A Privacy-Preserving Blockchain Platform for a Data Marketplace

PAULO VALENTE KLAINE[*] and HAO XU[*], James Watt School of Engineering, University of Glasgow, UK
LEI ZHANG, James Watt School of Engineering, University of Glasgow, UK
MUHAMMAD IMRAN, James Watt School of Engineering, University of Glasgow, UK
ZIMING ZHU, Bristol Research and Innovation Laboratory, Toshiba Europe Limited, UK

Recent data leak scandals, together with the under-utilization of collected data (estimated that around 90% of data never leaves a device's local storage), limits the applicability and potential of novel data driven applications. Thus, novel ways to treat data, in which users are guaranteed control, usability, and privacy over their generated data, are needed. In this paper we propose a novel privacy-preserving blockchain framework for a data sharing marketplace. The proposed framework allows users (such as sensors and devices) that generate data to store it in external servers, while the blockchain is utilized to record buy and sell transactions between parties, as well as perform access control by generating access sequences whenever trades are performed. A novel perspective over data ownership is presented, in which whoever generates the data has completed ownership and control over it and the blockchain transactions are only utilized to guarantee temporary access to it. The proposed blockchain framework also supports different types of data and provides, via the distributed and openness of the framework, quality, timeliness and similarity control over the data stored in the marketplace. In this context, different types of applications that can benefit from this framework are presented and open problems are discussed.

Additional Key Words and Phrases: Blockchain, Big Data, Internet of Things, Marketplace.

1 INTRODUCTION

The exponential growth in mobile traffic has led the wireless network industry to produce and collect an unprecedented amount of data [4]. It is expected that by 2025 the number of devices connected to the Internet will be around 42 billion, generating around 80 zettabytes of data in the same year [10, 23]. Due to this huge amount of data generated daily, several novel data driven applications have emerged in recent years, ranging from vehicular communications, smart healthcare, to personalized services. However, it is estimated that around 90% of the data generated by sensors never leave a device's local storage and is not utilized [11], forming numerous isolated data-islands and contradicting the behavior that data is the *new oil* of the digital economy. Several reasons contribute for this under-utilization, such as difficulties in terms of structuring the data, collection of redundant information, and political or economical reasons. On top of these issues, there are also concerns in terms of user's privacy and how major corporations utilize their data, which also has a direct impact in data collection, since less user's opt-in to these services [12, 34, 35].

In order to remove the public's concern and regain their trust in terms of data collection, as well as controlling who has access to their data (by allowing users to determine what type and with whom their information is shared), new approaches on how data is treated are needed. However, building a new data platform, such as a marketplace, in a centralized manner, is not ideal. While this model has worked until now, the exponential growth of Internet of Things (IoT) devices as well as the data collected, threatens to push centralized systems to its limits in terms of scalability [5, 23].

[*] Both authors contributed equally to this research.

Authors' addresses: Paulo Valente Klaine; Hao Xu, James Watt School of Engineering, University of Glasgow, Glasgow, UK, Paulo.ValenteKlaine@glasgow.ac.uk, Hao.Xu@glasgow.ac.uk; Lei Zhang, James Watt School of Engineering, University of Glasgow, Glasgow, UK, Lei.Zhang@glasgow.ac.uk; Muhammad Imran, James Watt School of Engineering, University of Glasgow, Glasgow, UK, Muhammad.Imran@glasgow.ac.uk; Ziming Zhu, Bristol Research and Innovation Laboratory, Toshiba Europe Limited, Bristol, UK, ziming.zhu@toshiba-bril.com.

Moreover, centralization can result in single points of failure, access inefficiencies and targeted attacks, which can lead to data leakages such as the ones from the Facebook-Cambridge Analytica and Google+, both in 2018. Moreover, having a single party control a tremendous amount of data is not economically healthy, as users would be limited to adhering to the guidelines imposed by a single organization to have their data in such platform [26]. Moreover, this behavior also threatens vendor lock-in, limits interoperability between parties and exposes the data to the vendor, reducing consumer's trust [3]. Thus, there would be no proper incentive for data sharing, hindering the advance of innovations and economic growth [12, 23].

To overcome these issues, a decentralized architecture can be considered, such as in blockchain [19, 28]. Originally proposed as the backbone of Bitcoin, blockchain has become a revolutionary decentralized data management framework that can transform the way in which information is shared. Since blockchain does not rely on a central server, it can provide more transparent access to information, while also guaranteeing privacy and being more secure against single point attacks. Blockchain is also able to enhance data integrity and security, contributing to restoring the general public's trust. In addition to its popularity in finance, blockchain has been applied in other sectors including energy, supply chain, healthcare, and wireless networks [14, 27, 28, 31]. In the context of a data marketplace, a few companies have already attempted to deploy such an ecosystem, such as [9, 24]. In [9] an IoT data marketplace where the blockchain is used to record transactions while data is stored in external servers is proposed. Whereas in [24], a private blockchain solution is developed in order to exchange data from supply chains, such as tracking shipments and their arrival locations. Despite the similarities with our work, the main differences between our proposed scheme and these solutions come from a technical aspect regarding the blockchain. In all of these implementations the blockchain assumes a simple role of just recording transactions or basic access control. Whereas in our framework the blockchain has a bigger role in terms of controlling the quality of data that is available in the platform, tracking reputation and assigning rewards to different parties, detecting similarities between data, as well as performing access control and enabling users to opt-out of the platform. Access control based on smart contract and blockchain-enabled key exchange are playing important roles in the proposed architecture, as they have been realised in a comprehensive privacy-preserving manner, where the users' data security and integrity are guaranteed and automated by the marketplace. Furthermore, our proposed framework is also composed of neutral third parties storage servers, which are responsible for storing data and preserving privacy. These storage servers can be verified and accessed by interacting with the blockchain smart contracts, through buy or sell transactions, and allow the download of the data uploaded to the marketplace. By using external storage in this marketplace, memory and computationally constrained devices, such as IoT devices can participate in this framework.

Other works that explore the idea of storing data in external servers when using a blockchain system are the works in [34, 35]. In [35] a peer-to-peer system that enables different parties to store and run computational operations on data stored in external servers while keeping the data private is proposed, whereas, in [34], the authors extend their ideas and propose a blockchain system that allows users to control the data the generated and that can autonomously perform access control to data stored in external servers. Similar to those works, in our marketplace external servers are used to store data and the blockchain is utilized to perform access control as well. However, in our proposed marketplace we do not store a pointer to the data in external servers in the blockchain, but rather we record the transactions performed by generating a sequence that hashes the encrypted data with a public-key of the buyer to verify a transaction. This way we guarantee that only the person that has requested to buy that data can access it. Another approach in [17], considers a privacy preserving data sharing solution for industrial IoT devices in a federated learning scenario, whereas in [21] a blockchain framework for data access control of IoT data is proposed. In [22] a blockchain-based data marketplace for IoT devices is proposed, where devices interact with machine learning providers and exchange data in a transparent manner. Also, in [12] a blockchain marketplace is proposed, in which user's data is uploaded to the blockchain together with a policy (dictating how other users can access the data). In [3] a data marketplace based on blockchain is also proposed. In their design buyers and sellers agree on a set price beforehand. After that, the seller performs a data transformation through a function and sends it to the buyer for verification. After receiving the verification confirmation, the seller then uploads the inverse of the transformed data to the blockchain, which is then received and decoded by the buyer. Another work that proposes a blockchain-based marketplace is in [26], which considers a platform based on the Bitcoin protocol to validate and record transactions between parties.

In [20] the authors evaluate different trading protocols in data trading IoT environments in a narrow-band IoT system, where massive sensing is supported. The authors provide a more fundamental work and analyse the performance of the blockchain system in terms of latency and energy consumption. Also, in] [2] a blockchain crowd sensed data trading system is proposed to overcome the challenges and inefficiencies of conventional crowd sensed data trading, which requires for all transactions to go through a broker. Their work shows that by incorporating blockchain into this system, trustworthiness can be guaranteed, preventing auction manipulation, while also motivating the upload of truthful and meaningful data, similar to our proposed approach. In addition to these works, Non-Fungible Tokens (NFT) have recently emerged as a revolution of digital art copyright [29]. However, NFTs currently suffer the heavy cost from storage of original art, whereas in our proposed data marketplace, it solves the authentication and confirmation of copyright while offering peace-of-mind storage solution to NFT creators and consumers. Lastly, several companies around the world are already integrating blockchain with IoT and artificial intelligence (AI) to share and trade data, which highlights the importance of this topic [23].

Differently than previous works, this paper proposes a novel blockchain architecture to enable a data marketplace. By utilizing blockchain as its enabler, the main goal of this proposed architecture is to guarantee that untrusted parties are able to make data transactions to one another, while keeping their privacy and data ownership. The proposed framework supports all types of data, and builds upon the ideas of [34, 35] to store the traded data into off-chain storage clusters, while the blockchain is utilized to register the transactions between parties. Access control is performed by generating a one-time sequence whenever a trade is made, which is used by the buying party to read the data stored in the off-chain storage.

By utilizing this sequence, the proposed framework guarantees that the data still belongs to the user who generated it, while keeping access to the data restricted from other non-buying parties. Moreover, whenever a user opts-out of the marketplace all sequences are revoked, guaranteeing that no unwanted transactions take place after the user withdraws. The proposed framework also supports a two-way system, in which users can not only share data, but also the byproduct generated by the consumers such as machine learning models can also be sold to enable specific applications (e.g., buyers and sellers' roles are interchangeable). Furthermore, mechanisms to control the quality, timeliness and similarity of data in the marketplace are also provided, by assigning incentives, penalties and monitoring each device's reputation level via the blockchain. Lastly, we present potential applications, open issues and future work directions. The contributions of this paper are:

- Proposing a blockchain framework for the trade of any kind of data in a private and unregulated manner, ensuring neither users' identity, access information and data integrity is compromised.
- Investigating how access control can be performed, guaranteeing that the data still belongs to the user that generated it and allowing users to opt-out of the scheme. A fully decentralized key exchange scheme is also included to improve the end-to-end data privacy and integrity in a fully decentralized manner.

- Providing a mechanism for data quality control and similarity detection in the form of incentives and reputation.
- Investigating the balance between client, data providers and miners and the economic aspects of such balance.
- Consideration of different levels of data, and presentation of different use-cases that the marketplace can use.
- Discussion and suggestions as to novel applications, open problems, and solutions.

2 PRELIMINARIES

2.1 Big Data

Due to the myriad of new data driven use-cases that have emerged in recent years, the production and collection of data has seen an unprecedented increase. As such, data is being considered as one of the most important assets in today's world, since it can generate powerful insights and predictions about someone's behavior, or interests [11]. However, due to the vast number of applications, the data generated can be quite different in terms of data sources, volume, and purposes. Since the proposed marketplace is capable of handling any type of data, it is only natural that a brief review and classification of big data types is performed.

2.1.1 Data Sources. With the increase in number of devices equipped with Internet connection capabilities and sensors, it has become quite easy to collect data. In this regard, data can come from several different sources, such as [4]:

- Data centers and servers, in the case of web applications.
- User generated content, either through streaming, embedded sensors, or mobile applications.
- IoT devices and wireless sensor networks, in the cases of healthcare, agriculture and industry.

2.1.2 Data Transmission. There are two main modes in which data can be transmitted, either wired or wireless, and, depending on the application requirements, one transmission mode is favored over the other. Wired transmissions are usually extremely reliable, whereas wireless transmissions suffer from attenuation from the wireless channel. This can potentially bring novel challenges to blockchain architectures in terms of the blockchain performance and consensus algorithms, since blockchains have been designed to work in reliable domains.

2.1.3 Data Volume. Depending on the type of application and data source, as well as the rate of data transmission and battery life, the volume of data produced can be quite different. Applications that require raw data, for example, are going to consume significantly more resources than applications that only require processed data to be transmitted to them. Volume can also differ depending if the data is utilized at the application or network side.

2.1.4 Data Types. Data can be divided into two main categories: raw data, or processed data. Raw data consists of unprocessed data collected from a source that is directly fed into its application, whereas processed data undergoes manipulations after it is collected in order to produce meaningful information, such as relevant features. Examples of raw data consist of images from a camera, which can be used in the context of deep learning to automatically discover representations and patterns in the dataset [15]. On the other hand, conventional machine learning algorithms, such as supervised and unsupervised learning, need to be fed with processed data that have hand-crafted features, to discover correlations and patterns in the data and make their predictions [13, 15]. Another example of processed data is in federated learning, in which each device learns based on its own local data [16]. However, depending on the type of data, different challenges can arise in the proposed marketplace, in terms of storage requirements as well as similarity detection.

2.2 Blockchain and Smart Contracts

2.2.1 Blockchain. Blockchain has been proven to be a powerful tool in multiple industries including finance, supply chains, energy and data trading, and in IoT, to name a few. It is widely regarded as the next technology to enable a new era of collaboration and innovation, due to its advantages in terms of replacing a central server with a distributed ledger, therefore preventing data from being tampered with, and allowing anyone with access to the blockchain to check any information inside it [19].

Blockchain consists of a distributed database responsible for keeping its data in open and tamper-proof distributed storage. This allows blockchains to not depend on a central server controlled by third parties, increasing trust and transparency [30]. These blocks are composed of a header, which establishes a retroactive connection from the most recent block to the first block created [19], thereby allowing users to verify the integrity and authenticity of any known block; and a body, which contains the data in the blockchain and can be composed of any payload. In order for participants to locate the latest block of the chain, every user of the blockchain must hold a copy of the headers of the longest chain, making the network distributed. In addition to this data structure, new blocks are added by a collective effort, through a consensus mechanism, in which participants can agree on the status of the ledger [19]. Blockchain also relies on the use of encryption, which enables different levels of authorities and interactions in the network, thus, due to these advantages, blockchains are a good alternative to enable a private and secure data marketplace.

It is envisioned that the proposed marketplace can work similarly to Bitcoin, in which blockchain acts as the backbone infrastructure in order to preserve privacy between users, and to allow untrusted parties to make transactions with one another [19]. However, instead of trading currency, access to data is exchanged. This is done through the utilization of external storage servers responsible for storing the data and also verifying when a request has been made. Alternatively, the seller may choose to expose itself for a limited time period and offer the data downloading service with Trusted Execution Environment (TEE), if supported by the host, though it is not the first choice. In the setup of using external storage, the data exchanged are well encrypted and sliced, thus illicit access to the data does not tamper the data, nor causes any breaches. Thus, the blockchain is responsible for recording the transactions between two parties, which represents who can have access to the data stored in the external servers. In addition, blockchain can also record whenever someone revokes access to its data, generating a complete timeline of when the data access was granted and revoked, thereby creating a transparent and controlled environment for its users. By creating a decentralized blockchain platform to buy and sell data, a highly scalable, secure, private, and trusted peer-to-peer data sharing framework can be achieved. This could potentially eliminate bottlenecks found in centralized systems, while also keeping user's ownership of the data, giving them increased power over their own data, as opposed to today's standards.

2.2.2 Smart Contracts. Smart contracts consist of self-executing scripts in the blockchain, in what can be described as protocols that executes the terms of a contract [7]. These contracts are codes that can be executed by either miners or clients, and have the objective to provide regulation without the need of trusted intermediaries between transacting parties. Since these contracts are part of the chain, whenever users want to use them, instead of addressing the transaction directly to another user, the transaction is addressed to the smart contract, which executes independently and automatically. This allows for parties that do not trust each other to exchange data in a common and trusted platform (the blockchain) [7].

In the context of the proposed data marketplace, smart contracts can be used to perform transactions between untrusted parties. Requests to buy / sell data can be sent to the smart contract, which is then responsible for making sure all requirements of the transaction are met, as well as for publicizing the transaction information in the blockchain.

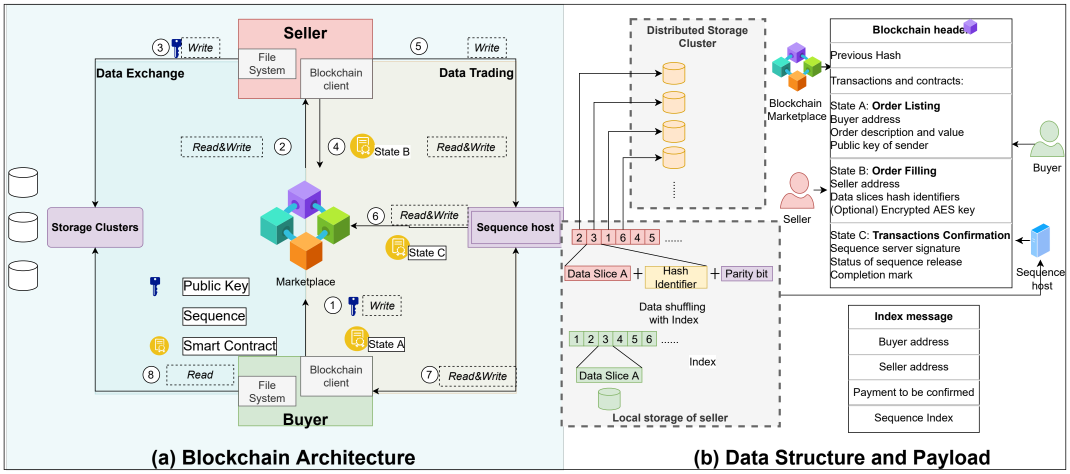

Fig. 1. Proposed Blockchain-enabled marketplace.

The transaction can have information about to whom the seller is granting access, but it can also involve a time-frame, in which the other party can have access to the data for a period of time. The utilization of smart contracts can lead to the automation of the marketplace, facilitating the exchange of goods, such as through deposit holdings whenever a transaction is pending.

3 BLOCKCHAIN-ENABLED DATA MARKETPLACE FRAMEWORK

The blockchain-enabled data marketplace is a holistic approach to build up the trust and privacy between anonymous sellers and buyers for different data applications, such as machine learning, data analytics and many others.

In this section, we introduce the proposed data marketplace framework with detailed definitions of its elements, architecture and the data structure. Moreover, we also include details in terms of the data structure and payload, as well as the services provided by the smart contracts, sequence hosts, and storage clusters. The orchestration of the state-of-the-art components combined with the multi-stage smart contract flow design are tailored for the desired privacy-preservation when two anonymous entities interacting on the actual data exchange.

3.1 Entities, functions and interfaces

The following parties are defined as the entities with their roles and interfaces:

- Blockchain-enabled data marketplace[1]: a distributed ledger that serves Buyers and Sellers by providing them with a smart contract and tamper-proof ledger. The marketplace offers the following basic functions to the users: listing data request, notification of filled data request order, listing the encrypted sequence (index) from sequence host[2], and access/quality control.
- Buyers: those who request data via the blockchain-enabled marketplace with a request order. During the request order creation, Buyers also attach their own public-key and the price for the order, if applicable.

[1] Such a system is also eligible for a blockchain-enabled data sharing service, where the service is hosted without trading functionality, the following stated Sellers and Buyers are basically data provider and enquirer

[2] Sequence is an example of transposition cipher methodology [18] used in data shuffling.

- Sellers: those who accept the request order listed on the blockchain marketplace and provide the intended data, which are encrypted by the Buyer's public-key, to the Buyer. The roles of Seller and Buyer are exchangeable. It is worth mentioning that in most scenarios, the Seller should fill the order from the Buyer, hence the reputation (which is known from the blockchain) and the link to real identity may be desirable for trading purposes. Moreover, it is also important to highlight that this framework is fully autonomous, where as Sellers generate data, it gets automatically advertised to the blockchain via configurable scripts n each node. Similarly, Buyers can also have certain scripts or functions that scan for a particular type of data and automatically place buy requests.
- Storage Services Providers: the hosted storage solution for the Seller's data alongside with the blockchain network. Storage Services Providers are neutral parties who co-exist with blockchain miners. In the scope of the proposed solution, they are voluntary services provided by miners, but the incentive of running hosted storage for the marketplace is also open to exploration. The privacy is preserved when the access information is not tagged with specific interest and individuals. The access to the storage clusters is considered as privacy-preserving because the user downloads a trunk of data, which includes the wanted data slices and some random slices that conceal the interest of intended data. In this way, the storage clusters provide unconditional data downloading service to any user, hence with all the knowledge obtained by the storage provider, it cannot link the user with specific data slices, even by knowing the specific data being downloaded. By incorporating an external storage in the marketplace, it allows devices that have limited memory and computational power, such as IoT devices, to participate. Moreover, the utilization of external storage also contributes to the key principle of this work, in which whoever generates data has full control over it. If data transfers were done directly, without external servers, this control would not be possible, since a direct transaction would transfer to the buying party full control over the data generated by the seller. By using an external server, we make sure that only the owner of that data has ownership over it and only a "right to use" the data is sold to the buyer, whereas the original data still belongs to the original owner.
- Sequence Host Service[3]: a third-party service provider that broadcast encrypted sequences using the Buyer's certificate. The Sequence Host Service is also a break-in point for regulators and policy makers without compromising users' privacy. Though it is described as an entity with centralized behaviour, in the practice, the sequence host can be decentralized by employing decentralized oracle or independent validator nodes in the blockchain network, randomly assigned in each transaction. The sequence host can operate in a decentralized manner by allowing users to act as sequence hosts with their commitment made to the community, by depositing or stacking activities in the community. The sequence host is able to create rules on data deletion period and access control, which verifies the Buyer's blockchain wallet address. As the blockchain marketplace is a self-regulatory community, the sequence server should be considered as an independent service provider who gains the trust from users. The sequence host operates on a minimal trust model, same as other entities, but it accumulates reputation by providing users with good quality of service. Blockchain-enabled mutual authentication protocols, such as [32], are encouraged to ensure the sequence host is authentic. An overview of the privacy visibility of all entities is presented in Table 1
- Data as the trading goods: the goods traded in this marketplace, as requested by Buyers and supplied by the Sellers. They are first produced by the Seller as raw or processed data, then are encrypted by the Seller using the public-key supplied by the Buyer and then shuffled with a sequence, resulting in the out-of-order data (which

[3] The Sequence Host Service is an example of host service for transposition cipher methodology, any alternative cipher tools can be introduced if it suits the architectural design of the proposed scheme

has an additional parity bit), before storing it on the external storage service in the form of fixed length slices (sliced data), a minimum number of slices of 256 is recommended for brute force attack prevention. Although we have designed some protection mechanisms over the data, such as similarity detection and access control, other issues such as the upload of sensitive data or data bought in the marketplace being offered outside the platform are out of scope of the paper.

- Cryptography measures: Public-key-based cryptography used to encrypt the designated information. Note that all trading parties shall have use of Trusted Platform Module (TPM) and TEE.
  - Public-keys are commonly used to encrypt data from Sellers and the decryption is done at the Buyer's end. It allows the exposing of encrypted data to third party storage services and prevents sequence server corruption.
  - Sequences[4] are random indexes used for the sliced data order randomization, it is also used to assure the integrity of the data stored on the storage clusters. In addition to the security of a slice order, a recommended number of slice of 256 is required to suit the needs of one-time pad and the distributed access performance.
  - (Optional) AES or other symmetrical encryption for enhanced security. The exchange of the symmetrical key is initiated by the Seller, who uses the Buyer's public-key to protect the symmetrical key.

The above described forms a tamper-proof and stage by stage content delivery system with strong credential and encryption control over it.

Table 1. Privacy visibility to all entities

	Seller IP	Seller ADD	Raw Data	Buyer IP	Buyer ADD	Encrypted Data	Data Key	Sequence/Index	Encrypted and Shuffled Data
Sellers	N/A	N/A	Y	N	Y	Y	Y(Public)	Y	Y
Buyers	N	Y	Y	N/A	N/A	Y	Y(Private)	Y	Y
Storage Clusters	Y (not linked to identity)	N	N	Y (not linked to identity)	N	N	N	Y	Y
Sequence hosts	Y	Y	N	Y	Y	Y	N	Y	Y
Public	N	Y	N	N	Y	N	N	N	Y

3.2 Architecture

The blockchain-enabled data marketplace is a holistic approach to share raw, processed data, or even the resulting models of its users with enhanced privacy and security features. There are three primary trust zones: Zero Trust Zone (ZTZ), Minimal Trust Zone (MTZ) and Trusted Zone (TZ). All entities are assumed in the ZTZ at the start of trading. Once the transaction happens between entities (seller, buyer, sequence host/market regulator, and third party storage, they are classified into MTZ with their presence of blockchain address. The MTZ sees through the State A, State B and State C in the smart contracts. Trusted Zone does not exist in the physical procedures of trading, it applies to entities who have completed mutual-authentications between each other, exclusively for Seller and Buyer, as they will have chance to exchange the key between each other.

It consists of 8 basic steps. Note that an arrow does not indicate read/write operations to the blockchain or file system, it is an indicator of access initiator, or in other words, it indicates who is the active party regarding the service performed and represents the process flow. The steps of the proposed framework are described as follows:

- Prior to a buyer making a buying request, it is assumed that nodes that are in this system advertise the available data through the blockchain. This is achieved by the nodes that generate data sending transactions to a smart

[4] Note that the shuffling method is a general idea of transposition cipher techniques [18], without deforming the data. A new sequence is generated for every data transaction with an equal size of the number of slices, hence it is unbreakable in theory with a safe minimum length of sequence, which is effectively a one-time pad [25].

contract in the blockchain containing information about the data, such as what type of data was collected, data tags, timestamps, minimum price, who is the owner, and other relevant information. More importantly, when advertising the data, the Seller can also upload data access information, if that data can be shared to anyone (public), or only to specific addresses (private).

- Whenever the Buyer wants to purchase data from the marketplace, in *step* 1, the Buyer needs to interact with a smart contract, which records the Buyer's public-key and awaits the Seller to complete the transaction, the contract is now in *State A*, also known as Order Listing state
- Data Sellers fill the contract of the requested data with optional symmetrical keys with the Seller's public-key encryption via the blockchain, in *step* 2.
- The relevant encrypted data with pseudo-random stream encryption (out-of-order data) is uploaded to hosted storage clusters, as in *step* 3.
- The data is confirmed to have been shared via smart contract in *step* 4 and achieves *State B* of the contract, which is known as Order Filling state.
- The Seller uploads the sequence order to a third-party server, which only releases the sequence order when the identities of both sides are matched as in Fig. 1 (a), *step* 5.
- The sequence server is required for access control purposes. It looks up the blockchain records and sends the relevant sequence of the out-of-order data upon requests, as indicated in Fig. 1 (a), *step* 6, where it also sends the contract completion notice to the blockchain contract (*State C*), as Transactions Confirmation state.
- After the hosted sequence is successfully released in *step* 7 of Fig. 1 (a), an optional payment verification can also take place, such as the sequence server requesting deletion of shared data by canceling the sequence, as requested by the agent of policy makers and regulators. Moreover, after the smart contract is fulfilled, the funds stored in it, due to the purchase of the data, are then transferred to the Seller, completing the transaction.
- The hosted storage service provides the general public with the downloading service, in *step* 8. Ideally the sequence host is co-hosted with the distributed storage server, in order to avoid access information leakage.

3.3 Data Structure of Blockchain-enabled Data Marketplace

The data structure of data shared by the blockchain-enabled marketplace and the blockchain data structure are, by design, pillars of privacy-preservation. In Fig. 1 (b), we have a breakdown of both structures with cryptographical measures explained. Note that the blockchain data structure is exclusive to the data stored in the block in the form of transaction records and it is not intended for sharing the actual data, but rather the contract information.

Data as the trade goods. The data, defined within the article, is the only good traded in the designated marketplace, and it is designed to fit into multiple lengths and sizes of information with robust privacy and security requirements, both raw and processed. An identifier is generated by hashing the data, in order to provide a unique ID for the unique data. Data types for the data marketplace are divided into two categories: 1) the sequence hosted by the sequence server; and 2) the sliced (in green) and shuffled (in red, encrypted) data stored in storage clusters with hash identifiers and parity bits.

Index of the sequence. The index of the sequence is used to further encrypt the sliced data. The index converts the ordered data slices into a random order data series, which provides an additional layer of data protection. The message from the Seller to the sequence server, as shown in Fig. 1 (b), has both addresses of the Buyer and Seller, thus allowing payment to be confirmed and the sequence used for decryption. The addresses in this message are used for sequence look up and Buyer verification. Thus, in the event of sequence server collusion, there is no concern of data or privacy breach.

Contracts. The contract used in the marketplace is set to be a fixed term contract, where all the parties are required to fill in all the fields with valid information. The contract consists of three states, dedicated by three parties during the transaction, as indicated in Fig. 1 (b). In State A, the Buyer advertises an open order request to the network with its public-key; in State B the Seller accepts the order and delivers the data as requested; and in State C the sequence host confirms the acceptance of the delivery and audits transactions and participants across the network.

3.4 Secured transactions and deposit holding

Due to the nature of virtual goods trading, transactions are not guaranteed in both sides, in other words, the Buyer cannot verify the goods before sending the payment, and the Seller has no way to secure the payment once the item has been dispatched. In our proposed marketplace, this can be overcome through the sequence host by setting the State C of the contract into an obligatory status.

Using the smart contract, the marketplace enables the Buyer to deposit rather than sending the payment directly to the Seller, and the Buyer knows beforehand that the payment has been setup for final confirmation. The sequence server is set to be the verifier of the transaction by signing the contract. The verifier is required to audit the payment from the Buyer and release the fund (by signing State C with a completion mark) to the Seller when the sequence has been retrieved by the Buyer, which is also the sign that the Buyer has accepted the deal. Note that, the Buyer has the right to download all the data from the storage clusters, hence the Buyer knows that the data has been delivered [5].

3.5 Incentives and Penalties

In addition to recording transactions and performing access control, the blockchain can also track reputation of different parties. For instance, depending on the number of requests performed and requests fulfilled, buyers and sellers can have a publicly visible reputation index (similar to vendor ratings offered in other marketplaces such as Amazon or eBay). This can then be used to increase the transparency and trustworthiness of the marketplace as well as by users, which can filter buyers/sellers based on reputation levels. In addition to reputation levels, other incentive systems can also take place. For example, depending on the quality and timeliness of the data collected (which can be monitored by the amount of requests generated to buy that specific data) certain data producers can be rewarded by for example, earning a percentage of the mining fee. This can encourage users to generate meaningful data in order to receive better rewards. Similarly, if low quality data (such as fake or similar data) is uploaded to the marketplace, penalties can be enforced, by having for example a percentage of the mining fee added to the transaction of the user that generated that data, effectively discouraging the upload of fake data to the marketplace. Moreover, due to the nature of blockchain consensus mechanisms, it has already been proven that through different consensus protocols and incentives/penalties, fake and similar data can be prevented in decentralized applications [6].

3.6 Similarity detection

Similarity detection for data resale is allowed in the marketplace. However, the feature is considered as an additional service in terms of generalized trading behaviors, where the similarity is not concerned. Data duplicates shall be rejected

[5] It is out of the scope to determine whether the data is fake or counterfeit, but the marketplace can punish the Seller using the consensus formed by the majority of users. Once the consensus has been made to block a certain user, the fund held by the user is subject to rejection by all participants.

by the marketplace by either monitoring an entity's reputation level [6] or comparing the hash of the data, as a piece of raw data is pointed to a unique hash key, hence preventing a second entity of selling the same data. On the other hand, the same piece of encrypted data can be re-shuffled (in case that a malicious user obtained all encrypted sliced data from a legal seller without finishing the contract, and attempts to be the owner of encrypted data), the parity bits for the particular slice of data are the same. The possibility of performing encrypted similarity detection is imminent with the progress of [8][33] and similar works, and the use of homomorphic encryption is considered for future work of the marketplace. Since the sliced data has a fixed length with parity bits, the plagiarism check can be easily conducted by checking the statistics of parity bits.

3.7 Access control and regulations

Access control is an essential service for the marketplace. Though access control blockchains can be divided into permission and permissionless blockchains, none of the conventional blockchain access control methodology fits the purpose of an open-trading marketplace, as is the proposed marketplace. For instance, despite Role-Based Access Control (RBAC) working well within organisations to limit data consumers to only access data that pertains to their job functions, in the case of the proposed marketplace such scheme would not be possible. Since the core idea behind the proposed marketplace is to give full control of the data to whoever generated it, users should be able to choose freely to whoever they want to share their data with, instead of limiting their ability to a few groups, as is the case in RBAC. To address this issue, we need to review the openness of all elements involved in the scheme:

Completely open information. All information in this group is accessible to all participants, regardless of their role and status. In the proposed marketplace, this consists of all public encrypted data, contracts on the blockchain, public-key of Buyers, blockchain wallet addresses of Buyers and Sellers, and the optional encrypted symmetrical keys.

Partially open information. The information in this group is visible to a third party and the exclusive Buyers and Sellers, but not to the general public. It includes the sequence.

Secret. The data in this group is kept secret by Seller or Buyer. It consists of raw data, processed data, Buyer's private-key, and a Seller's optional symmetrical key pairs. Under no circumstances is the private-key of the Buyer ever shared.

The proposed marketplace only handles the open information available to the general public and keeps the partially open information secure with an access control mechanism based on directory/wallet address look up. In the case of marketplace collusion, there is no risk of exposing any secret to third parties. However, the sequence is the key for data traded among the marketplace, thus naturally becoming a tool for regulators, to provide market supervision and intervention.

3.8 Experimental Validation

In order to verify the validity of the framework, a simple implementation of the proposed framework was built. The blockchain data marketplace was built utilizing the Go Ethereum (Geth) framework [1] and consists of an IoT node powered by a raspberry pi 0W and a DHT11 temperature/humidity sensor. The raspberry board is connected to an Amazon Web Server (AWS) instance to a Structured Query Language (SQL) private server, which acts as the storage clusters of the proposed framework and store the generated data by the IoT node. A private blockchain consisting of 2 interconnected nodes is deployed via Geth also in AWS cloud, and the nodes can access the blockchain by sending interacting with its smart contracts.

In this implementation, a total of 3 smart contracts were developed. The first smart contract is responsible for recording the advertisement of data made by Sellers, with information such as owner address, data tags, timestamps, etc. This contract also serves as record keeping, as it can be queried by others to check what data was stored at any point in time in the blockchain. This contract corresponds to the prior step described in the framework, consisting of the advertisement part of the framework only.

The second smart contract is responsible for processing data requests made by Buyers to buy the data. This contract is responsible for recording the Buyer's public-key, as well as waiting for the Seller's to fill in the required information, as in steps 1 and 2. This contract is also responsible for collecting the money from the Buyer whenever a buy request is performed, and will only send it to the Seller after all steps have been completed.

Steps 5-7 have not been implemented, so a simplification has been assumed. When the Seller uploads the data, it specifies if certain addresses are in its white or black list (can or cannot buy its data). This is handled by a third and final smart contract, responsible for storing the Seller's white or blacklist of users that can/cannot access the data. Thus, it is assumed that if the Buyer is in the Seller's whitelist, the contract is then fulfilled (goes to State C) and the Buyer is then sent a link to download the data, otherwise, the transaction gets rejected and the funds are returned to the Buyer.

In addition to the blockchain and the SQL server, a web application was developed in JavaScript, which is used to browse the data stored in the marketplace, as well as to manually generate buy requests in order to verify the proposed framework. Figure 2 shows the implementation of the blockchain marketplace.

The implemented solution has three main components: an IoT sensor node, a web interface (front-end) and a storage server. The cloud is responsible for storing the SQL server, as well as the private blockchain nodes and its smart contracts. It consists of 11 steps, which are described as follows:

- Step 1: Whenever an IoT sensor node collects data about its environment it uploads the data to the server.
- Step 2: The server returns some metadata, such as the file information, identity number, and other information which can be used to identify the file.
- Step 3: The IoT node then collects all this information, together with other metadata it generated, such as what type of data was collected, data tags, timestamp, its own address, and send it to the data recording smart contract, responsible for doing the advertisement of the data.
- Step 4: The data recording smart contract is an event-based contract, thus no information is stored directly at the contract, but rather its data is stored in the logs property of the blocks of the blockchain, whereas indexed event parameters are stored as topics of the blocks.
- Step 5: By storing certain topics, such as owner address, and data tags, it allows the web interface to query the blockchain for these topics and display the information about the data in the marketplace.
- Step 6: Whenever a user wants to buy the data, it sends a request to the transaction processing contract, which then verifies if the minimum amount requested by the seller was met by the buyer.
- Step 7: The Seller is then notified and replies to the buy request, updating the contract to State B.
- Step 8: The seller is also responsible for keeping a whitelist of addresses, which can or cannot download its data, which is then queried by the transaction processing contract. As such, this contract is responsible for access control, allowing the seller to revoke any permissions to its data, or even opt-out of the marketplace.

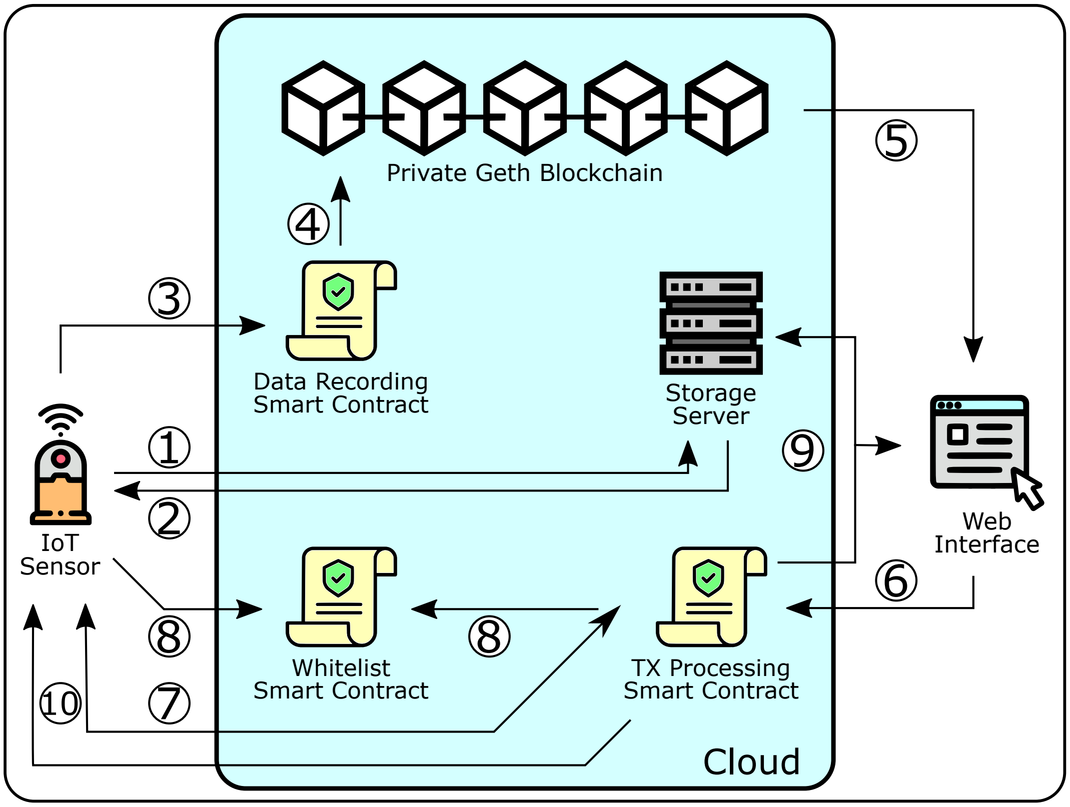

Fig. 2. Architecture of the implemented blockchain marketplace.

- Step 9: After checking the whitelist, the contract then notifies the server, which allows the Buyer to download the data.
- Step 10: Funds are transferred to the Seller for the Buyer's purchase.

With the development of this test-bed, we could verify the applicability of the proposed framework and that the concepts behind privacy preservation, access control and trustworthiness between parties work as intended [6]. As future work, we aim at improving this test-bed and considering applying to some specific scenarios, such as the ones mentioned in this paper (energy trading, or healthcare data).

4 APPLICATIONS

The creation of a blockchain framework for a data marketplace can drastically change how data is shared, generated, and consumed, which can lead to unprecedented innovations and novel applications.

By providing an immutable history of all the data generated and consumed in the marketplace, as well as providing privacy between participants, blockchain can act as a pillar for several data-oriented services. Next, we discuss some applications that can greatly benefit from this framework.

[6] For more information about the implementation, please check www.pristine-blockchain.com

4.1 Personalized Services

Current personalized services utilize all sorts of data stored by organizations to provide services such as targeted advertisements and product recommendations to customers. However, due to the centralized nature of such an approach, users of these services are not able to sell their data nor decide how it is used. On the other hand, organizations are not able to control what type of data is collected from its users, or its quality and timeliness. By utilizing the proposed blockchain framework this can change, as a decentralized and open marketplace for such data is enabled. This can bring potential benefits for both users and organizations. From the user's perspective, there would be full control and ownership of a user's data, allowing them to subscribe to services and recommendations that are meaningful or to opt-out of any unwanted schemes. For organizations, this new marketplace can motivate the collection of high-quality data from users to attract potential buyers.

4.2 Energy Trading

Energy trading is still in its infancy and some energy providers enable customers to sell the surplus energy that they have back to the grid for a fixed price. However, with the creation of a proper marketplace, a full ecosystem for peer-to-peer energy trading can be built, allowing users to trade surplus energy among themselves without the involvement of energy providers. This can lead to more competition in the market, leading to different price options as well as benefiting both customers and energy providers.

4.3 Healthcare

The proposed marketplace can also provide the privacy needed for several applications in the healthcare domain, in which potentially sensitive and personal data is shared between patients and physicians. Similar to the realm of personalized services, the trade of healthcare data in an open and decentralized marketplace can bring several benefits. From a user's perspective, they can decide to which healthcare provider their data is shared, while maintaining their privacy, allowing them to hear opinions from different experts in the field or to consider personalized treatments specifically for a condition. From a health provider's perspective, a decentralized data platform can avoid issues such as vendor lock-in, allowing access to different levels of data that were not previously accessible. Moreover, by monitoring all the information that is traded in the platform, health providers can build an entire history of health data from anonymous users, which can then be used to create certain services or to generate population-wide health information.

4.4 Artificial Intelligence

In terms of AI, it is well-known that AI models need to train with a massive amount of data to produce good outputs and that the quality of the data directly affects their performance of such models. Unfortunately, the value of good AI models is currently concentrated in a few global organizations, which have the time, money and personnel capable of collecting, storing and processing huge amounts of data. This reality, combined with the challenge of big data (in terms of volume and velocity) as well as the fact that most data contains protected, personal or sensitive information, means that there is currently no incentive to share all this information, hindering progress and innovation. The creation of a decentralized framework for data trading enables a much more level playing field for data producers, allowing them to share data with whomever they choose subject to legal limitations. Moreover, the proposed framework also enables the opposite flow by allowing different models to be bought. Lastly, the creation of a very large and broad blockchain-based data marketplace not curated by humans can also help decentralize AI, while its transparency could contribute to the explainability of AI.

4.5 IoT

It is expected that IoT will play an enormous role in the future, due to the vast number of devices deployed. However, if conventional centralized architectures are utilized, this can cause issues in terms of scalability, security and interoperability between devices and their data, forming several data islands and hindering progress and innovation. By considering a blockchain framework for data sharing, IoT can greatly benefit from its high scalability and security, as well as support peer-to-peer transactions. This can enable a much more autonomous IoT framework, while also supporting IoT interoperability by allowing untrusted devices to share and communicate data in a trustful and autonomous manner.

5 OPEN PROBLEMS

5.1 Blockchain, Machine Learning and Data Quality Control

The proposed marketplace can enable the transfer of different types of data as well as learned models. However, the complexity of such transactions as well as the overhead needed still need to be investigated. It is still not clear how the performance of the blockchain network can be affected in terms of delay and its consensus mechanisms. Thus, the investigation of different consensus mechanisms, as well as different types of blockchain (public or private) in a data marketplace context are needed. In terms of controlling the quality of both the data and the models generated in the marketplace, several questions are still open, such as how to balance the different requirements among different data collectors and consumers; how much the data is worth and how to reward meaningful data collection, and the similarity detection on trading data using homomorphic encryption inspired protocols.

5.2 How to build the Marketplace

Since the marketplace is fully decentralized, further studies are necessary in terms of how much data needs to be stored at each node, such as in IoT devices and servers. In addition, since the proposed marketplace relies on an external storage solution, it is still not clear who is to be responsible for controlling this storage or how this external server can guarantee its trustworthiness, though methods like stacking and deposits may help with commitment issues. Further concerns also lay on the automation of transactions, as current design elaborates the basics of decentralized marketplace, but the operation of such marketplaces require more effort to be user-friendly. These issues bring further questions in terms of the blockchain scalability, as well as throughput (in terms of transactions per second).

5.3 Economy and Different Players

Despite our proposed marketplace, there is still work to be done in terms of the market's economy, currency, and reward systems, as it is still not clear what type of currency will be used and if it can have any use outside of the marketplace. Another aspect that needs further investigation is how to enable this interaction between different players, since collaboration and data sharing between vendors is quite regulated. These bring potential questions to the marketplace, such as what type of blockchain can be utilized, or even if multiple blockchains should be used, such as a public blockchain overseeing other private chains supported by companies or other organizations. In such cases, the complexity of those systems can be a problem, especially in terms of the interaction between them.

5.4 Data ownership, data rights, and copyright

From the proposed framework it is clear that blockchain can help drive transparency and authenticity in data trading environments, as all data present in the blockchain is immutable. In addition, the blockchain platform is also able to enhance user experience, as now users are completely aware of who has access to their data, for what purpose and for how long. However, despite its advantages, it is still not clear how trading in this ecosystem will work. For instance, how will data trading be regulated, how will data access work, would it be similar to a copyright transfer or more like a licensing scheme? Moreover, using a blockchain environment would require companies to replace their existing systems and data utilization terms and conditions, which would require not only changes in terms of technical, but also regulatory and legal aspects. As such, this open novel and interesting future research ideas and applications, not only in terms of the development of the blockchain platform, but also in other areas, such as social sciences, regulations and copyright, and intellectual property protection.

6 CONCLUSIONS AND FUTURE WORK

In this manuscript we have reviewed current state-of-the-art in terms of big data, blockchains and smart contracts. Based on this review, we have proposed a novel blockchain based framework for data sharing, where data collectors can both securely and privately sell their data while also maintaining access control, and where data consumers can buy the generated data while keeping data quality control. We consider different types of data and also investigate the trade-offs in terms of blockchain performance, data storage and cost. As a future work, we are working on developing algorithms to validate this proposed framework as well as improving the current test-bed implementation in order to showcase a fully functional blockchain-enabled data marketplace considering some application scenarios, such as energy trading or healthcare.

ACKNOWLEDGEMENTS

This work has been supported by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK EPSRC under grant number EP/S035362/1.

REFERENCES

[1] [n.d.]. Geth Documentation. https://geth.ethereum.org/docs/. Accessed: 2021-04-20.

[2] Baoyi An, Mingjun Xiao, An Liu, Yun Xu, Xiangliang Zhang, and Qing Li. 2021. Secure Crowdsensed Data Trading Based on Blockchain. *IEEE Transactions on Mobile Computing* (2021).

[3] Prabal Banerjee and Sushmita Ruj. 2018. Blockchain Enabled Data Marketplace–Design and Challenges. *arXiv preprint arXiv:1811.11462* (2018).

[4] Suzhi Bi, Rui Zhang, Zhi Ding, and Shuguang Cui. 2015. Wireless communications in the era of big data. *IEEE communications magazine* 53, 10 (2015), 190–199.

[5] Bin Cao, Yixin Li, Lei Zhang, Long Zhang, Shahid Mumtaz, Zhenyu Zhou, and Mugen Peng. 2019. When Internet of Things meets blockchain: Challenges in distributed consensus. *IEEE Network* 33, 6 (2019), 133–139.

[6] Qian Chen, Gautam Srivastava, Reza M Parizi, Moayad Aloqaily, and Ismaeel Al Ridhawi. 2020. An incentive-aware blockchain-based solution for internet of fake media things. *Information Processing & Management* 57, 6 (2020), 102370.

[7] Konstantinos Christidis and Michael Devetsikiotis. 2016. Blockchains and smart contracts for the internet of things. *Ieee Access* 4 (2016), 2292–2303.

[8] Mateus S. H. Cruz, Toshiyuki Amagasa, Chiemi Watanabe, Wenjie Lu, and Hiroyuki Kitagawa. 2017. Secure Similarity Joins Using Fully Homomorphic Encryption. In *Proceedings of the 19th International Conference on Information Integration and Web-Based Applications and Services (iiWAS '17)*. Association for Computing Machinery, New York, NY, USA, 224–233. https://doi.org/10.1145/3151759.3151788

[9] Drasco Draskovic and George Saleh. 2017. Datapace-Decentralized Data Marketplace Based on Blockchain. *Whitepaper). Verfügbar unter https://www. datapace. io/datapace_whitepaper. pdf, zuletzt geprüft am* 13 (2017), 2019.

[10] IDC. 2019. The growth in connected IoT devices is expected to generate 79.4 ZB of data in 2025, according to a new IDC forecast. (2019).

[11] Heather Johnson. 2015. Digging up dark data: What puts IBM at the forefront of insight economy. *SiliconANGLE, October* 30 (2015).

[12] Noah Johnson. 2019. Building a Secure Data Market on Blockchain. USENIX Association, Burlingame, CA.

[13] Paulo Valente Klaine, Muhammad Ali Imran, Oluwakayode Onireti, and Richard Demo Souza. 2017. A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks. *IEEE Communications Surveys & Tutorials* 19, 4 (2017), 2392–2431. https://doi.org/10.1109/COMST.2017.2727878

[14] Paulo Valente Klaine, Lei Zhang, Bingpeng Zhou, Yao Sun, Hao Xu, and Muhammad Imran. 2020. Privacy-Preserving Contact Tracing and Public Risk Assessment using Blockchain for COVID-19 Pandemic. *IEEE Internet of Things Magazine* 3, 3 (2020), 58–63.

[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[16] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. 2020. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine* 37, 3 (2020), 50–60.

[17] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang. 2020. Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT. *IEEE Transactions on Industrial Informatics* 16, 6 (2020), 4177–4186.

[18] Robert A. J. Matthews. 1993. THE USE OF GENETIC ALGORITHMS IN CRYPTANALYSIS. *Cryptologia* 17, 2 (1993), 187–201. https://doi.org/10.1080/0161-119391867863 arXiv:https://doi.org/10.1080/0161-119391867863

[19] Satoshi Nakamoto. 2019. *Bitcoin: A peer-to-peer electronic cash system.* Technical Report. Manubot.

[20] Lam Duc Nguyen, Israel Leyva-Mayorga, Amari N Lewis, and Petar Popovski. 2021. Modeling and analysis of data trading on blockchain-based market in IoT networks. *IEEE Internet of Things Journal* 8, 8 (2021), 6487–6497.

[21] Aafaf Ouaddah, Anas Abou Elkalam, and Abdellah Ait Ouahman. 2016. FairAccess: a new Blockchain-based access control framework for the Internet of Things. *Security and Communication Networks* 9, 18 (2016), 5943–5964.

[22] Kazim Rifat Özyilmaz, Mehmet Doğan, and Arda Yurdakul. 2018. IDMoB: IoT data marketplace on blockchain. In *2018 crypto valley conference on blockchain technology (CVCBT)*. IEEE, 11–19.

[23] EU Blockchain Observatory and Forum. 2020. *Convergence of Blockchain AI and IoT.* Technical Report. 1–26 pages. https://www.eublockchainforum.eu/sites/default/files/report_convergence_v1.0.pdf

[24] IBM Corporation and GTD Solution Inc. 2020. *TradeLens Data Sharing Specification: Data Sharing Model.* Technical Report. 1–11 pages. https://docs.tradelens.com/reference/DSS_Data_Sharing_Model_V4.0.pdf

[25] C. E. Shannon. 1949. Communication theory of secrecy systems. *The Bell System Technical Journal* 28, 4 (1949), 656–715. https://doi.org/10.1002/j.1538-7305.1949.tb00928.x

[26] Hemang Subramanian. 2017. Decentralized blockchain-based electronic marketplaces. *Commun. ACM* 61, 1 (2017), 78–84.

[27] Y. Sun, L. Zhang, G. Feng, B. Yang, B. Cao, and M. A. Imran. 2019. Blockchain-Enabled Wireless Internet of Things: Performance Analysis and Optimal Communication Node Deployment. *IEEE Internet of Things Journal* 6, 3 (2019), 5791–5802. https://doi.org/10.1109/JIOT.2019.2905743

[28] Sarah Underwood. 2016. Blockchain beyond bitcoin. *Commun. ACM* 59, 11 (2016), 15–17.

[29] Qin Wang, Rujia Li, Qi Wang, and Shiping Chen. 2021. Non-fungible token (NFT): Overview, evaluation, opportunities and challenges. *arXiv preprint arXiv:2105.07447* (2021).

[30] Hao Xu, Paulo Valente Klaine, Oluwakayode Onireti, Bin Cao, Muhammad Imran, and Lei Zhang. 2020. Blockchain-enabled Resource Management and Sharing for 6G Communications. *arXiv preprint arXiv:2003.13083* (2020).

[31] Hao Xu, Lei Zhang, Oluwakayode Onireti, Yang Fang, William J Buchanan, and Muhammad Ali Imran. 2020. BeepTrace: blockchain-enabled privacy-preserving contact tracing for COVID-19 pandemic and beyond. *IEEE Internet of Things Journal* (2020).

[32] Hao Xu, Lei Zhang, Elaine Sun, and Chih-Lin I. 2021. BE-RAN: Blockchain-enabled Open RAN with Decentralized Identity Management and Privacy-Preserving Communication. *IEEE Journal on Selected Areas in Communications Special Issue on Private Information Retrieval, Private Coded Computing over Distributed Servers, and Privacy in Distributed Learning* (jan 2021). arXiv:2101.10856 http://arxiv.org/abs/2101.10856

[33] Jun Zhang, Shiqing Hu, and Zoe Lin Jiang. 2020. Privacy-Preserving Similarity Computation in Cloud-Based Mobile Social Networks. *IEEE Access* 8 (2020), 111889–111898. https://doi.org/10.1109/ACCESS.2020.3003373

[34] G. Zyskind, O. Nathan, and A. '. Pentland. 2015. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In *2015 IEEE Security and Privacy Workshops*. 180–184.

[35] Guy Zyskind, Oz Nathan, and Alex Pentland. 2015. Enigma: Decentralized computation platform with guaranteed privacy. *arXiv preprint arXiv:1506.03471* (2015).