Received May 10, 2020, accepted May 24, 2020, date of current version July 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3002823

Subscription-Based Data-Sharing Model Using Blockchain and Data as a Service

FAHAD AHMAD AL-ZAHRANI
Computer Engineering Department, Umm AlQura University, Mecca 24381, Saudi Arabia
e-mail: fayzahrani@uqu.edu.sa

The work was funded by grant number 12-INF2970-10 from the National Science, Technology and Innovation Plan (MAARIFAH), the King Abdul-Aziz City for Science and Technology (KACST), Kingdom of Saudi Arabia. We thank the Science and Technology Unit at Umm Al-Qura University for their continued logistics support.

ABSTRACT In modern times, many individuals, businesses and the Internet of Things (IoT) integrated industries collect huge amounts of meaningful data daily, which may be beneficial for other individuals and businesses as well. By utilizing this data, future trends to make the right decisions on the bases of facts and figures are analyzed efficiently. In addition to that, many new ways are paved for researchers to utilize this data in their upcoming research. However, due to some major issues like security, privacy and access control of data, data owners avoid sharing data among themselves. Another main problem is the selfish behavior of data owners. Businesses also act selfishly and invest huge amounts of money to collect and maintain the data for their benefits. Therefore, data owners are hesitant to share their data with others without the availability of a fair profit and secure data-sharing platform. Moreover, consumers are not much motivated to buy data from Data Providers (DPs) due to its bad quality and inconsistency. The data provided by data owners is mostly incomplete, outdated, heterogeneous and costly. In this paper, a subscription-based data-sharing model is proposed by leveraging the blockchain technology and Data as a Service (DaaS) concept. In this model, users subscribe to a DP for a specific period to get access to the data and pay according to the subscription plan. The DP keeps receiving revenue recurrently for a long-time, which has a huge profit margin in comparison with selling data at once. Furthermore, two major pricing models, Flat Rate Pricing (FRP) and Usage-Based Pricing (UBP), are discussed to set standards for data owners to monetize their data, and a new hybrid pricing model is also proposed. Blockchain technology is utilized in the proposed model to make it secure, transparent and immutable. To investigate the performance of the proposed model, a private blockchain network is deployed using a web interface provided by MultiChain blockchain. The simulation results demonstrate that the proposed model is feasible and efficient. The theoretical discussion proves that the proposed model is beneficial for both data owners and data consumers and has a good scope in the future for data management and trading processes.

INDEX TERMS Access control, blockchain, data-sharing, data as a service, pricing strategies, incentive mechanism, data subscription.

I. INTRODUCTION

Data is the most valuable asset of any business. The wisdom required for making correct decisions and taking the right actions depends upon the meaningful data. If the data is meaningful, relevant, complete and up-to-date, it helps in the growth of a business. If not, it is useless and resource-wasting product for the business. It is true that companies, which do not recognize the importance of data utilization, probably fail to survive in the current economic scenario because the modern marketplace's environment is data-driven. Business leaders need data to analyze the market trends on the bases of facts and figures. For this purpose, they need to get the right information on the right time to make the right decisions for the growth of the business.

One of the major problems is that businesses neither share data internally (within departments) nor externally (with other businesses). A company or a department may collect data that is beneficial for other companies or departments; however, the lack of trust and communication, eliminates this possibility. Sometimes, an individual or business must access the data of other businesses to analyze the trends and make

The associate editor coordinating the review of this manuscript and approving it for publication was Patrick Hung.

strategies and decisions accordingly. Therefore, businesses need to buy data from other businesses to fulfill their requirements. However, there are still many reasons due to which businesses avoid sharing their data with others, even though it is a profitable practice. These reasons are privacy, security, access control, data governance, selfish behavior of data owners and lack of proper business models and policies to maximize the revenue of data owners. To tackle these issues, many researchers have proposed different data-sharing models. Recent research shows that blockchain is a suitable technology to resolve these issues.

Blockchain technology, a distributed and decentralized ledger, was proposed by Satoshi Nakamoto in 2008 [1]. It is a sequence of data blocks, produced and joined together chronologically. Blockchain technology is the combination of distributed ledger, consensus mechanism, Peer-to-Peer (P2P) network and smart contracts. Blockchain and its concepts are discussed in detail in the preliminaries section (Section III). It is the main underlying technology of Bitcoin [2]. Nevertheless, it is no longer just about cryptocurrencies or Bitcoin in general. Instead, it can be seen as an emerging and innovative technology, which has major influences on various aspects of our lives. Due to the design features of blockchain-like transparency, immutability and traceability, the research community is applying blockchain in many fields such as the Internet of Things (IoT), medicine, economics and so on. To solve the security, privacy and data access problems, many researchers proposed blockchain-based data-sharing models. The list of abbreviations used in this paper is given in TABLE 1.

A. MOTIVATION

Data-sharing has become important for almost every field of life. The blockchain-based data-sharing models are gaining popularity as they make data secure and reliable while protecting it and defining several access levels. A secure data-sharing model is proposed in [3] using blockchain. This model is specifically designed for medical records sharing of patients while protecting the sensitive information. It encourages the data owners to share data; however, it lacks data authentication and communications rules between two parties. Liu *et al.* proposed a similar model for patients' medical record sharing [4]. Here, data is stored on cloud servers and their respective indexes are stored in a blockchain. Another data-sharing model is proposed in [5]. This model uses two types of blockchains: private and public. The former is used to store the actual data and later is used to store the indexes of the stored data. In the aforementioned data-sharing models, the authors did not define incentives or pricing mechanisms for data-sharing. The incentive mechanisms play a very important role in motivating the data owners to share the correct data. On the other hand, the pricing mechanisms discourage the users to request for the unnecessary data and it results in the reduced computational overhead of the system. In [6], a data-sharing model for vehicular network is proposed. The digital signatures are used to ensure the reliability and integrity of shred data. The vehicles get a reward in the form of data coins when they share data honestly. However, the authentication of vehicles is ignored in this model. Another similar data-sharing model for vehicles is proposed in [7]. In this model, all the vehicles are first registered and then share data. To prevent the vehicles from sharing false information, a reputation mechanism is designed. Using this mechanism, the vehicles sharing correct information are awarded with positive points. Whereas, the vehicles sharing false information are given negative points. In this way, the malicious nodes are identified in the network. Motivated by the aforementioned work, a blockchain-based data-sharing model with decentralized storage is proposed. This model ensures data reliability, integrity and security. Additionally, it also includes the access control method for authentication and authorization of data.

TABLE 1. List of abbreviations.

Abbreviations	Description
ABE	Attribute-Based Encryption
B2B	Business to Business
B2C	Business to Consumer
C2C	Consumer to Consumer
DaaS	Data as a Service
DP	Data Provider
DS	Data Subscriber
DoS	Denial-of-Service
EPPAS	Enhanced Privacy-Preserving Auction Scheme
FQ	Fixed Quota
FRP	Flat Rate Pricing
GHC	Gartner Hype Cycle
HPM	Hybrid Pricing Model
ICS	Indicator-Centric Schema
IoT	Internet of Things
IPFS	Inter Planetary File System
ISP	Internet Service Provider
JSON	JavaScript Object Notation
LCs	Lightweight Clients
MPC	Multi-Party Computing
P2P	Peer-to-Peer
PHI	Public Health Information
PoW	Proof-of-Work
PoA	Proof-of-Authority
PayG	Pay-as-you-Go
PPAS	Privacy-Preserving Auction Scheme
SAP	System Applications and Products
SHA	Secure Hashing Algorithm
UBP	Usage-Based Pricing
VANETs	Vehicular Ad-hoc Networks
WSNs	Wireless Sensor Networks

B. PROBLEM STATEMENT

From the literature review, it is identified that major problems addressed by researchers in data-sharing are security, privacy, access control, decentralization (single point of failure), immutability, decentralized storage, etc. In [3], [4], [8], [9], the authors propose blockchain-based data-sharing models to address the privacy, security, access control and decentralized storage problems. However, the selfish behavior of data owners is neglected. Large businesses invest huge amounts of money to collect, store and maintain data for their benefits. The question is, why data owners share their data without any benefit? To answer this question, there must be a proper data-sharing model to motivate the data owners to share their data with others securely and profitably. To further motivate the data owners, many researchers propose different incentive mechanisms. The authors in [10], propose a decentralized framework to motivate data owners to share their data by giving them incentives. The authors use blockchain and smart contracts to provide security, scalability and privacy to their model, which ensure secure data-sharing among users. However, the incentive mechanism, proposed in this model, is more generalized and is not suitable for the data-market. In such an incentive mechanism, there is no fair distribution of incentive, e.g., a person who shares 1MB data gets the same reward as the person who shares 1GB data at once. Thus, in addition to security, privacy and transparency, the proper pricing schemes and business strategies are very important to maximize the benefits of data owners. Due to heterogeneity, low quality, incompleteness and high cost of data, consumers are not willing to buy data from Data Providers (DPs). It is difficult to utilize heterogeneous data to fulfill the business demands. Data generated by IoT devices is often heterogeneous in nature [11], which creates compatibility issues on different platforms. According to a recent Gartner's research, poor data quality costs businesses an average of $ 15 million of losses per year [12]. Poor quality of data does not only impact the financial resources, but it also has a negative impact on the productivity and credibility of businesses.

C. CONTRIBUTIONS

To solve the aforementioned problems, a subscription-based data-sharing model is proposed, which is the combination of blockchain technology and Data as a Service (DaaS) concept. DaaS helps in solving issues related to real-time data access by enabling companies to access real-time data streams from anywhere using the Internet. It further eliminates the restrictions of data sources by developing proper Application Programming Interfaces (APIs). These APIs provide data in specified formats, which resolve the issue of heterogeneity. Big companies like Amazon and Systems Applications and Products (SAP) are now tapping into this slowly developing area [13]. Blockchain provides security, immutability and transparency to the data-sharing network as all the communications and deals are done through the blockchain using smart contracts (smart filters). Public key cryptography ensures the authentication and authorization of the users. The major contributions of this paper are summarized as follows:

1) A subscription-based data-sharing model is proposed that integrates blockchain and DaaS to share data in a secure and cost-efficient way. The DaaS improves the quality of data and provides ease of access. The data in DaaS is provided in a specific format, which resolves the issue of data heterogeneity. Blockchain provides security, immutability and transparency to the data-sharing network. All the communications and data-sharing are done using smart contracts.
2) In order to ensure data governance, integrity and security, the access level of data (authentication and authorization) is implemented.
3) The business and pricing models are defined for data-market to set standards for data owners to monetize their data and maximize the profit of both data owners and data consumers.
4) In order to secure the API call (communication between the data provider and consumer), digital signatures, encryption and double side verification of the sender and receiver are implemented using public-key cryptography and blockchain.
5) The private network is designed for data-sharing with a round-robin consensus mechanism. This mechanism requires low computations and energy resources as compared to the Proof-of-Work (PoW) consensus mechanism. Furthermore, the block size limit is applied to slow down the generation of blocks, which can save storage resources.
6) To investigate the performance of the proposed model, simulations are conducted and results are generated by executing a high volume of transactions. Meanwhile, the network behavior and the mining process are monitored very closely. The proposed model is implemented using MultiChain blockchain and its interface web-demo for writing a PHP script to monitor the performance of the network after a specific interval of time [14].
7) A detailed discussion on theoretical analysis and simulation results is presented to demonstrate the feasibility and efficiency of the proposed model.

Throughout the paper, the terms 'businesses and companies', 'data owner and data provider', 'data seekers, users, consumers and Data Subscribers' (DSs), 'API call and request' and 'smart contract and smart filter' are used alternatively.

The rest of the paper is organized as follows: Section II presents the literature review. Section III describes the preliminaries of the paper. Section IV presents the proposed system model. Section V shows the simulation results and gives their discussion. Finally, the conclusion and future work are presented in Section VI.

II. RELATED WORK

In this section, the literature review of blockchain-based data-sharing, access control and incentive mechanisms is given.

A. BLOCKCHAIN-BASED DATA-SHARING, PRIVACY AND ACCESS CONTROL

In [3], the authors proposed a blockchain-based model to facilitate access between users and a pool of shared sensitive data. The main purpose of the model was data-sharing in a secure manner to protect the privacy of data. However, the communication and authentication protocols and algorithms between entities were not fully defined. Authors in [4] proposed a blockchain-based model for sharing medical records to preserve the privacy of patients. To solve the problem of centralization, the authors stored data on cloud and its indexes in blockchain. Attribute-based access control mechanism and the content extraction signature schemes were used for privacy preservation in data-sharing. Furthermore, smart contracts were defined to set access permissions and to ensure data access security. In [5], authors proposed a blockchain-based secure Public Health Information (PHI) sharing model. The motivation behind this model is the improvement in health diagnostics. In the proposed work, two types of blockchains are used, i.e., private and consortium. Former is used to store the PHI and later is used to keep records of the PHI indexes. The data in PHI and patients' identities are stored using the public key cryptography, encrypted with keyword search mechanism. The proposed model ensured improvement in health diagnostics. However, the conjunctive keyword search mechanism is not used.

In [6], authors proposed a secure data-sharing blockchain based model in Vehicular Ad-hoc Networks (VANETs). The proposed model is termed as data security sharing and storage system based on consortium blockchain. In the proposed model, the data integrity and reliability is ensured using the digital signatures and bilinear pairing techniques. The vehicles, which take part in data-sharing, are given the data coins upon successful and honest data transmission. The proposed model ensured reliable data-sharing and storage. However, the authentication of the vehicles is not considered. Authors in [7] used consortium blockchain model in vehicular network to achieve secure data-sharing among vehicles. To ensure sharing of high quality data, a reputation-based model is used. To manage the reputation system, a three-weight subjective model is used in the proposed model. In the proposed work, security analysis is also performed, which further promotes security and trust.

Authors in [8] proposed a blockchain-based data-sharing model between cloud service providers. The proposed model used smart contracts and an access control policy to efficiently trace the behavior of the data owners and revoke access in case of rule and permission violation. The proposed model enabled cloud service providers to securely achieve data provenance and auditing; while, sharing medical data among users without any risk of data privacy.

Authors in [9] proposed data-sharing model by combining the Inter Planetary File System (IPFS), Ethereum blockchain and Attribute-Based Encryption (ABE) technologies. The main purpose of the model was to provide privacy and fine-grained access control of data. In addition, the keyword search function in the cipher text of the decentralized storage system was implemented and the problem of cloud server not returning correct search results was solved. However, the scheme did not define the mechanism of user's permission revocation to update access policy. Authors in [15] proposed a framework for data collection and secure data-sharing by combining blockchain and deep reinforcement learning for mobile crowd-sensing. They used blockchain technology to share data among mobile terminals with different security levels.

Authors in [16] proposed a blockchain-based secure data-sharing platform. The proposed model used IPFS to store the data in a secured manner. In the proposed model, metadata is stored in IPFS, which is further divided into sectors. The users are authenticated in the proposed model using digital signatures. The authorized users are then asked to update reviews about the data, after it is successfully shared among them. The users are then incentivized on the basis of the reviews they updated in the network.

Authors in [17] proposed an information management system to handle the patients' data, termed as MedBlock. The proposed model used blockchain technology and the distributed ledger. The consensus mechanism used in the proposed work achieved reduction in both the energy consumption and the network congestion. According to authors, MedBlock played an important role in sensitive medical information sharing in a secured manner.

To tackle the privacy issues, authors in [18] proposed Privacy-Preserving Auction Scheme (PPAS). In the proposed model, the third party is compromised by two independent entities: auctioneer and intermediate platform. The auction is done in the proposed model using homomorphic encryption technique. To further promote security, an enhanced model is proposed termed as Enhanced PPAS (EPPAS). Both the proposed techniques are assessed for the resilience against different attacks. Extensive performance evaluations were carried out to ensure the feasibility of the propose techniques. Authors in [19] proposed a secure service provisioning mechanism for the Lightweight Clients (LCs). Blockchain technology is used to provide security and privacy to the network. In [20], authors proposed PageRank mechanism to assign reputation values to the customers. The customer with the highest reputation value is then authorised to add blocks in the blockchain network. Three different types of attacks are addressed using the modified form of Proof-of-Authority (PoA).

B. INCENTIVE MECHANISMS

In order to motivate users to participate honestly in network activity and to mitigate the selfish behavior of users, researchers designed many incentive mechanisms for users.

Incentives were mostly provided in the form of cryptocurrency in these mechanisms.

Authors in [21] proposed a blockchain-based truthful incentive mechanism for distributed P2P networks. Users were rewarded after successfully delivering data to other users. The main purpose of this work was to motivate intermediate nodes (nodes between sender and receiver) to share the data. They tackled the issue of selfish behavior of user by designing proper incentive mechanism and pricing strategies. However, they did not address the issues of a sender colluding with receiver and high computational overhead. Authors in [22] proposed an incentive mechanism to improve the efficiency and utility of mobile crowd-sourcing systems. In addition, they provided a mechanism for privacy protection of users. They utilized both offline and online incentives to propose an incentive mechanism that selects the user statically and then selects a winner dynamically after bidding. However, they did not tackle the fluctuation behavior of those users, who started acting untrustworthy after achieving high reputations.

In [23], the authors designed multi-market double auction mechanism for mobile crowd-sensing network. The system was truthful and efficient to address the multi-market nature of the mobile crowd-sensing system, which did not exist in the dynamic double actions. They demonstrated that the proposed system was efficient and reliable through simulations. Authors in [24] proposed blockchain-based incentive mechanism for data storage in Wireless Sensor Networks (WSNs). The node that stored the data was rewarded with digital currency. The incentive depended on the storage of data on node. The main purpose of this mechanism was to motivate nodes to store the data of other nodes, which could resolve the issue of storage in WSNs.

To ensure privacy provisioning in crowdsensing, authors in [25] proposed a blockchain-based secure incentive mechanism. In the proposed work, the high quality contributors are rewarded with incentives. To ensure privacy and tackle the impersonation attack, a node cooperation verification scheme based on k-anonymity technique is used. The miners are prevented from comprising the users' privacy using signcryption mechanism. The proposed model ensured privacy protection. However, other forms of attacks like collusion attacks, whitewashing attacks, etc., are not considered, leaving a security loophole. Similarly, authors in [26] tackled the location privacy issue in crowdsensing. A mixed incentive mechanism is proposed in the paper, which combined virtual credit and privacy protection. The proposed model is a three layered network, which are: blockchain, confusion mechanism and intelligent crowd sensing network. The performance evaluation of the proposed model in a real environment proved the efficacy in users' privacy protection.

For ensuring secure service provisioning in IoT, it is required that proper incentives are given to motivate the users. Authors in [27] used a consortium blockchain-based secure provisioning model, which involved a fair payment system for LCs. In the proposed model, PoA consensus mechanism is used to authorize the users. The service providers are rewarded with cryptocurrency upon successful service provisioning to the users. The proposed model is successful in achieving secure service provisioning to users. However, the incentive mechanism is not efficient, as it relied only the reputation values.

C. PRICING SCHEMES FOR DATA-SHARING

Authors in [28] defined that how big data can be priced fairly and reasonably. They proposed a pricing model for big personal data based on tuple granularity by comparatively analyzing the existing strategies and pricing models. Authors in [29] presented the insights about data market by interviewing seven data vendors about the key challenges related to data pricing strategies. Furthermore, they discussed about the potential market situations, pricing models and strategies. Authors in [30] explained the pricing schemes implemented by Internet Service Provider (ISP). They explained the current as well as the new pricing strategies. Furthermore, they discussed about the pricing schemes based on connection speed and congestion sensitivity.

In this paper, a blockchain-based data-sharing model is proposed. This model is very different from existing solutions. It is based on the subscription business model and DaaS model. Blockchain is used for security, privacy and transparency in the proposed model. Furthermore, motivated from existing pricing strategies, a new pricing strategy is proposed to set standards for data monetization.

III. PRELIMINARIES

In this section, the background knowledge of the blockchain is discussed.

A. BLOCKCHAIN

The blockchain is a distributed and decentralized ledger that records all the transactions of the P2P network. All network participants keep the same copy of the ledger. The main purpose of blockchain technology is to remove the third party, i.e., the central authority so that no single member can control the whole network [9]. Blockchain is a sequence of blocks, which are linked together through cryptographic hashes. Each block contains the hash of the previous block and it resists the modification of data in the blockchain. It is the main underlying technology of cryptocurrency but it is no longer just about cryptocurrency. Blockchain has grasped the attention of many industries like medical, financial, identity management, asset management, government agencies, etc., [31]. The basic concepts and definitions are defined below:

- **Block:** A block is a collection of the verified transactions, which is appended at the end of a blockchain. It has two parts: block header and body. Former contains hashes of the current and previous blocks, nonce and timestamp and later contains valid transactions.
- **Node:** Any computer or electronic device connected to the Internet can be the part of the blockchain network

and act as a node. A node maintains a copy of blockchain and takes part in consensus mechanism to validate transactions.

- **Miner:** Miner is a blockchain node that mines (creates) new blocks. The miner processes and validates transactions, packs these transactions in a block, adds that block in the blockchain and gets rewarded for each block's creation.
- **Mining:** The main purpose of mining is to verify the transactions and make the blockchain immutable. Miners solve a complex mathematical problem to compute valid nonce for the block. When the mining is done, the block is broadcasted to every node of the network.
- **Nonce:** It is a random number added in a block's hash to calculate the hash in a specific range depending on the difficulty level of the mining.
- **Smart Contract:** A smart contract consists of a set of rules, which are important to complete a transaction between two or more parties. It is a computer program written in a special programming language, i.e. solidity. Once it is deployed on the network it cannot be modified. So, the rules written in a smart contract are the same for every user [32], [33].

B. PUBLIC KEY CRYPTOGRAPHY

Public key cryptography, also known as asymmetric cryptography, is an encryption method that uses two keys: a public key and private key. Data is encrypted using the public key and decrypted using the private key. It is not possible to know the private key from the public key because both are independent keys. Users can share public keys freely with others to allow them to encrypt data and to verify digital signatures. The private key must be kept secret, which ensures that only an authorized owner can create digital signatures and decrypt data.

- **Digital Signatures:** Digital signature is created with the user's private key and is verified using the user's public key.
- **Encryption:** Data is encrypted using the receiver's public key.
- **Decryption:** The specific receiver decrypts the data using his private key.

C. DaaS

DaaS is a data delivery and distribution model in which data is made available to consumers over the Internet. It uses the cloud as primary technology that supports APIs for communication. The data in this model is stored on the cloud and is accessible through APIs from different devices. DaaS provides the ability to move data easily from one place to another. It also provides ease of administration, collaboration, global accessibility and compatibility among different platforms. Moreover, it also reduces the maintenance and delivery costs.

IV. SYSTEM MODEL

In this section, the proposed subscription-based data-sharing model based on blockchain and DaaS is explained. The proposed model is presented in FIGURE 1. To elaborate the proposed model, it is divided into six modules, i.e., entities, business models, subscription-based data-sharing model, pricing models, data access mechanism using blockchain, authentication and authorization. These modules are further elaborated in the following subsections.

A. ENTITIES OF PROPOSED MODEL

The system model consists of two main entities [9]:

1) DATA PROVIDER

DPs are individuals or organizations that own or collect data on daily bases such as IoT device owners, industries, hospitals, social networking sites and other businesses, which are ready to monetize their data.

2) DATA USER/SUBSCRIBER

DSs are individuals or businesses that need data. They subscribe to a DP to fulfill their data requirements. After subscription, they are authorized to access the specific data provided by DPs. DSs periodically pay DPs according to their chosen pricing model at the time of subscription.

B. BUSINESS MODELS

There are three types of data-sharing business models, which are as follows:

1) BUSINESS TO BUSINESS (B2B)

The data generated by one business might be beneficial for other businesses as well. In this model, one business (organization, industry, institute or any other business) can subscribe to one or more businesses simultaneously to access their data and utilize it to get benefits.

2) BUSINESS TO CONSUMER (B2C)

Businesses can also share their data with individuals. A person who is looking for a data can simply subscribe to a DP. The individual will be authorized to access data after he/she pays the subscription fee to the DP.

3) CONSUMER TO CONSUMER (C2C)

In the proposed model, an individual can share data with another individual in a P2P fashion. The Data owner designs a proper interface for its data and hosts it on any webserver. Other consumers can access these APIs after buying subscriptions. The data owner receives digital currency from consumers' wallets when they use the data.

C. SUBSCRIPTION-BASED DATA-SHARING MODEL

The proposed model is a subscription-based data-sharing model, which is implemented using the DaaS concept to deliver data to consumers. So, every DP has pre-defined

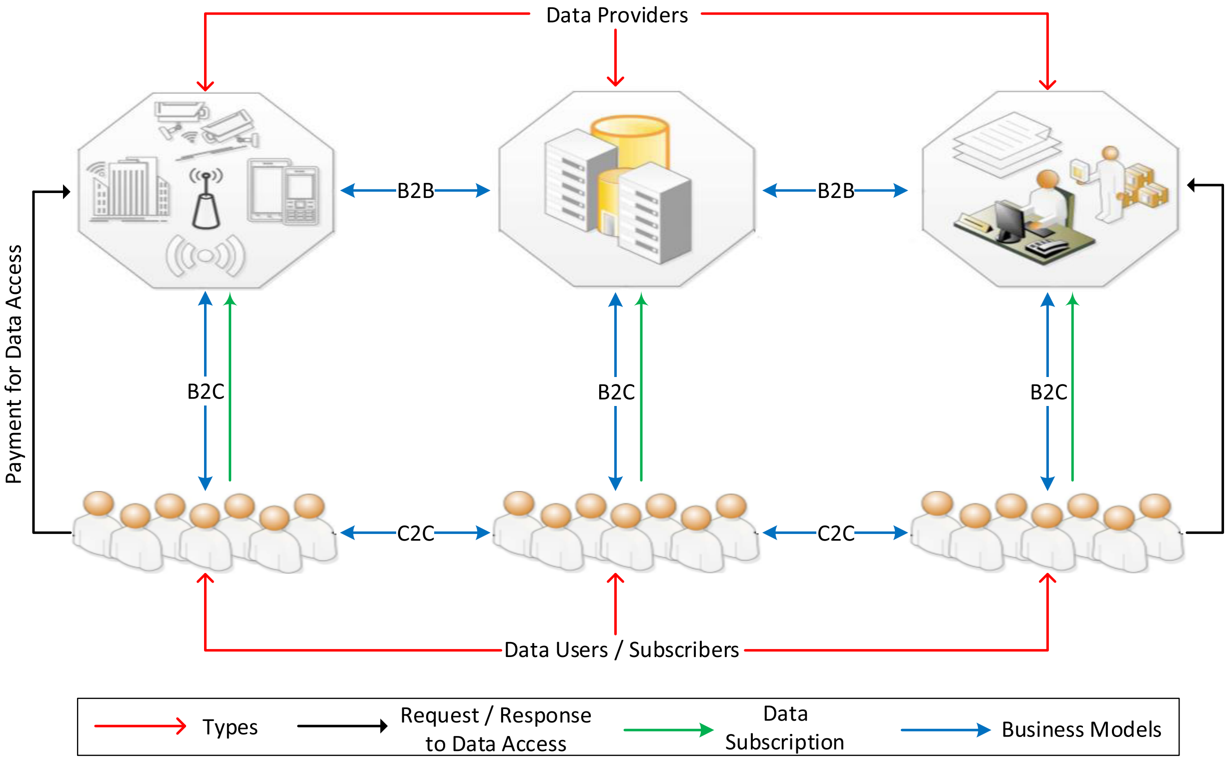


FIGURE 1. **Proposed system model.**

subscription plans for users to subscribe to its data services. In this work, subscription plans are not discussed in detail because these plans are designed by the DPs according to their policies and benefits. Anyone can view and subscribe the data services (subscription plans) provided by the DPs. To subscribe to any subscription plan of DP, the user must complete the subscription process as shown in FIGURE 2, which is common for all business models. The DPs and their subscription plans are publicly available on the blockchain. Every DP has a unique public key (DP_{PK}) and its subscription plan has a unique id (Sub_{id}). Equation 1 shows that how a unique identifier can be derived for a specific subscription plan.

$$Sub_{id} = SHA256(DP_{PK} \cup Sub_{Name} \cup T) \qquad (1)$$

where, Secure Hashing Algorithm (SHA)-256 converts data of any size into the hash of 256 bits, Sub_{Name} is the name of the subscription plan and T is the current timestamp. $\cup$ symbol is used as a concatenation operator.

The user selects a DP and suitable subscription plan by providing public key ($User_{PK}$) and signs the message with its private key ($User_{PrK}$) as shown in step 3 of FIGURE 2. The subscription request is stored in the blockchain and at the same time, DP receives a subscription request. The DP verifies the signature using the user's public key. At step 5, DP sends a payment request to the user that contains the subscription price, wallet address of DP, Sub_{id}, pricing model (defined in section IV-D) and subscription period (Sub_{Period}). The user pays subscription amount using the wallet address of a DP and record is stored in the blockchain. After the confirmation of the payment, DP grants access to the user for the data stream for a certain period and the record is stored in the blockchain. Step 10 checks the subscription expiry. When the subscription period ends, the user has to renew the subscription by repeating steps 5-9. Else, DP revokes user's access.

To provide DaaS to the users, the DP must design different types of APIs for different categories of data, users and devices. The data must be provided in a specific format to avoid compatibility issues. Nowadays, JavaScript Object Notation (JSON) format is commonly used, which is a lightweight format for interchanging data. It is popular due to its lightweight, robustness, compatibility and support for any size of audio and video. This approach provides data that is easy to parse for all types of devices and browsers.

D. PRICING MODELS

Selecting the price of a data service is an important task to undertake. However, it is both a tedious and time-consuming task. A business may diminish its profit by choosing the price of services too low or it may lose customers by keeping the prices too high. This is a very important and challenging task to balance the prices of services to maximize profit without losing customers. To address this challenge, the literature related to pricing models of data-market [29] and ISP [30] is reviewed to get motivation from their pricing models.

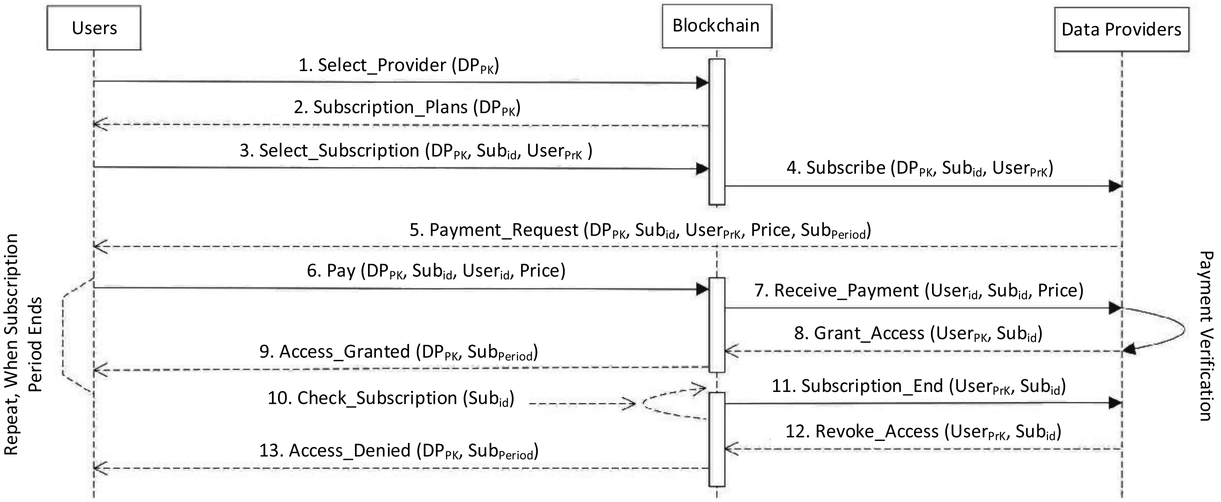

Users
Blockchain
Data Providers
1. Select_Provider (DP_{PK})
2. Subscription_Plans (DP_{PK})
3. Select_Subscription (DP_{PK}, Sub_{id}, $User_{PrK}$)
4. Subscribe (DP_{PK}, Sub_{id}, $User_{PrK}$)
5. Payment_Request (DP_{PK}, Sub_{id}, $User_{PrK}$, Price, Sub_{Period})
6. Pay (DP_{PK}, Sub_{id}, $User_{id}$, Price)
7. Receive_Payment ($User_{id}$, Sub_{id}, Price)
8. Grant_Access ($User_{PK}$, Sub_{id})
9. Access_Granted (DP_{PK}, Sub_{Period})
10. Check_Subscription (Sub_{id})
11. Subscription_End ($User_{PrK}$, Sub_{id})
12. Revoke_Access ($User_{PrK}$, Sub_{id})
13. Access_Denied (DP_{PK}, Sub_{Period})
Repeat, When Subscription Period Ends
Payment Verification

FIGURE 2. **Subscription process sequence diagram.**

There are two major pricing models, which are: Flat Rate Pricing (FRP) and Usage-Based Pricing (UBP).

1) FRP

In FRP, consumers pay an initial subscription fee for registration and then pay a fixed amount after a specific period, like weekly, monthly or yearly. They do not pay for every bit of data, they consume. FRP model has some advantages like it is easy to calculate and avoid extra costs of administration, tracking and usage. However, users in this model are motivated to use more resources, which lead to the issue of congestion.

2) UBP

In the UBP model, the initial subscription fee is fixed and the usage fee is variable. A consumer has to pay for each request for data. A UBP model works very well for congestion control because the consumer utilizes data with care to minimize the bill. However, its main problem is that it enforces usage price, whether the network is congested or not. It can collapse the whole revenue model because a large number of consumers may be driven away.

To address the problems mentioned above, three simple categories for DP to price their APIs seem promising. These categories are Fixed Quota (FQ), Pay-as-you-Go (PayG) and Hybrid Pricing Model (HPM).

1) **FQ:** FQ model [30] allows consumers to buy a fixed number of API calls for a specific time. Consumers are restricted and they cannot exceed the limit of allowed calls. The subscription is over when the number of calls or time is over. The flow of the FQ model is shown in Algorithm 1.
2) **PayG:** In PayG pricing model [30], a consumer has to pay for each request (API call). Consumers can consume the services until they have balance in their wallets. The price increases linearly with data usage and the price for a specific period is hard to assess. The flow of PayG model is demonstrated in Algorithm 2.
3) **HPM:** In HPM, the features of both FQ and PayG price models are combined to develop a hybrid model. HPM allows consumers to buy a fixed number of API calls for a specific period. However, if a consumer exceeds the subscription limits, he will have to pay a small fee per call (when FQ is over, then PayG comes into action). The benefit of this model over FQ and PayG is that it does not stop suddenly with the end of the subscription. Algorithm 3 shows the flow of HPM.

Algorithm 1 FQ Pricing Model

```
initialization: User = Requester
if isUserRegistered(User) == true then
    if $User.RequestCount \leq Subscription.AllowedRequests$
    $\&CurrentTime \leq Subcription.EndingTime$ then
        User.RequestCount++;
        Return "Access granted";
    else
        Return "Access denied, subscription is over";
    end if
else
    Return "Access denied, invalid user";
end if
Result: Access Granted/Access Denied
```

E. DATA ACCESS MECHANISM USING BLOCKCHAIN

After the completion of the subscription process, DS can access the data securely through APIs provided by DP. The main steps of data access mechanism are given below:

1) Before sending a request, the user digitally signs and encrypts the request using his private key and DP's public key.

Algorithm 2 PayG Pricing Model

```
initialization: User = Requester
if $isUserRegistered(User) == true$ then
    if $User.Balance \leq Request.Price$ then
        Transfer(amount:Request.Price, from: User, To: DP
        Return "Access granted";
    else
        Return "Access denied, insufficient balance";
    end if
else
    Return "Access denied, invalid user";
end if
Result: Access Granted/Access Denied
```

Algorithm 3 HPM Pricing Model

```
initialization: User = Requester
if $isUserRegistered(User) == true$ then
    if $User.RequestCount \leq Subscription.AllowedRequests \& CurrentTime \leq Subcription.EndingTime$ then
        User.RequestCount++;
        Return "Access granted";
    else
        if $User.Balance \leq Request.Price$ then
            Transfer(amount:Request.Price, from: User, To: DP
            Return "Access granted";
        else
            Return "Access denied";
        end if
    end if
else
    Return "Access denied, invalid user";
end if
Result: Access Granted/Access Denied
```

2) The user sends the request to DP.
3) DP decrypts the request using the user's public key and verifies the digital signature. If the signature is not valid, DP simply discards the request.
4) If the digital signature is valid, then DP verifies the user 's registration and subscription by executing the algorithms, which are defined in section IV-D. This step is different for different users, depending on the pricing scheme they selected while buying a subscription.
5) After executing a specific algorithm, the user gets a response from the DP. If the conditions given in the algorithm are fulfilled, the user gets access to data, else the DP sends an error message according to the failed condition.
6) Before sending the response, DP encrypts it with the user's public key and digitally signs it with its private key.
7) The DP sends the response of API to the user.
8) After receiving the response, the user ensures that the response is coming from an authentic DP by verifying the digital signature using DP's public key and then it decrypts the request.
9) A user can view the history of his subscriptions and the status of current subscriptions by querying the blockchain.

These steps are also shown in FIGURE 3. The process of verifying users and messages is defined in section IV-F.

F. AUTHENTICATION AND AUTHORIZATION

In this section, authentication and authorization processes are discussed in detail.

1) AUTHENTICATION

The public key cryptography and digital signatures are used to authenticate users and messages in a blockchain. Before sending a message, the sender signs the message using his private key and the receiver verifies the signature using the sender's public key. If the signature is valid, the transaction is completed. Else, the message is discarded by the receiver. The basic steps of the authentication process are signing the message and verifying the message, which are discussed below.

Signing the Message:

1) The one-way hash of the data (message) is calculated first.
2) The hash of the data is then encrypted using the sender's private key.
3) The encrypted hash along with metadata is used to generate a digital signature. The signed message is then sent to the receiver.

Verifying the Message:

1) Firstly, the hash is decrypted using the sender's public key.
2) Secondly, the receiver generates the hash of the same data.
3) Lastly, the receiver compares these two hashes (decrypted hash and generated hash). In return, the signature is either valid (both the hashes are the same) or invalid (the signature is not created with the sender's private key).

2) AUTHORIZATION

In a multi-user network, permissions and access control procedures are very important to control users. Access levels help an administrator to allow users to access limited resources and restrict them from using sensitive resources. In order to ensure the authorization, some access levels (permissions), provided by MultiChain blockchain [34], [35], are implimented. Some of them are discussed below:

- **Connect:** It allows users to connect to the blockchain and see its contents. This permission is for users who want to access the specific data. After buying the subscription, the DP grants connect permission to the users.

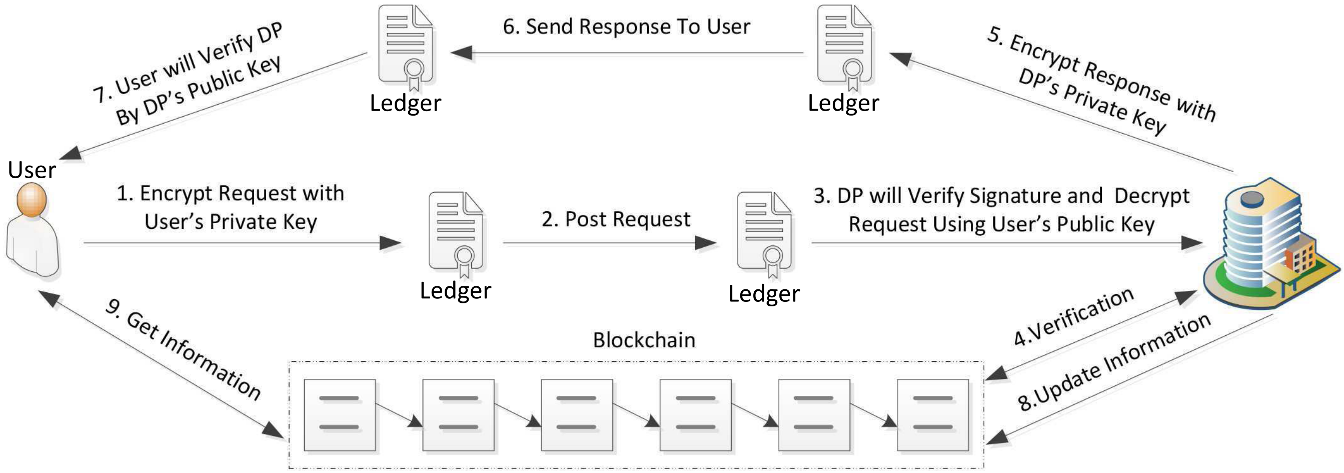

User
7. User will Verify DP By DP's Public Key
Ledger
6. Send Response To User
Ledger
5. Encrypt Response with DP's Private Key
1. Encrypt Request with User's Private Key
Ledger
2. Post Request
Ledger
3. DP will Verify Signature and Decrypt Request Using User's Public Key
9. Get Information
Blockchain
4.Verification
8.Update Information

FIGURE 3. Data access mechanism through APIs using Blockchain.

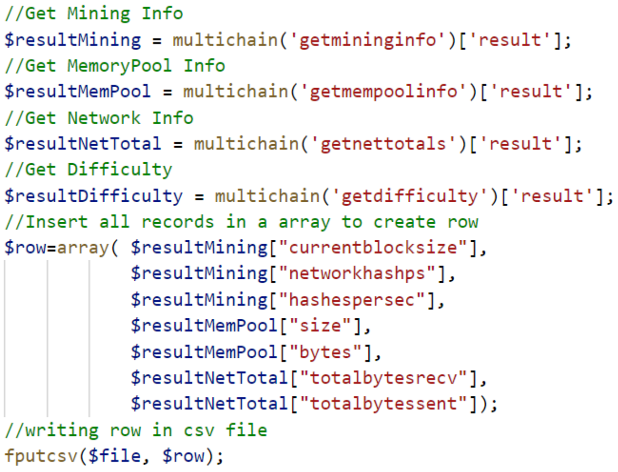
- **Send:** It allows the user to send digital currencies to others. In the proposed scenario, the send and receive permissions are granted by default.
- **Receive:** It allows the user to receive digital currencies from others.
- **Issue:** This permission is assigned to a person, who is responsible for creating the new digital assets, which are further issued to any other user for a specific purpose.
- **Create:** It allows to create data streams. DPs grant this permission to their authorized staff members to enable them to create and manage new data streams and subscription plans.
- **Mine:** It is used to allow network node to mine new blocks. Admin grants this role to one or more nodes of the network.
- **Activate:** This permission is granted to an authorized person of a DP to change connect, send and receive permissions for other users. This permission is also granted to the staff members of DP to manage the permissions of users. It is just like a sub-admin, who has limited administration privileges.
- **Admin:** This permission is assigned to the higher authority of DP's administrator. Admin permission allows DP to change all permissions for other users, including issue, mine, activate and admin.
- **Custom:** MultiChain also allows to create custom permission. Rules for custom permissions are defined and enforced by smart contracts. DPs define their own permissions to further refine the access policies.

V. SIMULATION RESULTS AND DISCUSSION

In this section, the simulation results of the proposed model are discussed in detail.

A. SIMULATION SETUP

The proposed blockchain-based data-sharing model is implemented on the machine having Windows 10 operating system, 4GB RAM and Intel(R) Core (TM) i3 CPU @ 1.7GHz processor. A private blockchain network is deployed using MultiChain [34], which is an open-source platform for developing blockchain. The proposed private blockchain includes four nodes: one of them is DP and the others are DSs. All the privileges are granted to DP, which further grants and revokes privileges to other nodes. In addition, there are two miner nodes: one is DP itself and the other is selected by DP. Furthermore, MultiChain-web-demo [14] is used for user interface, which is a simple web interface written in PHP for MultiChain blockchains. In order to investigate the performance of the proposed model, a large number of transactions are generated and tested against Denial-of-Service (DoS) attacks. The performance of blockchain is monitored by writing a simple PHP script as shown in FIGURE 4. Finally, the results are ploted and discussed in the following subsections.



```
//Get Mining Info
$resultMining = multichain('getmininginfo')['result'];
//Get MemoryPool Info
$resultMemPool = multichain('getmempoolinfo')['result'];
//Get Network Info
$resultNetTotal = multichain('getnettotals')['result'];
//Get Difficulty
$resultDifficulty = multichain('getdifficulty')['result'];
//Insert all records in a array to create row
$row=array( $resultMining["currentblocksize"],
            $resultMining["networkhashps"],
            $resultMining["hashespersec"],
            $resultMemPool["size"],
            $resultMemPool["bytes"],
            $resultNetTotal["totalbytesrecv"],
            $resultNetTotal["totalbytessent"]);
//writing row in csv file
fputcsv($file, $row);
```

FIGURE 4. PHP code to get network information.

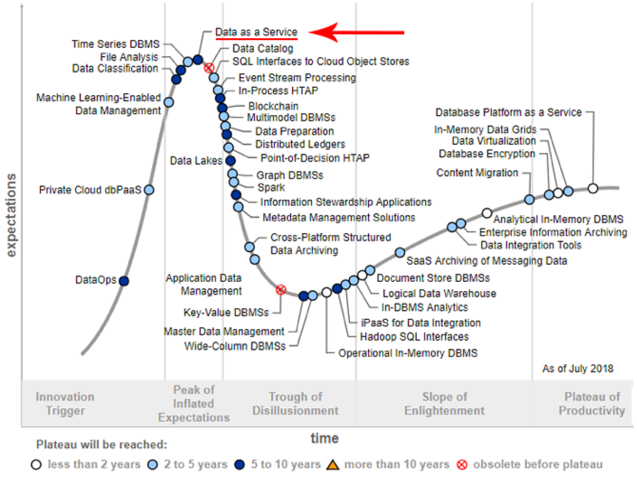
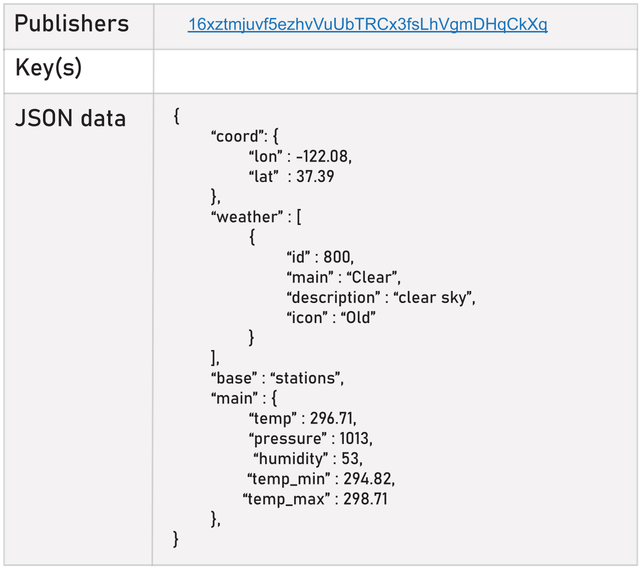


FIGURE 5. Gartner hype cycle for data management.



Publishers	16xztmjuvf5ezhvVuUbTRCx3fsLhVgmDHqCkXq
Key(s)	

JSON data

```
{
    “coord”: {
        “lon” : -122.08,
        “lat”  : 37.39
    },
    “weather” : [
        {
            “id” : 800,
            “main” : “Clear”,
            “description” : “clear sky”,
            “icon” : “Old”
        }
    ],
    “base” : “stations”,
    “main” : {
        “temp” : 296.71,
        “pressure” : 1013,
         “humidity” : 53,
        “temp_min” : 294.82,
       “temp_max” : 298.71
    },
}
```

FIGURE 6. API Response in JSON Format.

B. ANALYSIS OF DaaS AND SUBSCRIPTION-BASED DATA-SHARING

In this section, the importance of DaaS and subscription-based data-sharing are discussed. In addition, some advantages and future trends of these services are also highlighted.

1) DaaS

DaaS is an emerging technology and many big companies are adopting this technology for data trading [13]. However, spending too much money on new technologies has associated risks, e.g., if the technology does not stick around for a long time, then it is a great loss. In order to mitigate this risk, the business trends and technologies are analyzed with the help of the Gartner Hype Cycle (GHC) for data management 2018 [36]. According to GHC 2018 as shown in FIGURE 5, taken from [36], DaaS is at the peak of inflated expectations in data management strategies. This hype cycle shows that DaaS has some serious sticking power because of its ability to provide big data in human-readable form and it also offers extraordinary insights about consumers' behavior. It is expected that in the near future (5 to 10 years), DaaS will attain great maturity and will be on the plateau of productivity.

2) WHY SUBSCRIPTION-BASED DATA-Sharing?

There are several advantages of subscription-based data-sharing. Some of them are mentioned below:

- **Recurring Revenue Streams:** Instead of selling data as a product at once, subscription-based data-sharing has a huge profit margin. Businesses keep receiving revenue for a long time on a daily, monthly or yearly bases.
- **Availability:** Instead of visiting the data houses or waiting for the data to come from another source, the data is easily available on any web browser or mobile device.
- **Application Interfaces:** It provides possibility to download information and provides real-time connectivity of data. So, information and presentations are updated in different applications, in a very short time.
- **Higher Consumer Retention:** A subscription-based model guarantees that the business is retaining a large number of consumers on a consistent basis. The subscription-based model provides flexibility to increase revenue from the existing consumers by upgrading their policies.

The DaaS technology and subscription-based business model are the perfect combination for data management and monetization as they provide ease of administration, collaboration and compatibility among different platforms along with the increase in the revenue of data owners.

3) COMPATIBILITY AND HETEROGENEITY

The heterogeneity of data and compatibility of different data formats are some other issues faced by developers and service consumers. DaaS provides proper APIs for data access and returns data in a specified format. The most commonly used format is JSON, which has several advantages like lightweight and fast. It also allows server-side parsing and provides a wide range of compatibility with different operating systems, browsers and devices. Due to increased compatibility, it also resolves the issue of heterogeneity. JSON is the best format for data-sharing of any size and type. FIGURE 6 shows data in JSON format.

C. PRICING MODELS

The pricing strategies have a very huge impact on the profit as well as on the performance of the proposed model. It is obvious that defining good and fair pricing strategies can increase the profit margin for the business. In addition, by implementing FQ, PayG and HPM strategies, the network traffic can be controlled and network congestion can be avoided.

As it is discussed in subsections IV-D1 and IV-D2, users in FRP are motivated to utilize more resources, which create congestion problems. Whereas, UBP is costly and diminishes the whole revenue model because users are not motivated to adopt this model. FQ, PayG and HPM play a vital role in establishing a sense of equilibrium between resource utilization and revenue. These models restrict users to use the limited resources according to their requirements. By keeping a balance between the price and resource utilization, a data owner can retain customers for a long time and maximize its revenue. Furthermore, by restricting consumers to the limited number of calls and resources, both the network congestion and the downtime of servers are decreased. These models can also resolve the issues of traditional mechanisms and provide revenue to DPs as they deserve. The revenue generated by any DP is fair and justifiable.

D. SECURITY AND PRIVACY ANALYSIS

According to MultiChain white paper [37], the public key cryptography enables each transaction to be signed by a sender to prove that it owns the private key matching to a particular public key. MultiChain uses this property to control blockchain's access to a list of authorized users only, by expanding the handshaking process that happens when two nodes connect. The main steps of the handshake process of MultiChain blockchain are as follows [35], [37]:

1) Every node presents its public key address on the permitted list.
2) Every node verifies the public key address in its permitted list.
3) Every node sends a challenge message to the other nodes.
4) Finally, every node sends back a digital signature of the challenge message to prove its ownership of the private key related to the public key address. If any node is not satisfied with the signature, then it terminates the connection.

Before establishing a connection, each node verifies the identity of the other nodes using the extended handshake process. This process ensures the authorization of a node before any transaction, which reduces the risk of any malicious activity. Every API request is encrypted and digitally signed by a user and the response is encrypted and digitally signed by a DP. This double-sided security check avoids unauthorized access of data.

In MultiChain blockchain, all access rights are granted and revoked through network transactions. The administrator of the network manages privileges of other users, which restrict users to their authorized area only. Access rights ensure that users can only access limited resources, to which they are authorized.

Furthermore, by keeping the block size of blockchain smaller (limited), the network becomes more distributed. A more distributed network is difficult to tamper as compared to a less distributed network. This point is discussed in section V-E1 in detail.

E. PERFORMANCE OF BLOCKCHAIN NETWORK

In order to monitor the performance of blockchain, some metrics of the network like block size, hashes per second, memory pool size and bytes sent and received by the network are analyzed carefully.

1) BLOCK SIZE

In a blockchain, the total number of transactions or size of all transactions in the block is called block size. In the early stages of blockchain, the block size was not limited. However, the block size is now limited to reduce the threat of spams and DoS attacks. This has created a dispute and divided developers into two parts: individuals who are in favor of limited block size and individuals who are against the block size limit. There is a trade-off between small and large block sizes. Their pros and cons are discussed below.

a: LARGE BLOCK SIZE (WITHOUT SIZE LIMIT)

There are many reasons to keep the block size bigger, e.g., due to the large block size, a miner receives more incentives and does not need to pay high transaction fee. Furthermore, it decreases the mining time of a transaction by packing all the transactions of the memory pool in a single block. It increases the network performance due to the increase in transactions per second. However, due to the larger block size, full nodes become more expensive to operate. It decreases the number of full nodes, which leads the network towards centralization. In addition, it also creates mining problems like forking.

b: LIMITED BLOCK SIZE

The limited block size makes full nodes cost-effective as they require less computational power. It makes the network more distributed. As the network becomes more distributed, it mitigates the risk of spam activities and it becomes difficult for a hacker to tamper any information. In addition, it also resolves issues related to mining, storage and scalability. However, it slows down the network when more users make transactions over the network. The reason is that the limited block size reduces the number of transactions per second, which increases the mining time of transactions. A user pays more transaction fee to get priority in confirmation of the transaction, which makes the transaction expensive.

c: ANALYSIS

The block size of proposed network with respect to time is shown in FIGURE 7. The maximum block size limit is set to 1 MB. As discussed earlier, the limited block size makes the network distributed due to the low cost of full nodes. This helps to mitigate the risk of DoS attacks, data tampering and other mining issues like a hard fork. In addition, the size of the blockchain does not increase exponentially. The proposed model is a private network for data-sharing, so, the transaction fee and reward for miners is not considered in this scenario. The main concern is storage cost that is why the limited block size option is adopted for this scenario.

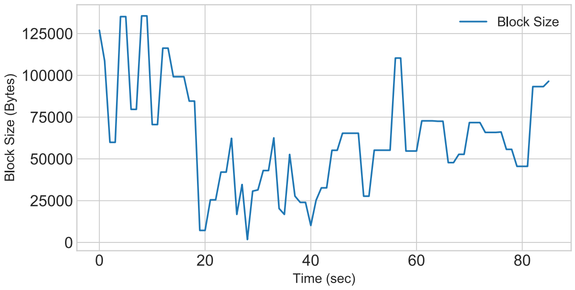


FIGURE 7. Block size of private network.

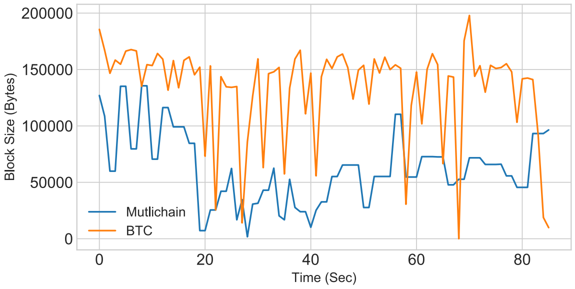


FIGURE 8. Block size comparison between Bitcoin and MultiChain.

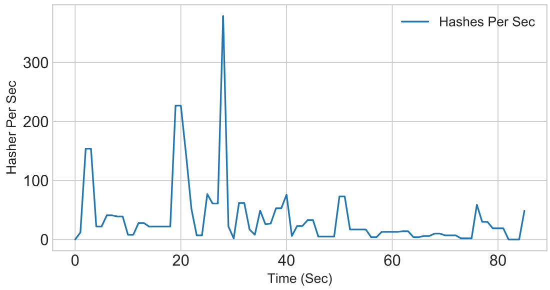


FIGURE 9. Hashes generated per second.

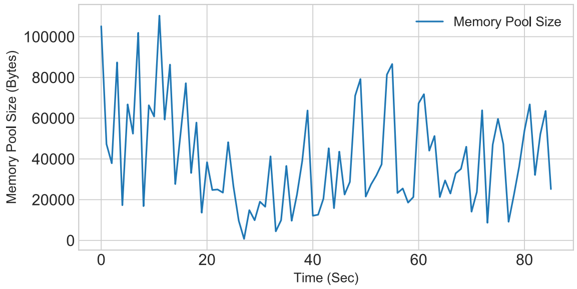


FIGURE 10. Size of memory pool.

The block size dynamically increases with the increase in the number of transactions per second. However, the maximum limit of block size is 1MB. Finally, the block size of the proposed blockchain is compared with Bitcoin's blockchain as shown in FIGURE 8. This figure depicts the block size of multichain changes over time; however, it does not cross its limit of 1 MB. On the other hand, the block size of bitcoin is more than multichain and it also has more fluctuations. The limited block size prevents the malicious nodes from adding invalid transactions and inefficiently increasing the size of a block. This size limitation also saves storage space.

2) HASH RATE AND MINING DIFFICULTY

Hashes generated per second to solve the mathematical puzzle by miners is known as hash rate. Hash rate and mining difficulty are directly proportional to each other. The mining difficulty depends on the total hashing power of the network. So, if the hash rate of the network increases, then difficulty level is adjusted to slow down the block generation rate. Bitcoin limits the block generation rate to 1 block per 10 minutes. FIGURE 9 shows the hashes generated per second in the proposed network. The unusual spikes show that the miners made more guesses to mine the block. Guessing the hash for the block, which is less than or equal to mining difficulty, is a random process. Sometimes, it is solved in few guesses and sometimes it takes millions of guesses to generate a valid hash for the block. That is why the plot in FIGURE 9 shows the irregular behavior. As the aim of this work is to keep the block generation rate on a moderate level, so, the difficulty is set is such a way that the hash rate is mostly low. Although, spikes are also present as the miners solve the puzzle randomly, so, the pattern of this plot cannot be smooth. This transaction rate is ideal for the proposed model as the block generation speed is low.

F. MINING POOL AND NETWORK UTILIZATION

FIGURE 10 and FIGURE 11 show the size variation of mining pool. The size of the mining pool varies from time to time. Its size increases when the number of transactions per second increases and vice versa. If we compare the size of mining pool, as shown in FIGURE 10 and FIGURE 11, with the size of the block as shown in FIGURE 7, then it is concluded that size of mining pool is directly proportional to the size of the block. As the size of the mining pool increases, the block size also increases, accordingly. The reason behind this is that an increase in the number of transactions requires a miner to pack more number of transactions in the block, which further increases the block size. However, in the case of the proposed model, the block size is limited to 1 MB only, so, the increase in the mining pool will not increase the block size. Instead, the remaining transactions will be added to the next block. The patterns of plots in both Figure 10 and Figure 11 are showing the same behavior. The reason for their similarity is that when the number of transactions increases in the memory pool, the required space for these transactions is also increases accordingly. Moreover, during the 60*th* second, the memory pool size in Figure 10 shows less increment as compared to the number of transactions on same interval in Figure 11. The reason is that the size of all transactions is not the same, so, this type of variation is natural to occur.

TABLE 2. Comparison with existing schemes.

References	[8]	[3]	[9]	[4]	[10]	[21]	Proposed Model
Blockchain	✓	✓	✓	✓	✓	✓	✓
Access Control	✓	✓	✓	✓	✗	✗	✓
Security	✓	✓	✓	✓	✗	✗	✓
Privacy	✓	✓	✓	✓	✗	✗	✓
Incentive	✗	✗	✗	✗	✓	✓	✓
Pricing Models	✗	✗	✗	✗	✓	✓	✓
Heterogeneity	✗	✗	✗	✗	✗	✗	✓
Selfish Behavior of DP	✗	✗	✗	✗	✓	✓	✓
Consensus Algorithm	Not-mentioned	Not-mentioned	PoW	PoW	PoW	PoW	Round Robin

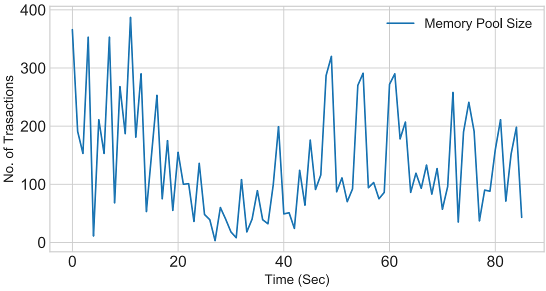


FIGURE 11. Number of transactions in memory pool.

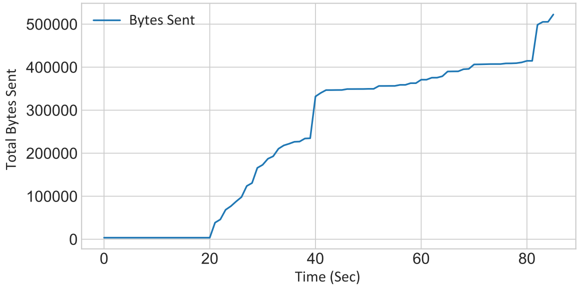


FIGURE 12. Total bytes sent by one node to network.

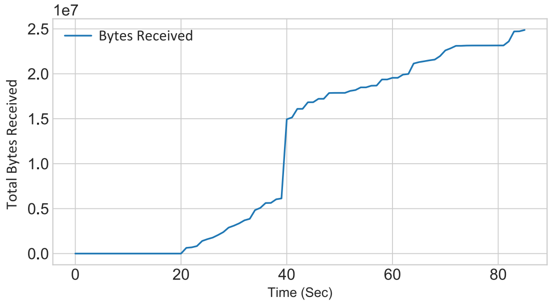


FIGURE 13. Total bytes received by one node from network.

In order to analyze the network utilization, the bytes sent and received by a network node are observed carefully, as shown in FIGURE 12 and FIGURE 13, respectively. Initially, the network is in the idle state. When transactions are performed, bytes are sent and received accordingly. The number of bytes increases gradually with the increase in the number of transactions. If the number of sent and received bytes are compared, then it is seen that the bytes received are much greater than the bytes sent in the blockchain network. The ratio between bytes sent and received depends on the number of nodes in the network. As the number of nodes increases, the difference between bytes sent and received will also increase. The reason behind this difference is that when a sender sends a message on the network, it is received by the whole network. As nodes are connected in a peer to peer manner, so, all the nodes present on the network receive the message and the sender is only one. The difference between sent and receive messages increases with the increase in the number of network nodes.

Table 2 depicts the comparison of our proposed model with existing data-sharing models for afore discussed features. For a fair comparison, all data-sharing models are blockchain-based. Table 2 depicts that authors in [10], [21] did not implement any access control method. Security and privacy of nodes are also neglected in these studies. On the other hand, authors in [3], [4], [8], [9] included these features. Moreover, incentives and pricing mechanisms are defined in [10], [21] to tackle the selfishness of DP and encourage them to share their data. The heterogeneity of data is not addressed by any of the existing models. Our proposed model includes all the aforementioned features. Besides, the computational overhead of the mining nodes is reduced by using the round-robin algorithm for consensus. This comparison shows the significance of the proposed model from existing models.

VI. CONCLUSION

In this paper, a a subscription-based data-sharing model is proposed by combining blockchain and DaaS to make data-sharing more beneficial and secure for both data owners and data consumers. The subscription-based model has more profit margin as compared to selling data at once. The data owner received revenue recurrently for a long time, which is more beneficial than one-time selling of data. Besides, the concept of DaaS is used, which has lots of benefits. DaaS provides the ability to move data easily from one place to another. It also provides ease of administration, collaboration, global accessibility and compatibility among different platforms. It also reduces the maintenance and delivery costs. Furthermore, to set a standard for monetizing data, the pricing models like FQ and PayG are discussed and a new pricing model named HPM is proposed. To add security, privacy, transparency and immutability to the proposed model, blockchain is used. Public key cryptography ensured the authentication and authorization of DSs and DPs. The digital signatures are used to verify every transaction, which made the network more secure. For simulations, a private blockchain network and an interface web-demo are implemented; both provided by MultiChain. Finally, through simulation results and discussions, the feasibility and rationality of the proposed model are demonstrated. The discussion proved that the proposed model has good scope in the future and it could be very beneficial for both data owners and consumers.

Deployment of the proposed data-sharing model can be beneficial for businesses, industries, individuals, healthcare, academia, etc. The DaaS ensures the availability and access of data in real-time using the Internet. It also solves the problem of data heterogeneity. The typical blockchain system is not suitable for data storage as it faces scalability issues due to its limited storage availability. The blockchain is used to keep track of the transactions related to data-sharing between DP and its subscribed user. To encourage the DPs to share data, a novel pricing model is used, which is a hybrid of FQ and PayG pricing models. Each user has to pay some price to DP for accessing the data. This data is accessible for only a limited period and if a user exceeds its permissible access time then it has to pay an extra price. In this way, the DP gets incentives for sharing data and this is not a one-time incentive. In this way, the access of data is restricted to authorized users only, i.e., subscribed users who have enough balance in their account. Besides, the computational overhead is reduced by implementing the round-robin consensus algorithm instead of the commonly used PoW consensus algorithm. The block size is also limited to 1 MB only, which makes the blockchain decentralized as required computational power to work as a miner is reduced. From the perspective of business, healthcare and academia, a blockchain-based efficient data-sharing model is crucial. In the healthcare department, the patients' health history and information related to their medication can help the doctors to treat the other patients with similar health problems and also they can figure out which medicine is best for curing a specific disease. From the academic point of view, the research information can be shared efficiently using this model. The researchers can share their research achievements and findings with new researchers and make their data available. Also, academic institutions can share their educational policies. From a business perspective, a business company can share its data with its suppliers, partners and employees. It can also share its market analytics and platform models with other organizations.

The practical implementation of a blockchain-based system is still a challenging task and poses several challenges. In the future, I plan to design a reputation mechanism for DP. This mechanism will help the users to select the best DP according to their requirements. The reviews about each DP would be stored from their registered users. A mechanism to detect fake reviews will also be developed to detect misbehaving users.

ACKNOWLEDGMENT

The authors would like to thank the Science and Technology Unit at Umm Al-Qura University for their continued logistics support.

REFERENCES

[1] S. Nakamoto. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Accessed: Nov. 18, 2019. [Online]. Available: http://bitcoin.org/bitcoin.pdf

[2] BlockchainHub. (2019). *Explained—Intro—Beginners Guide to Blockchain*. Accessed: Nov. 20, 2019. [Online]. Available: https://blockchainhub.net/blockchainintro/

[3] Q. Xia, E. Sifah, A. Smahi, S. Amofa, and X. Zhang, "BBDS: Blockchain-based data sharing for electronic medical records in cloud environments," *Information*, vol. 8, no. 2, p. 44, Apr. 2017.

[4] J. Liu, X. Li, L. Ye, H. Zhang, X. Du, and M. Guizani, "BPDS: A blockchain based privacy-preserving data sharing for electronic medical records," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[5] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-Health systems via consortium blockchain," *J. Med. Syst.*, vol. 42, no. 8, p. 140, Aug. 2018.

[6] X. Zhang and X. Chen, "Data security sharing and storage based on a consortium blockchain in a vehicular ad-hoc network," *IEEE Access*, vol. 7, pp. 58241–58254, 2019.

[7] J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan, S. Xie, and Y. Zhang, "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4660–4670, Jun. 2019.

[8] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "MeDShare: Trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, pp. 14757–14767, 2017.

[9] S. Wang, Y. Zhang, and Y. Zhang, "A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems," *IEEE Access*, vol. 6, pp. 38437–38450, 2018.

[10] Shrestha, Ajay Kumar, and Julita Vassileva, "Blockchain-based research data sharing framework for incentivizing the data owners," in *Proc. Int. Conf. Blockchain*. Cham, Switzerland: Springer, 2018, pp. 259–266.

[11] Wang, Lidong, "Heterogeneous data and big data analytics," *Autom. Control Inf. Sci.*, vol. 3, no. 1, pp. 8–15, 2017.

[12] S. Moore. (2019). *How To Create A Business Case For Data Quality Improvement*. Accessed: Nov. 20, 2019. [Online]. Available: https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement

[13] (2019). *Data As A Service: The Big Opportunity For Business*. Accessed: Nov. 22, 2019. [Online]. Available: https://www.forbes.com/sites/danielnewman/2017/02/07/data-as-a-service-the-big-opportunity-for-business

[14] (2019). *Multichain/Multichain-Web-Demo*. Accessed: Nov. 25, 2019. [Online]. Available: https://github.com/MultiChain/multichain-web-demo

[15] C. Harold, Q. Lin, and S. Wen, "Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3516–3526, Jun. 2019.

[16] M. Naz, F. A. Al-zahrani, R. Khalid, N. Javaid, A. M. Qamar, M. K. Afzal, and M. Shafiq, "A secure data sharing platform using blockchain and interplanetary file system," *Sustainability*, vol. 11, no. 24, p. 7054, Dec. 2019.

[17] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "MedBlock: Efficient and secure medical data sharing via blockchain," *J. Med. Syst.*, vol. 42, no. 8, p. 136, Aug. 2018.

[18] W. Gao, W. Yu, F. Liang, W. G. Hatcher, and C. Lu, "Privacy-preserving auction for big data trading using homomorphic encryption," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 776–791, Apr. 2020, doi: 10.1109/TNSE.2018.2846736.

[19] M. Rehman, N. Javaid, M. Awais, M. Imran, and N. Naseer, "Cloud based secure service providing for IoTs using blockchain," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–7.

[20] O. Samuel, N. Javaid, M. Awais, Z. Ahmed, M. Imran, and M. Guizani, "A blockchain model for fair data sharing in deregulated smart grids," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1–7, doi: 10.1109/GLOBECOM38437.2019.9013372.

[21] Y. He, H. Li, X. Cheng, Y. Liu, C. Yang, and L. Sun, "A blockchain based truthful incentive mechanism for distributed P2P applications," *IEEE Access*, vol. 6, pp. 27324–27335, 2018.

[22] Y. Wang, Z. Cai, G. Yin, Y. Gao, X. Tong, and G. Wu, "An incentive mechanism with privacy protection in mobile crowdsourcing systems," *Comput. Netw.*, vol. 102, pp. 157–171, Jun. 2016.

[23] H. Zhang, B. Liu, H. Susanto, G. Xue, and T. Sun, "Incentive mechanism for proximity-based mobile crowd service systems," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, Apr. 2016, pp. 1–9.

[24] Y. Ren, Y. Liu, S. Ji, A. K. Sangaiah, and J. Wang, "Incentive mechanism of data storage based on blockchain for wireless sensor networks," *Mobile Inf. Syst.*, vol. 2018, pp. 1–10, Aug. 2018, doi: 10.1155/2018/6874158.

[25] J. Wang, M. Li, Y. He, H. Li, K. Xiao, and C. Wang, "A blockchain based privacy-preserving incentive mechanism in crowdsensing applications," *IEEE Access*, vol. 6, pp. 17545–17556, 2018.

[26] B. Jia, T. Zhou, W. Li, Z. Liu, and J. Zhang, "A blockchain-based location privacy protection incentive mechanism in crowd sensing networks," *Sensors*, vol. 18, no. 11, p. 3894, Nov. 2018.

[27] T. A. Alghamdi, I. Ali, N. Javaid, and M. Shafiq, "Secure service provisioning scheme for lightweight IoT devices with a fair payment system and an incentive mechanism based on blockchain," *IEEE Access*, vol. 8, pp. 1048–1061, 2020.

[28] Y. Shen, B. Guo, Y. Shen, X. Duan, X. Dong, and H. Zhang, "A pricing model for big personal data," *Tsinghua Sci. Technol.*, vol. 21, no. 5, pp. 482–490, Oct. 2016.

[29] Muschalle, Alexander, Florian Stahl, Alexander Loser, and Gottfried Vossen, "Pricing approaches for data markets," in *Proc. Int. Workshop Bus. Intell. Real-Time enterprise*. Berlin, Germany: Springer, 2012, pp. 129–144.

[30] H. Almay. *Pricing in the Internet*. Accessed: Nov. 25, 2019. [Online]. Available: https://pdfs.semanticscholar.org/39c9/6f8b4e93690ff6771 38dfddc4ef81e20be08.pdf

[31] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, "Blockchain technology: Beyond bitcoin," *Appl. Innov.*, vol. 2, pp. 6–10, Jun. 2016.

[32] Z. Zheng, S. Xie, H.-N. Dai, W. Chen, X. Chen, J. Weng, and M. Imran, "An overview on smart contracts: Challenges, advances and platforms," *Future Gener. Comput. Syst.*, vol. 105, pp. 475–491, Apr. 2020.

[33] S. Voshmgir. (Jan. 2020). *Token Economy By Shermin Voshmgir*. [Online]. Available: https://shermin.net/token-economy-book/

[34] (2019). *Multichain|Open Source Blockchain Platform*. Accessed: Nov. 25, 2019. [Online]. Available: https://www.multichain.com/

[35] S. Bistarelli, I. Mercanti, P. Santancini, and F. Santini, "End-to-End voting with non-permissioned and permissioned ledgers," *J. Grid Comput.*, vol. 17, no. 1, pp. 97–118, Mar. 2019.

[36] (2019). *Gartner Hype Cycle For Data Management Positions Three Technologies In The Innovation Trigger Phase In 2018*. Accessed: Nov. 30, 2019. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2018-09-11-gartner-hype-cycle-for-data-management-positions-three-technologies-in-the-innovation-trigger-phase-in-2018

[37] G. Greenspan. (2015). *MultiChain Private Blockchain—White Paper*. Accessed: Nov. 25, 2019. [Online]. Available: http://www.multichain.com/download/MultiChain-White-Paper.pdf

FAHAD AHMAD AL-ZAHRANI received the B.Sc. degree in electrical and computer engineering from Umm Al-Qura University, Makkah, Saudi Arabia, in 1996, the M.S. degree in computer engineering from the Florida Institute of Technology, in 2000, and the Ph.D. degree in computer engineering from Colorado State University, in 2005. From 2011 to 2016, he was the IT Dean of Umm Al-Qura University and has had several other responsibilities thereafter. He is currently an Associate Professor with the Computer Engineering Department, Umm Al-Qura University. He has taught several computer network courses and supervised related research projects. His research interests include high-speed network protocols, sensor networks, optical networks, performance evaluation, the IOT, and blockchain architecture and performance analysis. He is a member of the International Society for Optical Engineering, and the Optical Society of America.