetizing IoT data using blockchains. Their solution focuses on renting real-time access to IoT data and are using smart contracts on the Ethereum blockchain as a platform to achieve this. The system uses an MQTT broker [6], a lightweight messaging protocol for IoT, to aggregate the data and interact with the Ethereum blockchain. The customers receive their data through MQTT subscriptions, to verify customers the broker node verifies valid customers through interaction with the smart contract between device owner and customers. Badreddine et al. [7] continue their work using MQTT and smart contracts on Ethereum. They then propose and compare three different solutions for verification of data transfers, based on maximum traceability, minimum traceability

and bloom filter-based traceability.

Ali et al. [8] instead proposes a solution using a mix of public and private blockchains, using IPFS for file storage and transfer. This solution aims to address issues regarding scalability and privacy by isolating the IoT devices into Edge-tier private blockchains and allowing for transactions and monetary services to be handled on a public Core-tier blockchain.

Özyilmaz et al. [9] aim for their platform to function as both an always on data store and a marketplace for IoT data. They similarly to the previous option opt not to use data streams, instead using Swarm, for their blockchain they also opt to use Ethereum, like other technologies discussed here.

Gupta et al. [10] aim for a blockchain solution that allows for more specialization in the market, instead of using the blockchain mechanism to help consumers find providers they relegate this responsibility to brokers. These brokers then use smart contracts to transparently facilitate data trading. Furthermore they allow for intermediaries in their system, whom act as both a provider and consumer, attempting to add value to the data in the process.

3 Methodology

First and foremost the immutable nature of blockchains and smart-contracts must be stressed, as updating and improving these technologies in the future can prove difficult or even impossible. Because of this there is a need for rigorous and traceable research and design in the process of creating a framework or prototype. Recommendations from [11] for the structure of this study will be used. As a research framework design science has been chosen for this paper as described in [12], this means the paper will follow these guidelines:

1. The research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
2. The objective of the research is to develop technology-based solutions to important and relevant business problems.
3. The utility, quality and efficacy of the produced artifact must be rigorously demonstrated via well-executed evaluation methods.
4. The research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
5. The research relies upon the application of rigorous mehtods in both the construction and evaluation of the artifact.
6. The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
7. The research must be presented effectively both to technology-oriented as well as management-oriented audiences.

To achieve this the paper aims to review all technologies listed in the related works section, aswell as implementations in industry. Where possible these solutions will be compared in aspects such as their maximum transactions per second (tps), latency, transaction cost, energy consumption and non quantitative differences, such as resistance to different attacks etc.

In the conclusion this paper aims to provide practical recommendations for future frameworks, and will provide an artifact in the form of a natural language description of such a framework.

4 Background

This section of the paper provides an introduction into Internet of Things, blockchain and smart contracts. It discusses definitions and basic workings of these technologies and discusses some alternative technologies that are further considered outside the scope of this research.

4.1 Internet of Things

The internet of things is a network of computing devices capable of communicating with each other without the need for human to machine or human to human interaction[1]. IoT devices are used a variety of application domains, such as smart home, agriculture and supply chain. IoT devices are generally small in nature and have very limited computational power. Because of this they often make use of cloud-based infrastructure, sending the data they collected to be used in the intended use case for the device.

It is predicted that in 2025 the IoT will consist of 75.44 billion connected devices [13]. This generates an enormous amount of data, the interest in which is outlined by [3], this also regards interest in the data outside of the original use case of the device. This demand for IoT data can be satisfied using a marketplace, where data producers can sell their data to other parties, who then use it for analysis, or to drive decision making within their own IoT systems etc. However, in a world with growing concerns about privacy and misuse of data, trust becomes of uttermost importance in building such a system. Blockchain may provide a solution in decentralizing such a marketplace.

4.2 Blockchain

The idea of blockchain as we know it today first appeared in the bitcoin white paper, published under the pseudonym Satoshi Nakamoto [14]. The paper does not explicitly use the term blockchain, but describes a system for sharing transactions within a peer-to-peer network. This system has been taken as an start in developing further cryptocurrencies. Other uses are also being discovered and investigated, such as the use described in this paper. In this section I will discuss the data structure in blockchain, the networking architecture and consensus algorithms, and the programmability of blockchains using smart-contracts.

Distributed ledgers

The main aspect of a blockchain is the underlying data structure, this is where it gets its name from. A blockchain consists of a list of timestamped blocks, each block storing the hash value of it's predecessor. This structure of blocks is distributed throughout a peer-to-peer network. Because of this when a block is tampered with, all subsequent blocks must be tampered with as well, or it becomes easily detectable by

verifying the hash values. In the blocks we can store transactions, which are signed using digital signatures. This list of transactions then allows other parties to verify ownership of tokens being spent, thus enabling a decentralized payment system such as bitcoin. [14]

Blockchain types

Blockchains can be divided into three categories: public blockchains, private blockchains and consortium blockchains [15].

Public blockchains are open for participation by anyone that has the technology to interact with them, this is also know under the term permissionless. That means that anyone is allowed to maintain a copy of the blockchain and validate new blocks. These types of blockchains are well known for being used in cryptocurrencies such as bitcoin [14]. Popular consensus algorithms for these blockchains are Proof of Work (PoW) and Proof of Stake (PoS). In order to incentivize nodes in the network to validate blocks transaction fees are used, although more methods of providing incentive exist [15].

Private blockchains are blockchains of which all participants are from a single organization, this would be considered a permissioned blockchain. These blockchains function more like a database, as validators do need need incentives and access to the data is controlled by a centralized entity. They also aren't immutable by definition, as an organization can choose to roll back to any point in the past [15].

Consortium blockchains are another form of a permissioned blockchain. They instead span multiple organizations and are generally used as a synchronized database between these organizations. Whilst they are not entirely decentralized or censorship-resistant they do increase auditability and synchronization [15].

Smart contracts

Smart contracts allow for programmability to be integrated into a blockchain. They have their own accounts and addresses on the blockchain, allowing them to hold assets. Invoking the functions of a smart contract costs execution fees, this helps mitigate flooding attacks. Some of the use cases of smart contracts include:

- Allowing transactions that must be signed by multiple parties.
- Automate transactions triggered by events.
- Provide utility for other smart contracts to build further on.
- Allow application-specific data to be stored on the blockchain.

It is important to note that smart contract code is stored on the blockchain, thus visible to everyone and immutable. Because of this correctness of the code is of vital importance, as a potential attacker has access to the code and it cannot be patched. However if used well smart contracts enable many different applications to benefit from the decentralized nature of blockchains [15].

5 Analysis

In this section I'll discuss the drawbacks and benefits of the different technologies discussed in this paper. In doing this it aims to answer the sub-questions posed for this research.

5.1 Sensing as a service

In 2016 Opher et al. [3] noted the development of data as a new asset in modern business. Furthermore they mention IoT as a driving force behind this development, due to the rapidly growing amount of IoT devices [13]. Opher et al. [3] then proceed to describe the current state of the data economy as a stack, consisting of the following parties:

- Data presenters: Provide the user experience, can also play a role in the discovery of new insights.
- Insight providers: Analyze the data, provide valuable insights to generate revenue.
- Platform owners: Provide APIs, development tools and cloud services.
- Data aggregators: Normalize data and group data from different devices.
- Data producers: Allow for access to their data and responsible for collection.

It is important to note that a single company or technology does not have to stay in a single layer of this stack. This underlines the interest of this research for a decentralized marketplace for IoT data. By connecting data producers to insight providers such a marketplace can occupy the space of data aggregators and platform owners. In these spaces this can tackle issues related to trust, privacy, misuse of data and waste generated by redundant sensors. In order to show how such technologies achieve these benefits, some technologies that have been developed from 2016 will be discussed and analysed.

5.2 A centralized marketplace

The first solution to discuss is an IoT marketplace as proposed by Mišura et al. [4]. This marketplace proposes a central server, to which both data producers and consumers connect. This server runs a database of devices that are willing to produce data for a financial incentive. Consumers can submit queries to this database specifying parameters such as age, location, budget, amount of sources etc. The server then selects a best fit of devices to serve that query and proposes this option to the consumer, negotiations can be continued by modifying the query.

After negotiating what devices to purchase data from the server requests this data, then forwards it to the consumers. This allows the server to be the authority verifying whether data transfers actually occurred. If devices fail to fulfill requests from the server, this is tracked in the database and used to compute a credibility score, which can then influence query results.

In terms of the data stack this solution occupies the data aggregator and platform owner layer. It provides some simple solution for connecting producers to consumers. A more powerful querying solution would increase the aggregation

capabilities of such a system, however as the system uses a MySQL database such capabilities do not come without security concerns.

Furthermore such a concern raises issues regarding trust, privacy and misuse of data. The party running this server can choose to start providing insights into this data themselves, with potentially malicious intent. Data providers also are unaware of who is buying their data, providing transparency into this would help alleviate some concerns, but doesn't solve the issue as the relationship between provider and platform is not trustless.

5.3 Decentalization

As has been discussed in the previous section, the lack of transparency into the sales of data can be problematic. However there is also demand into flexibility implementation, as to not lose the variety of services offered in the data aggregator and platform owner parts of the stack. This paper looks at blockchain as a potential solution to solve these issues, being both transparent and programmable it seems like a natural fit. Since 2016 several blockchain solutions to this issue have been proposed and will be covered in this section.

The analysis of these frameworks is done by first identifying the different roles in said system. It is then discussed what devices can be used to fulfill these roles and how they are placed in the data economy stack. Mechanisms for trustless exchange of data and payment will be explained, and cost saving measures will be outlined.

IDMoB

The first platform to discuss is IDMoB [9], a platform based on smart contracts running on the Ethereum blockchain. This platform is only interacted with by data producers and consumers, this creates a platform that fulfills the data aggregator and platform owner roles in the data economy stack. Data producers connect their IoT devices to some form of gateway node, this is then capable of interacting with the platform and publishing data on to it.

A single smart contract provides all functionality, to provide this functionality it defines vendor, customer and payload structures. By defining a vendor, the smart contract can function as a registry, this also enables the ability to vote on vendors after a purchase, creating a reputation system. Customers are registered such that the smart contract can keep track of payments, and verify to what data the customer has access. Finally payloads are the data structure encapsulating the data transmitted, it keeps track of timestamps and other metadata, the data is then stored on Swarm [16] and can be referenced using the payload. In order to prevent unauthorized access to data on the Swarm file system encryption of files is encouraged, and a method to achieve this using symmetric keys is described in the paper. It should also be noted that the use of Swarm here has it directly competing with cloud infrastructure as an option for data storage. If costs are low enough, a device manufacturer could opt to replace their cloud infrastructure with Swarm, making it easier to sell data on such a marketplace.

To address the responsibilities of this platform as a data aggregator, it provides a basic querying ability allowing for filtering based on data type. More specific descriptions of the data are also available on a per sensor basis. This means the customer chooses specific devices to purchase data from in this system, some metrics such as timestamps, geolocation, and vendor reputation can be used to make an informed decision. This solution does not allow for an extra party in the marketplace that attempts to add value to data by applying their own data aggregation techniques and selling larger data sets.

In handling payment the marketplace uses Ether, the paper suggests the use of a custom ERC-20 [17] token. Such a token could provide a more stable form of currency in this market and help widespread adoption. Raiden Network [18] is another technology suggested by the paper, it is an off-chain solution for handling payments offering near-instant and low-fee payments. By using Raiden as a payment channel, pay-as-you-go or subscription based solutions may become viable.

Data Subscription Contracts

The next platform to discuss is that proposed by Gupta et al. [10]. Outside of the roles that we have seen in the previous platform, there are also broker nodes participating in the network, these fulfill the registration and querying responsibilities of the smart contract. In order to address trust issues regarding the process of the broker it takes part in a permissioned blockchain. Search and discovery algorithms are then smart contract based, this allows brokers to validate each others behaviour. To handle decision making within this network tokens are deployed, tokens are used for voting on allowing new brokers into the network.

The use of brokers in this framework creates issues with trust around these brokers, in order to solve this issues it is suggested they partake in a permissioned blockchain and act by the rules of a smart contract. These responsibilities can also be handled on a permissionless blockchain, this allows for cost to be handled at a function specific level, instead of over the whole process.

Data subscription contracts handle several responsibilities in this system, outside of the broker network. First of all negotiations are started by either the producer or consumer, who send out a bid, every node to receive a bid responds with a counter bid, at this point the requester decides what bids to accept. After the negotiation phase the contract keeps track of active subscriptions. The contract handles data exchanges directly from producer to consumer for the duration of the subscription. Finally it handles payment and ratings before terminating.

To handle all separate data subscription contracts, an registry contract is maintained. In this contract addresses of producers and consumers are registered, as well as the application binary interface of the contract. This allows nodes to query in which contracts they are taking part at any time after execution. However such a contract also raises concerns over scalability, as expired contracts cannot be removed.

Smart contracts on Ethereum

Ethereum is a widely adopted blockchain platform with smart contract capabilities. These scripting capabilities can be leveraged to create an IoT data marketplace, as demonstrated in [5]. Similarly to the previous architectures, a form of

gateway node is once again added in the form of an MQTT [6] device which collects data generated by IoT devices, the framework is designed to allow multiple gateway devices per owner. The framework then builds on top of the publish/subscribe architecture of MQTT, automating monetization of these data streams by leveraging smart contracts.

This framework does not propose the use of a registry contract, the smart contract functionality proposed only deals with the logic of subscribing to an MQTT topic. producers, consumers and gateway nodes all have their own Ethereum addresses in this framework. The constructor of the contract is invoked by the owner and is created for a single gateway node. This gateway contract then advertises topics it offers subscriptions to. When a customer wishes to subscribe to a data stream, they first deposit ether into the contract, after which they can purchase subscriptions and request to access active subscriptions. If a valid access request goes through, a token is granted and the event and access duration are logged. The system does not propose any mechanism to verify whether the gateway actually sends the data however, which opens the system up to malicious behaviour.

In order to solve these issues, [7] proposes a similar framework, but instead introduces third-party broker nodes to route data through which then, submit traceability information to the blockchain. As this solution still employs smart-contracts for connecting producers and consumers, data transmitted this way can be encrypted. The paper proposes three different solutions for traceability.

Trace-MAX provides maximum traceability, in this solution, for every publication the publisher writes message id, topic name, data hash, data size and timestamp to the blockchain. Similar steps are taken when delivering from broker to subscriber. This allows for exhaustive auditing of data transactions throughout the system, however it also incurs high gas fees.

Trace-MIN aims to provide some traceability for minimum cost, it achieves this by having the broker publish the total amounts of data successfully received from publishers and sent to subscribers to the smart contract. In doing this we can periodically compare these numbers. This system only discovers dishonest logging in the relationship between the broker and subscriber, and fails to catch cases of malicious behaviour where the amounts are equal.

Trace-BF aims to provide a middle ground between these two solutions, it works by creating three Bloom Filters [19] that are stored in the smart contract, one for registering outgoing traffic from publishers, one for outgoing traffic from brokers and one for acknowledgements from subscribers. With any data transaction the participant hashes the MQTT topic name concatenated with the data. Because each bloom filter is controlled by only a single party, malicious use would show non-similarity with two other similar bloom filters. A drawback of this is that checking the similarity of the bloomfilters does require halting communication. Potential misuse can occur when the publisher and broker conspire, claiming to have delivered data that hasn't been delivered or to publish bad quality data. Such an attack could be somewhat mitigated through the use of a reputation system, where sellers and brokers are reviewed and misbehaviour is publicly recorded.

IPFS as a file store

The last framework we'll review is by Ali et al. [8] and uses a tiered blockchain architecture and IPFS for file storage. The blockchain in this case consists of 2 tiers, a single public core-tier blockchain and multiple private edge-tier blockchains. Inter-blockchain gateway nodes (IBGWs) participate in this core-tier blockchain to connect it with edge tier private blockchains. The core-tier is intended to handle transactions and advertising of available data. The edge-tier blockchain serves as a community of nodes that can engage in smart contracts in order to govern data access rights.

An edge-tier blockchain keeps track of the data produced by publishing devices within it. In order to address concerns regarding DoS attacks, transactions can be limited by issuing a tokens at a maximum rate, which can slow down the more tokens a node owns. By having edge-tier nodes engage in smart contracts with IBGW nodes they can determine access control privileges of nodes on the core-tier blockchain.

On the core-tier blockchain a registration smart contract is hosted, here IBGWs and requester nodes can register themselves to engage in data trading, this contract also handles registering reviews of sellers, forming a reputation system. When a requester wishes to buy data from an IBGW they register with their smart-contract, this stores information about the data requested, the value as set by the requester and stores their blockchain address. The IBGW then checks the access-rights on the edge-tier, if the privileges and price check out, the IBGW continues to send the data by using IPFS.

5.4 Drawbacks and benefits

Looking at these different frameworks we can outline some key differences, which will be highlighted and discussed in this section.

Communicating with IoT devices

The entry point for IoT devices into the blockchain is an important factor to consider in designing such systems, as IoT devices tend to have limited computational power and memory. Most solutions reviewed here involve some form of gateway node that connects multiple IoT devices to the blockchain, although [8] would allow IoT devices to participate in private side chains with limited permissions, where an IoT device could be given the permission to execute code in a smart contract and publish data onto the network.

Data storage

After IoT devices have some way to communicate to the blockchain, the next question to answer is what data to store in this system. Before looking at data storage solutions chosen it is important to consider where data is stored in current IoT architectures. The discussed technologies either make use of a decentralized file system or require data storage capabilities at the gateway node, choosing not to transmit it into the peer-to-peer network, instead establishing a direct connection with the customer. The use of decentralized file systems requires providing incentive to nodes participating in this network however. Because of this it is suggested to only use it when transacting data, or to use it as a primary platform for data storage, [9] lists making this economically viable a key factor in the adoption of decentralized marketplaces.

Data transmission

IoT data is worth most when it is traded in real-time allowing it to be used for actionable triggers, as underlined by [10]. As such, the ability to subscribe to real-time data streams is an important functionality in a decentralized marketplace. Considering this, the model as proposed in [5] and [7] offers a valuable feature, as it handles data exchange through subscribing to an MQTT topic. However this model is shown to either have security flaws or very expensive traceability requirements to work, as all exchanges of data must be recorded for this system to be secure, with Trace-MAX costing about $5 * 10^7$ gas for 60 transactions[1]. This may cause solutions using decentralized file systems to be more attractive, as less verification is required in making these transactions happen, thus saving on gas costs. Close to real-time subscription can also be realized with somewhat higher latency by regularly querying for and buying available data. It is important to note that accessing smaller amounts of data at higher frequencies does drive up cost, as a set cost is related to data accesses and transactions. As processing fees are payed by the consumer, this creates a situation where a consumer can make an educated decision between low-latency in receiving data and higher costs vs lower costs and higher latency.

Security

A decentralized marketplace would offer financial incentives to various parties participating in such a system. This also creates an incentive for malicious behaviour, as such the security and potential misuse of such a system should be discussed. One potential misuse is running scams where data is advertised through smart contracts, but the producer never sends the data promised, or sends bad quality data. Decentralized file systems offer a solution to this by allowing a third party to verify the existence of a file in the network. The traceability submissions of different nodes in the network, and routing through a third party as suggested in [7] are an option for solutions using real-time data streams.

A general solution to help fight bad actors in such a system is the use of a reputation system. By allowing consumers to vote on producers, bad actors in the system can be more easily identified, allowing consumers to avoid them.

6 Results

In the previous section different decentralized marketplaces for IoT data were discussed and analyzed. We have seen different approaches, each with their own drawbacks and benefits. This variety goes to show a one size fits all solution does not exist for the problem. This section aims to gather these lessons about the benefits and drawbacks.

An overview of the analysis is given in Table 1, looking at this we see several frameworks offering a variety of features, but also representing a number of drawbacks. Comparing these frameworks we aim to identify features that are beneficial to this technology, and also look at ways these technologies aim to address their drawbacks.

[1]no conversion to other currencies have been made due to the volatile market

Table 1: Comparison

Framework	Data storage	Registry	Verification
IDMoB [9]	Swarm	yes	Swarm
[10]	At gateway	yes	none
[5]	At gateway	no	none
[7]	At gateway	no	Trace-submission
[8]	IPFS	yes	IPFS

Registration contracts are a feature identified in multiple frameworks that aims to address one of the drawbacks in such a system, the incentive towards malicious behaviour. By identifying bad actors in the system, consumers can make more informed decisions as to which data to buy. Furthermore it can motivate producers to get higher ratings and compete for customers.

We see different solutions for data exchange throughout the proposed frameworks. MQTT, Swarm and IPFS are discussed as potential options for transferring data between producer and consumer. MQTT does not provide mechanisms to detect malicious behaviour in such an architecture, this responsibility is solved by submitting traceability information into smart contracts as suggested by [7]. This paper also showed their Trace-BF solution scales linearly in cost with publications. Swarm is free to use up to a set bandwidth cap, from here costs scale with bandwidth [16], considering publications will be of a set size, this means Swarm also scales linearly in cost with publications. As for IPFS, the original protocol does not enable paying the network for bandwidth, however protocols such as Filecoin [20] have been built for this. It is important to note here that the cost of smart contracts and decentralized file systems is subject to change, which means no best system can be chosen from this.

An important aspect of creating this system is the gateway node that enables interaction with a blockchain. This node is given the responsibility of being able to participate in the blockchain and have sufficient capacity to engage in smart contracts where required. A private blockchain as suggested in [8] can help in governing access rights and keeping track of devices, but is not required, as proposed register contracts handle registry per gateway node and not per device.

The last important feature is separating the registry contract and data trading contracts, as is shown in [10] and [8]. By creating this separation, and giving producers their own contract to handle sales a system is created where a single registry can be kept over different types of data trading contracts, allowing more innovation and control in this area.

It is important to note the Ethereum blockchain still has volatile gas prices and the idea of a data marketplace is still developing, innovation in the mechanisms of data aggregation and sale is important in driving this market forward. Furthermore with Ethereum 2.0 [21] will help decrease gas costs by introducing Proof of Stake and sharding, which may help drive forward innovation in this field as well.

7 Responsible Research

In this section, the ethical implications and reproducibility of this research will be addressed.

This research concerns decentralized marketplaces for IoT data, this raises concerns about privacy and misuse of data. These concerns are inherent in to a marketplace for data. Some of these concerns are addressed by using blockchain. Blockchains provide pseudonymity to offer some form of privacy, however identities could still be inferred from data sold on this marketplace. It aims to address concerns for misuse of data by removing intermediaries and creating a transparent market. [22] discusses these aspects and considers blockchain technology beneficial to concerns raised in the case of an IoT data marketplace, it however does not consider these to bring the technology within ethical limits, raising concerns about increased incentive to gather private data.

There are also concerns raised about the environmental impact of blockchain. In 2014 O' DWyer and Malone [23] estimated the total energy consumption of bitcoin, the most popular cryptocurrency at the time, to be approximately equal to that of Ireland. This is largely due to the nature of the Proof of Work consensus mechanism, which represents share in the consensus with computational power. alternative consensus algorithms, such as Proof of Stake have been proposed, and Ethereum is migrating to it with their Ethereum 2.0 upgrade [21].

It is also important to consider the reproducibility of this research. As this research is a literature study no experimental data or setups can be shared. The methodology, as described in section 3, aims to provide an overview of the steps followed to ensure the research was done rigorously, by adopting a research framework and setting guidelines to follow. The process of finding papers and technologies used are described in the discussion section.

8 Discussion

In this section the steps undertaken in this research, as well as technologies used are discussed. Finally the results of this paper will be compared to earlier work.

This research started off by reading [15] for orientation of the subject of blockchain and IoT. From here the topic of IoT Data monetization was chosen, after which literature was gathered google scholar and using various combinations of the keywords: "IoT", "data", "blockchain", "monetization", "marketplace" and "decentralization". The frameworks from the resulting papers were then compared on several criteria, after which an analysis of different features was done. Finally as a conclusion a set of guidelines was introduced.

These final guidelines aim to summarise the lessons learned from this research. It is important to note that these guidelines are simply a product of the analysis done in this paper, and designing and verifying a system using these guidelines is left as a suggestion for future work.

9 Conclusion and Future Work

At the start of this paper a set of sub-questions was set to help answer the main research question. This section aims to answer those sub-questions, then provides an answer to the main research question based on that.

What challenges are there when implementing blockchain on IoT devices?

IoT devices are typically small, embedded devices with limited computing capabilities. Blockchain can require large amounts of memory and computational power, which IoT devices do not have. In this research this is solved by not having IoT devices participate in the blockchain, but delegating this responsibility to a gateway node.

How to sell IoT data, and how can blockchain play a role in this?

To sell IoT data a platform is needed to handle data transmission and payment between data producer and consumer. Doing this in a centralized fashion raises concerns about privacy and data misuse. Blockchain provides a technological framework for a more transparent marketplace, allowing users to exercise more control over their data.

What methods exist to monetize IoT data using blockchain?

This paper reviews five methods for creating an IoT data marketplace using blockchain.

How can these methods be improved?

After taking the lessons learned this paper aims to provide a set of guidelines in designing decentralized IoT data marketplaces:

1. The marketplace must operate a register contract to disincentivize malicious behaviour from data producers.
2. A mechanism to verify data transfer must be in place.
3. Requirements for gateway nodes must be kept to a minimum.
4. Registry contracts should allow for flexibility in data trading contracts.

What are the drawbacks and benefits of currently existing blockchain technologies for IoT data marketplaces?

Current blockchain technologies for IoT data marketplaces offer several benefits over centralized solutions. It provides a trustless manner to trade data between producer and consumer, and enable users to gain insights into where their data is going. However the use of smart contracts still faces a significant cost barrier, improvements made with upgrades such as Ethereum 2.0 may bring improvements into this. Ethical concerns about enabling the trade of data in this fashion are also being raised.

Suggestions for future works

This research provides an overview of literature and comparison on aspects of discussed technologies. Based on this it proposes a set of guidelines for future frameworks, future research could focus on defining the technical aspects of a framework within such guidelines and implementing a proof of concept. Furthermore the comparison could be extended to include more quantitative measures for comparison, such as transaction and smart contract execution costs.

References

[1] P. Goyal, A. K. Sahoo, T. K. Sharma, and P. K. Singh, "Internet of things: Applications, security and privacy: A survey," *Materials Today: Proceedings*, vol. 34, pp. 752–759, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221478532033385X

[2] A. Richterich, *The Big Data Agenda : Data Ethics and Critical Data Studies*. University of Westminster Press, 2018.

[3] A. Opher, A. Chou, A. Onda, and K. Sounderrajan, "The rise of the data economy: driving value through internet of things data monetization," *IBM Corporation: Somers, NY, USA*, 2016.

[4] K. Mišura and M. Žagar, "Data marketplace for internet of things," in *2016 International Conference on Smart Systems and Technologies (SST)*, Conference Proceedings, pp. 255–260.

[5] A. Suliman, Z. Husain, M. Abououf, M. Alblooshi, and K. Salah, "Monetization of iot data using smart contracts," *IET Networks*, vol. 8, no. 1, pp. 32–37, 2019. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-net.2018.5026https://ietresearch.onlinelibrary.wiley.com/doi/pdfdirect/10.1049/iet-net.2018.5026?download=true

[6] "Mqtt," 2020. [Online]. Available: https://mqtt.org/

[7] W. Badreddine, K. Zhang, and C. Talhi, "Monetization using blockchains for iot data marketplace," in *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, Conference Proceedings, pp. 1–9. [Online]. Available: https://ieeexplore.ieee.org/document/9169424/

[8] M. S. Ali, M. Vecchio, and F. Antonelli, "A blockchain-based framework for iot data monetization services," *The Computer Journal*, vol. 64, no. 2, pp. 195–210, 2021. [Online]. Available: https://doi.org/10.1093/comjnl/bxaa119

[9] K. R. Özyilmaz, M. Doğan, and A. Yurdakul, "Idmob: Iot data marketplace on blockchain," in *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, Conference Proceedings, pp. 11–19.

[10] P. Gupta, S. Kanhere, and R. Jurdak, "A decentralized iot data marketplace," *arXiv pre-print server*, 2019. [Online]. Available: https://arxiv.org/abs/1906.01799

[11] H. Treiblmaier, *Toward More Rigorous Blockchain Research: Recommendations for Writing Blockchain Case Studies*, 2018.

[12] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004. [Online]. Available: http://www.jstor.org/stable/25148625

[13] T. Alam, "A reliable communication framework and its use in internet of things (iot)," vol. 3, 2018.

[14] "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[15] M. S. Ali, M. Vecchio, M. Pincheira, K. Dolui, F. Antonelli, and M. H. Rehmani, "Applications of blockchains in the internet of things: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1676–1717, 2019. [Online]. Available: https://ieeexplore.ieee.org/ielx7/9739/8727625/08580364.pdf?tp=&arnumber=8580364&isnumber=8727625&ref=

[16] "Swarm," 2021. [Online]. Available: https://www.ethswarm.org/swarm-whitepaper.pdf

[17] E. Yilmaz, R. Cordell, S. Richards, and A. Gontijo, "Erc-20 token standard," Report, 2021. [Online]. Available: https://ethereum.org/en/developers/docs/standards/tokens/erc-20/

[18] "Raiden network." [Online]. Available: https://raiden.network/101.html

[19] J. Blustein and A. El-Maazawi, "Bloom filters. a tutorial, analysis, and survey," *Halifax, NS: Dalhousie University*, pp. 1–31, 2002.

[20] P. Labs, "Filecoin," 2017. [Online]. Available: https://filecoin.io/filecoin.pdf

[21] "Ethereum 2.0," 2021. [Online]. Available: https://ethereum.org/en/eth2/vision/

[22] G. Ishmaev, "The ethical limits of blockchain-enabled markets for private iot data," *Philosophy Technology*, vol. 33, no. 3, pp. 411–432, 2019.

[23] K. J. O. Dwyer and D. Malone, "Bitcoin mining and its energy footprint," in *25th IET Irish Signals Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014)*, Conference Proceedings, pp. 280–285.