Hindawi
Journal of Nanomaterials
Volume 2022, Article ID 2219006, 14 pages
https://doi.org/10.1155/2022/2219006

Research Article

BSPPF: Blockchain-Based Security and Privacy Preventing Framework for Data Middle Platform in the Era of IR 4.0

Chuqiao Chen,[1,2] **S. B. Goyal**,[1] **and Kiran Ramaswamy**[3]

[1]*Faculty of Information Technology, City University, Petaling Jaya 46100, Malaysia*
[2]*City College of Huizhou, School of Information, Huizhou 510025, China*
[3]*Department of Electrical and Computer Engineering, Ambo University, Ambo, Ethiopia*

Correspondence should be addressed to Kiran Ramaswamy; kiran.ramaswamy@ambou.edu.et

Received 20 June 2022; Revised 21 July 2022; Accepted 1 August 2022; Published 30 August 2022

Academic Editor: Samson Jerold Samuel Chelladurai

Security and privacy issues about Big Data have emerged as one of the most pressing concerns in both academic circles and the business world in the era of IR 4.0. These worries are made worse because businesses are beginning to construct their own Big Data-based Data Middle Platform. The research team has summarized the security and privacy protection issues of Big Data-based Data Middle Platform by combining previous research and providing a table that lists some contributions that the blockchain uses for security and privacy protection in Big Data. A general Data Middle Platform architecture is also be given in the Methodology section. Based on this architecture, the team unveiled blockchain-based security and privacy protection framework on the Data Middle Platform in this study. This approach emphasizes mitigating the security and privacy issues posed by sensitive data entering the Data Middle Platform and data published by the Data Middle Platform. Following that, the researchers developed a testing system based on this framework to determine whether or not the framework was both feasible and practicable. The trials revealed that the framework is applicable, even though it requires further development and refinement.

1. Introduction

Today, privacy breaches are no longer news. In 2021, there were 44 high-impact security and privacy breaches worldwide. Victims are not only large and well-known commercial companies like Facebook and Microsoft but also include government agencies such as the California State Controller's Office. Although these incidents are isolated, it is clear that as data science continues to evolve, the risk of data security and privacy is gradually increasing, while people enjoy the convenience of Big Data. In particular, since 2018, many organizations in China have started to build their own Data Middle Platform (DMP), which is based on Big Data (BD) technologies. How to ensure data security and privacy on the Data Middle Platform is one of the key concerns in academia. In this article, the researcher will discuss the use of blockchain technology to help solve this challenge.

It is crucial to point out that privacy protection is relevant in the era of Industry 4.0. There are many reasons for this, but the most important is that Industry 4.0 is a process of using the cyber-physical system (CPS) to collect data and improve the supply and manufacturing information in production and, finally, to achieve rapid, efficient, and personalized product supply. Because of this, new information technologies such as Big Data and the Internet of Things will be combined with cloud computing and artificial intelligence. There will be a reduction in the number of information silos as a result of this convergence, but there will also be an increase in security risks. To solve security and privacy concerns, this study will leverage blockchain technology.

1.1. Blockchain. The concept of Blockchain was produced by Nakamoto [1], and an instance called bitcoin to verify the blockchain is also available to him. Blockchain can be seen

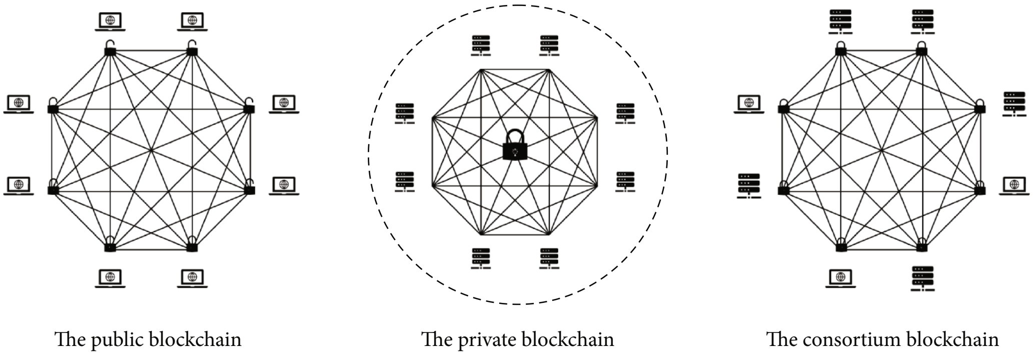


Figure 1: Three types of blockchains.

Table 1: Comparison among three type of blockchain.

Attribute	Blockchain types Public blockchain	Consortium blockchain	Private blockchain
Scope	Internet	Internet and intranet	Intranet
Permission	Public nodes	Consortium nodes	Private nodes
Speed	Slow	Medium	Fast
Decentralized	High	Partial	Partial
Cryptocurrency	Yes	No	Yes

as a decentralized database system [2, 3], which is characterized by distributed, open-source, tamper-proof, anonymity and traceability.

Figure 1 indicates that the blockchain can generally be classified into three main types, the public blockchain, the private blockchain, and the consortium blockchain. The public blockchain is open to all Internet users. Anyone can jump on the public blockchain network to read the raw data in the block, send and confirm valid transactions, and compete for the authority of the package blocks to get the award. The public blockchain is immutable because it is a completely decentralized blockchain. No one can control the public chain and tamper with data in theory unless they can control more than 50% of the total computing power of the machines in the blockchain network [4, 5]. The public blockchain uses cryptocurrency as an award to encourage people to help package data. Not only bitcoin but also include other kinds of public blockchains use this strategy. Contrary to the public blockchain, the private blockchain is completely closed to the public. It is only open for authorized nodes such as servers in a multinational organization. The dashed circles in Figure 1 evince the private blockchain which consisted of eight authorized nodes in the same group. The private blockchain is running on these authorized nodes. Because these nodes are limited and controllable, the private blockchain is faster and has more privacy and less attack than the public blockchain. The third kind of blockchain is called the consortium blockchain which hybridizes characteristics of the public blockchain and the private blockchain. Nodes in the consortium blockchain can come from the inside organization while can come from the outside. However, regardless of where it comes from, all nodes have been authorized to combine a consortium. The data in the consortium blockchain is only for the consortium. By default, it is not open to the public. To make these three types of blockchain clear, Table 1 gives a comparison among the three types of blockchain from five aspects such as the scope, the permission, and the speed, whether decentralized, whether used for cryptocurrency.

The development of blockchain has gone through four stages [6, 7]. Figure 2 shows the four-stage development path. The first stage is called Blockchain 1.0. At that time, the main usage of blockchain is cryptocurrency. For quite a long time, cryptocurrency has been the only field in which blockchain has been used. Bitcoin is a prominent example of the application of blockchain, and it is regarded as a promoter of cryptocurrency. The second stage is called Blockchain 2.0. In this stage, the smart contract is the mainstream. This is because there are two key issues in Blockchain 1.0, one is the blockchain wasting computing resources, and the other is that the blockchain lacks network scalability. The smart contract fills both of these gaps. The smart contract is a small real-time program. That makes the blockchain programmable. The best prominent example of smart contracts is Ethereum, which provides a platform on which developers can create distributed applications for blockchain networks. Blockchain 3.0 is used for decentralized applications (DApps). It is developed based on smart contract technology. Now there are several DApps like CryptoKitties already running on the Internet. The new stage of blockchain usage is called Blockchain 4.0. The Blockchain 4.0 pays close attention to the realtime decentralized applications. It faces Industry 4.0 with a perfect ecosystem. These four stages do not replace each other. They coexist. Therefore, researchers should carefully consider which technique should be chosen for practical applications.

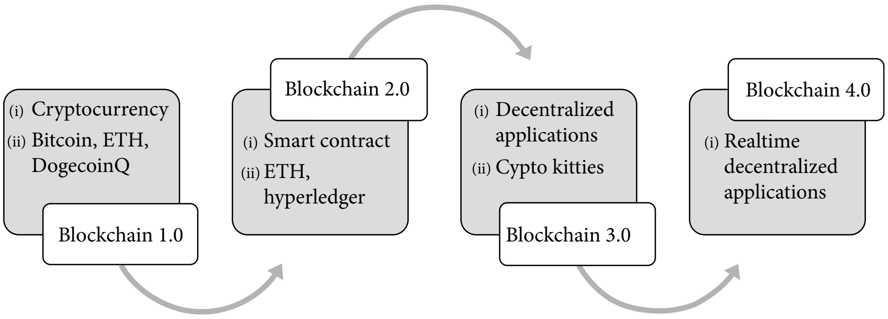


Figure 2: Four stages of development of blockchains.

1.2. Data Middle Platform. The Data Middle Platform was first used by Jack Ma, who was once the CEO of Alibaba [8, 9], and now this concept is popular in the Chinese Information Technology (IT) circle. According to Alibaba's practical experience, the Data Middle Platform acts as the interface between business services on the front end and computing services on the back end [10–12]. It requires both Big Data technology to manage and store data, as well as governance and data integration [10]. Therefore, the Data Middle Platform is not only a technical concept but also a management concept [13, 14]. In short, the Data Middle Platform is like an engine that drives data from records to information.

1.2.1. Relation between Big Data and Data Middle Platform. As mentioned in the definition of DMP, the Data Middle Platform contains both the Big Data and the Data Governance related concepts. The real reason why enterprises are keen to build their DMP instead of BD is that DMP make Big Data truly usable for business rather than mere technology.

The three papers published by Google laid the foundation for the subsequent Big Data technologies Hadoop, HBase, and MapReduce [15–17]. However, in the subsequent development of Big Data, there has been a disconnect between technology and business. A key feature is that the core data model changes relatively slowly, but at the same time business innovation and requirements for data are rapid and diverse. As an example, Taobao, China's largest online B2C platform, had a peak order count of 583,000 orders per second for the Double 11 shopping festival in 2020, and a range of business issues such as sales analysis and prediction and courier order delivery had to be met in real time while fulfilling their orders. Traditional Big Data can meet the storage of data, but cannot quickly meet the actual business needs, resulting in a gap between actual business and Big Data theory. The Data Middle Platform fills this gap where the speed of data development does not match the application development. Data middle platform relies on Big Data platforms like Apache Hadoop, and the Data Middle Platform adds data governance and data services to the Big Data platform, which can dismantle the data chimney and truly serve the business.

1.3. Layout of the Paper. The rest of the paper is organized as follows. In the related work section, the research team has summarized the security and privacy protection issues of the Big Data-based Data Middle Platform by combining previous research and providing a table that lists some contributions that the Blockchain uses for security and privacy protection in Big Data. A general Data Middle Platform architecture is also be given in the Methodology section. Based on this architecture, the team unveiled blockchain-based security and privacy protection framework on the Data Middle Platform in this study. This approach emphasizes mitigating the security and privacy issues posed by sensitive data entering the Data Middle Platform and data published by the Data Middle Platform. Based on this framework, the research team builds a test system and loads some sensitive data to test whether this framework can work or not. After that, comparison and discussion have been done. And finally, the team summarized their work and gave the direction for the next research.

2. Related Works

Related works are concerned with two major things. The first is what are the security and privacy issues in the DMP and the second is what kind of security issues have been addressed by blockchain in previous studies.

2.1. The Security and Privacy Issues in the DMP. Although DMP is based on BD, its privacy and security concerns are analogous to the Big Data security and privacy concerns. However, because of the variety of technologies involved in Big Data, security and privacy protection problems have always existed from many study viewpoints. Fang et al. [18] investigated Big Data security and privacy problems from a data lifecycle point of view. This article focuses on four aspects, such as anonymization techniques, storage encryption techniques, privacy protection in data mining, and access control techniques. Fan [19] addresses Big Data security and privacy challenges in terms of trustworthiness, authentication, etc.

Figure 3 exhibits there are three layers of Big Data security and privacy protection, the Big Data infrastructure security layer, the data safety layer, and the privacy protection layer, which are carried out in a sequential relationship from the bottom up [20]. The Big Data infrastructure security layer is the bottom layer of the whole Data Middle Platform, which not only needs to guarantee the security of its essential components but also provides security mechanisms for

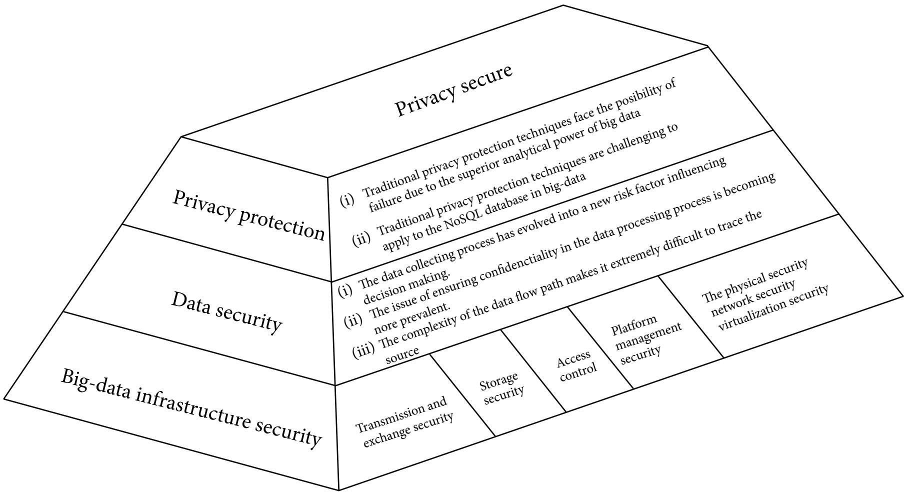

Privacy secure
Privacy protection
(i) Traditional privacy protection techniques face the posibility of failure due to the superior analytical power of big data
(ii) Traditional privacy protection techniques are challenging to apply to the NoSQL database in big-data
Data security
(i) The data collecting process has evolved into a new risk factor influencing decision making.
(ii) The issue of ensuring confidenctiality in the data processing process is becoming nore prevalent.
(iii) The complexity of the data flow path makes it extremely difficult to trace the source
Big-data infrastructure security
Transmission and exchange security
Storage security
Access control
Platform management security
The physical security network security virtualization security

Figure 3: The Big Data security and privacy protection system.

the data and applications running on it; above the Big Data infrastructure is data security. In addition to data safety, it primarily offers security protection for the data flow process in business applications; privacy security protection is the security protection of sensitive information of persons or organizations.

2.1.1. Big Data Infrastructure Security. Big Data infrastructure security is concerned with protecting the transmission, the storage, the processing, and other resources and operations of the Big Data platform. There are five aspects such as the transmission and exchange security, the storage security, the computing security, the platform management security, and the infrastructure security that has to be covered.

(i) Transmission and exchange security refer to ensuring the security and control of the process of exchanging data with external systems, which necessitates the use of mechanisms such as interface authentication to verify the legitimacy of external systems, as well as channel encryption and other methods to ensure the transmission process' confidentiality and integrity

(ii) Storage security entails implementing data backup and recovery systems as well as data access control measures to prevent unwanted data access

(iii) Access control refers to the provision by the computer components of relevant authentication and access control methods from the computer components by the computer components by the computer components to ensure that only authorized users or applications may start data processing requests

(iv) Platform management security includes the security configuration of platform components, resource security scheduling, patch management, and security audit

(v) In addition, the physical security, the network security, and the virtualization security of the platform software and hardware infrastructure are the foundation for the Big Data platform's secure functioning

Composed of five key points mentioned above, threats to Big Data infrastructure security can be summarized as the following three issues:

(i) Under the open-source Hadoop approach, the Big Data platform lacks an overarching security plan, and its security mechanism has limitations

(ii) The Big Data platform serves a large number of users and different scenarios, making standard security methods impossible to satisfy the demands

(iii) The large-scale distributed storage and computing model of the data platform has caused the difficulty of security configuration to increase exponentially

2.1.2. Data Security. Data security refers to the platform's security services that enable data flow security, such as data categorization and classification, metadata management, quality management, data encryption, data isolation, leak prevention, traceability, and data destruction, among others.

Big Data has caused the data life cycle to steadily grow from a classic single-chain form to a complex multichain form, boosting sharing and trading linkages and diversifying data application scenarios and participation roles. Data security is the main need in the complex application environment to ensure that sensitive data such as national significant data, confidential corporate data, and user personal privacy data are not disclosed.

As a pool of data resources serves numerous data suppliers and data consumers at the same time, the new requirement for data security in the Big Data environment is to enhance data isolation and access control and to actualize data "accessible but not visible." The results of mining and analyzing large amounts of data using Big Data technologies may contain sensitive information on national security, economic operations, social governance, etc. It is necessary to strengthen security management for sharing and disclosure of analysis results.

Big Data, due to its vast volume and variety, introduces new dangers to data security in the Big Data context that are distinct from traditional data security. Specifically, there are three points as follows:

(i) The data collection process has evolved into a new risk factor influencing decision-making

(ii) The issue of ensuring confidentiality in the data processing process is becoming more prevalent

(iii) The complexity of the data flow path makes it extremely difficult to trace the source

2.1.3. Privacy Protection. The privacy preservation mentioned in this paper refers to the use of technologies such as de-identification, anonymization, and cryptographic computation to safeguard personal or organizational data from disclosing privacy or other information that individuals or organizations do not want to be known by the outside world during the process of processing and flowing on the platform. Privacy preventing is a more advanced security need that is based on data security protection to preserve the privacy of persons or organizations. However, we recognize that privacy protection in the age of Big Data is more than just preserving the privacy rights of individuals or organizations; it is also about protecting the self-determination rights of data subjects regarding information during the gathering and use of information. Privacy protection has become a systematic project covering product design, business operation, security protection, etc., and is not a mere technical issue. Because this study focuses on the use of blockchain technology to secure private data in the Data Middle Platform, while discussing the protection of data subjects' privacy rights, we choose to begin our research with privacy protection technology, which has a clearer research path.

(i) Traditional privacy protection techniques face the possibility of failure due to the superior analytical power of Big Data

Enterprises in the Big Data environment can recover anonymized data through correlation analysis and deep mining of numerous datasets from multiple sources, allowing them to identify specific persons or obtain important personal information. Traditional privacy-prevention techniques and parameters are selected by data controllers for individual datasets in isolation to protect personal data. However, this practice, particularly the use of de-identification, masking, and other techniques, cannot deal with privacy leakage problems caused by multisource data analysis and mining in the aforementioned Big Data scenario.

(ii) Traditional privacy protection techniques are challenging to apply to the NoSQL database in Big Data

Data in Big-Data is dynamic, semi-structured, and unstructured. It is typically used to protect the privacy of unstructured data, which represents more than 80% of the total data. Nonrelational database (NoSQL) storage technology is typically used to gather, manage, and analyze large data for those unstructured data. There is no rigorous access control mechanism or reasonably complete privacy-protection tools, and existing privacy protection technologies, such as de-identification and anonymization, are primarily suited to relational databases. Existing privacy safeguards, like de-identification and anonymization, are most relevant to relational databases.

2.2. Contributions of Blockchain to Big-Data Security and Privacy. The distributed, open-source, tamper-proof, anonymous, and traceable nature of blockchain makes it inherently good for security and privacy. As a result, blockchain-based security and privacy protection research has proliferated. In this paper, we review Google Scholar, IEEE Access, ACM, and other databases to illustrate the literature on blockchain in Big Data storage, blockchain in Big Data auditing, and blockchain in Big Data collection, to understand the impact of blockchain on Big Data security and privacy protection from a macro perspective. To facilitate reading, we have mapped out the contribution of blockchain to Big Data security and privacy in a form in Table 2.

3. Methodology

An introduction to the general architecture of the Data Middle Platform is necessary. This is because, in concrete practice, the different organizations have a different understanding of what the Data Middle Platform is. A Chinese blog summarizes several kinds of Data Middle Platforms that already used in industry. And given these different Data Middle Platform models, a typical Data Middle Platform shown in Figure 4 should contain the following three layers, that is, the foundation layer, the processing layer, and the publishing layer.

3.1. The Typical Data Middle Platform. The foundation layer is the bottom layer. It is combined by a series of infrastructure components such as the high-speed Internet, the 5G, the Virtualization Technology (VT), and the Database Technology. The functions of this layer have two. One is for supporting the upper layer, and the other one is receiving the internal and external data which is produced from the other system or IoT devices.

The processing layer is between the foundation layer and the publishing layer. A set of data governance theories is used in this layer to help manage the data. Normally, data processing can be divided into four steps. The first step is to collect different sources of data into the data lake. The data will then be extracted for different purposes. The extracted data are computed using machine learning (ML) or artificial intelligence (AI), which is the third step. And the last step, the data has to aggregate and then become a series of data warehouses (DW) based on the business requests. Except for these four steps, the data operating system and the data security system are also necessary due to the escorts of this process.

Table 2: The contribution of blockchain in Big Data security and privacy.

Ref no.	Information	Result	Involve Big Data area					
			Storage	Collection	Transaction	Audit	Access control	Encryption
[21]	Xia et al., 2017	MeDShare uses smart contracts and an access control mechanism to examine the user who accesses data from a data custodian system for potentially hazardous behavior	√		√	√	√	
[22]	Chowdhury et al., 2018	Implemented a prototype framework that provides privacy, integrity, and fine-grained access control on shared data	√	√	√	√	√	√
[23]	Mora et al., 2018	There is a use case to discuss using Blockchain to conciliate security versus privacy. Security and privacy issues are mentioned in this article	√	√	√	√	√	
[24]	Yang et al., 2020	A Big Data exchange and transaction architecture built on blockchain's decentralization and transparency. Propose a data-tamperproof approach that includes a cryptographic algorithm to prevent transaction data from being tampered with during user storage. Assure transaction security and data dependability when trading on the blockchain				√		√
[25]	Tan et al., 2020	BacCPSS permission, allotment revocation, admission control, and analysis processes are built and maintained in the blockchain				√	√	√
[26]	Li et al., 2020	The author makes use of blockchain technology to provide a unique public auditing method for ensuring data integrity in cloud storage. Data owners store the lightweight verification tags on the blockchain to decrease the overhead of computation and communication for integrity verification	√			√		
[27]	Stodt and Reich, 2020	Using blockchain technology to prevent the disclosure of sensitive data while simultaneously providing auditable proof of data exchange. It defines providing ownership of the data to (possibly untrustworthy) other parties as a kind of privacy breach. The decentralized data storage method used in this article ensures data secrecy in Blockchain smart contracts using peer-to-peer communication and data processing			√	√		
[28]	Zhang et al., 2020	A Big Data security protection scheme proposes for Hadoop. Features are metadata decentralization and data that is difficult to tamper with. Smart contract reasonably allocates user roles based on the assessment of user tag and risk value. Building a risk value tracking chain to monitor user activity in real time	√			√	√	
[29]	Yu et al., 2021	BCBLPM focuses on the multichain environment and uses smart contracts to isolate access domains to achieve access control				√	√	

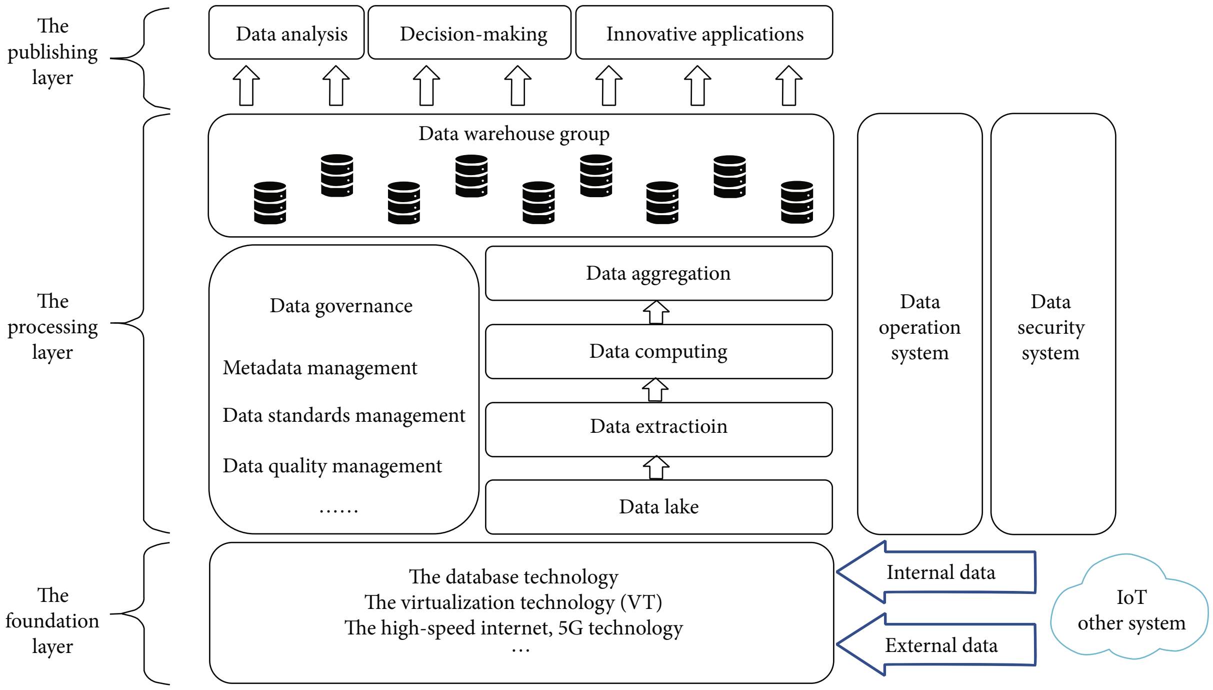


Figure 4: The general architecture of the Data Middle Platform.

The top layer is the publishing layer. After processing, the data will be regrouped and clear for use in various areas, such as using for data analysis, decision-making, and support innovative applications. For fixing the flexible and diverse business, the platform will use API as the union interface. Users call different interfaces; the platform will call the corresponding data warehouse for feedback on the right data.

3.2. The BSPPF Structure. Consider the core purpose of Data Middle Platform is to serve the business. However, in a practical situation, there are many unavoidable problems in the circulation of data. For example, communication companies, Internet giants, and many government agencies have a lot of first-hand data. These data are not only of high value but also of good quality and very meaningful for analysis and mining. However, most organizations are not willing to open up their data for Big Data analysis and mining due to fear of the adverse impact on the organization. This results in data that are often left unused and become garbage. On further analysis, the reason for this situation is mainly due to trust. Rashly opening data in an untrustworthy environment can easily cause data security and privacy protection as problems. Therefore, we try to build blockchain-based security and privacy preventing framework (BSPPF) in DMP to guarantee that sensitive data is secure. The framework focuses on two issues; first is how data, especially privacy, can be securely accessed in the DMP and second is how to ensure that data published after a series of aggregation and processing will not be illegally accessed and used.

The following is an overview of the DMP structure with BSPPF. As shown in Figure 5, the Data Middle Platform has a wide range of data sources. Cell phones, data manually uploaded by users, IoT devices, and other systems can be the data sources of the DMP. Of course, there are various types of data, such as logs, files, audio and video files, and ID information, fingerprint data. This requires data differentiation. If the data belong to the unsensitive data, then it can be used directly using the existing security means, and then it can enter the Data Middle Platform for data flow and processing after checking that it is correct. If these data belong to the sensitive data or the data that need to be guaranteed that they have not been tampered with, then they are passed on to the Data Middle Platform through the blockchain approach.

When the data is processed in DMP, the next step is regrouped and will become a series of Data Warehouses to serve the business. Typically, privacy attacks based on data release are carried out at this stage. Therefore, to solve this problem, we design a smart contract to control user access, thus solving the access control problem. The access record will be recorded in the blockchain, which is also of great help for future auditing. Specifically, it consists of five steps. The first step is the user initiates private data requests. Then, the access control mechanism will verify whether the user is authorized or unauthorized. If the user is the right person, the Data Middle Platform will return the privacy data to the smart contract. The smart contract will record this request and upload to the Blockchain to store. The advantage of this is that the privacy request information is effectively preserved by taking advantage of the difficult nature of the blockchain to modify, and in the event of a future privacy breach, the identity of the person who compromised the data can be easily traced by retrieving the records in the blockchain.

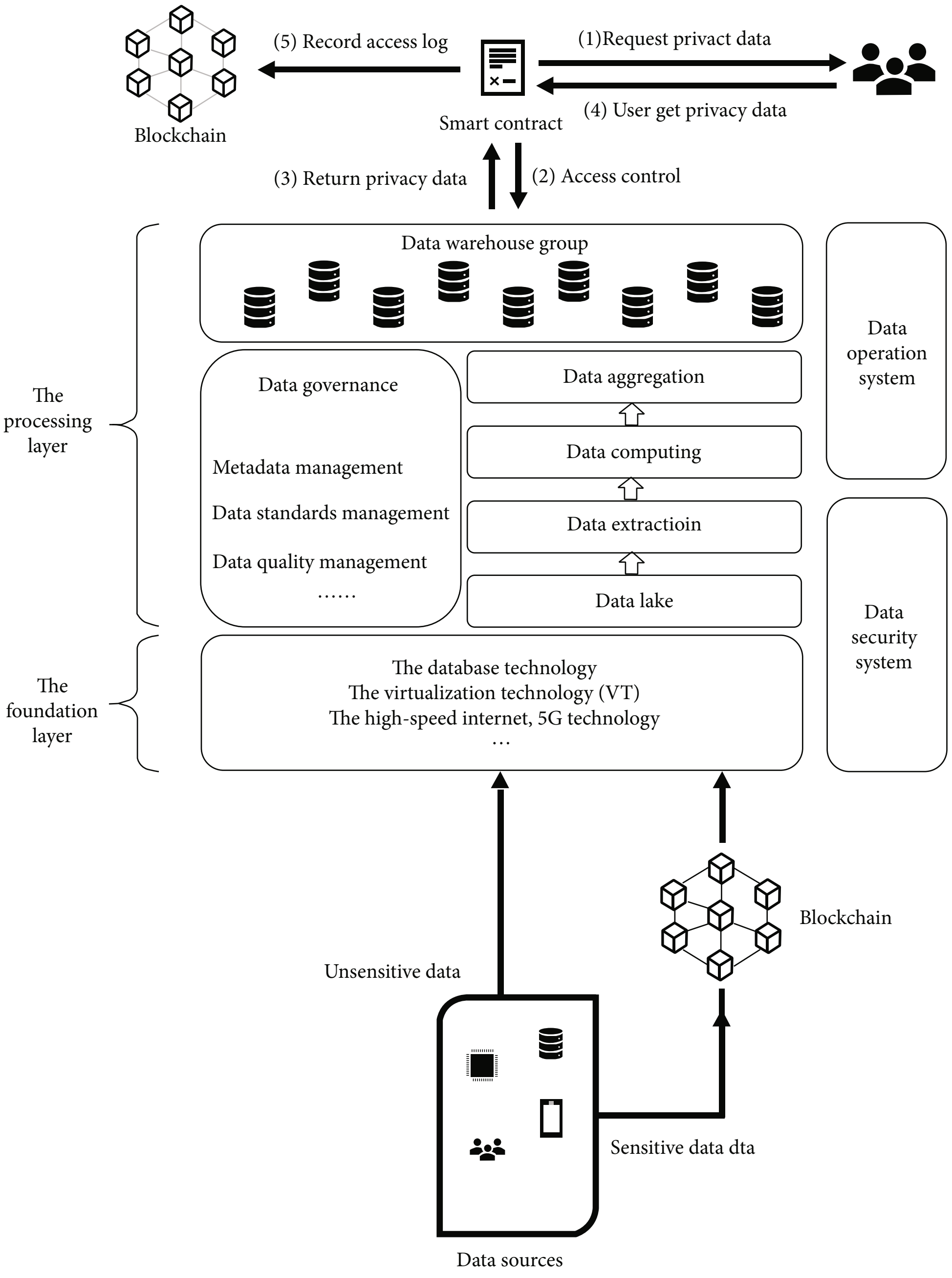


Figure 5: The BSPPF structure.

3.3. Data into Blockchain. Figures 6 and 7 depict the processes how data goes from the endpoint to the Data Middle Platform. The endpoint data is differentiated before entering the DMP, and blockchain has been used as a transmission medium to deliver privacy data into the DMP. Before the actual transfer of data, the key-exchange process must be done in Figure 6. The endpoint represents those data sources that need to send data to the DMP. It will create three data, the encryption key for the endpoint (EKeyEP), the decryption key for the endpoint (DKeyEP), and the endpoint address (EA). The DKeyEP and the EA will transfer to the DMP by a secure tunnel to create a key-address dictionary as the data validation tool.

Figure 7 describes how the data flow from the endpoint to the DMP. The original data will be separated into the privacy section and the nonprivacy section. The privacy section will be encrypted using EKeyEP. Then, the cipher text will be packaged with the nonprivacy section and the EA as a message transfer to the distributed hash table (DHT) Network. And the message will be aggregated into the block and then added to the blockchain. The DMP server will read the data in the blockchain and determine whether to decrypt the data based on the key-address dictionary.

3.4. Data Published from the DMP. Figure 8 gives the process of the privacy data publishing from the DMP. A smart

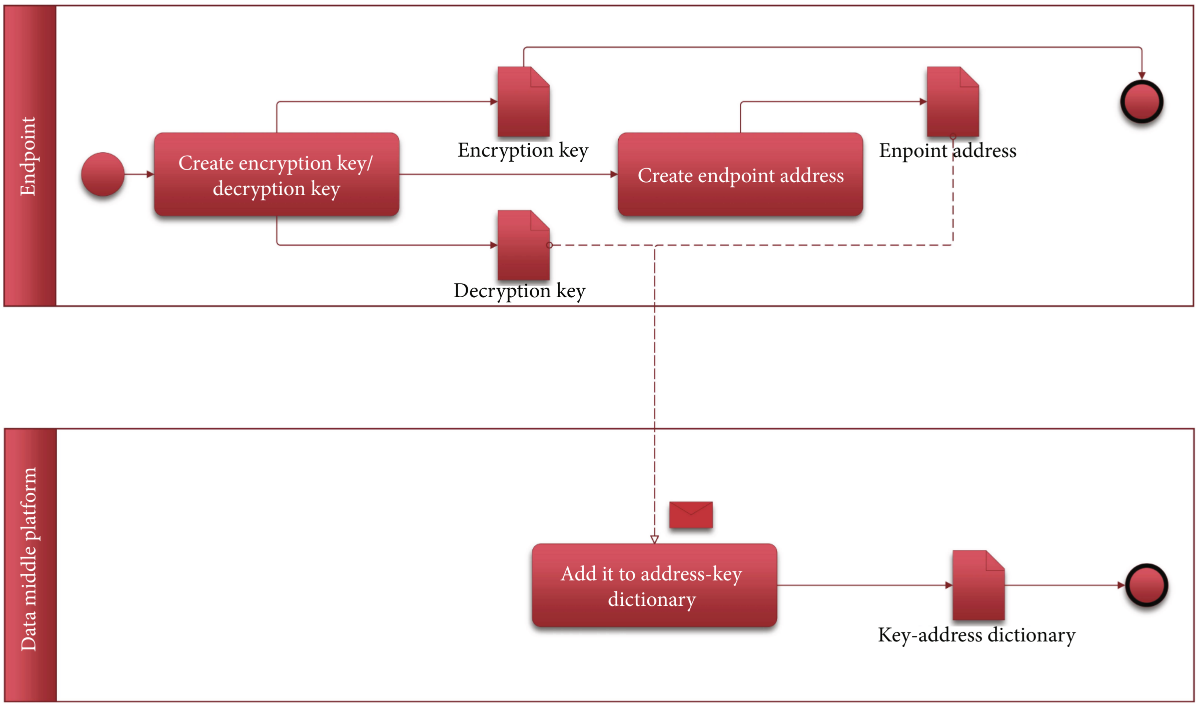


Figure 6: The key-exchange process.

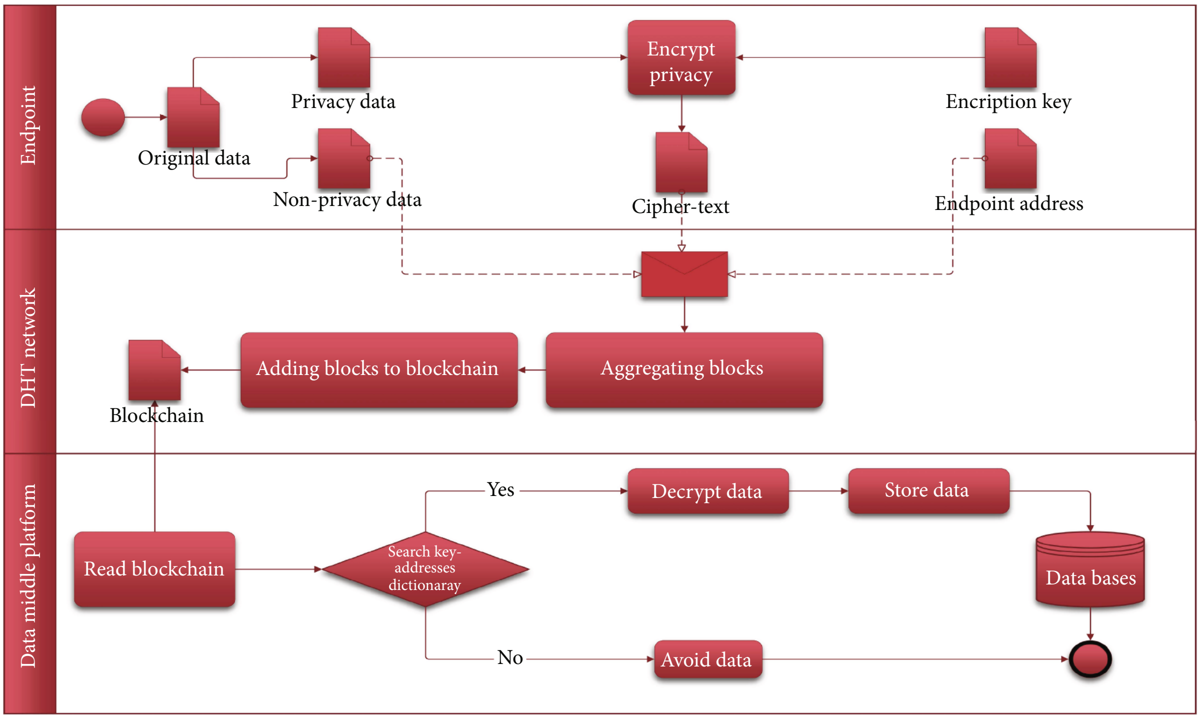


Figure 7: Data flows from the endpoint into The DMP.

contract will control the permissions to publish the data. Whether the data is published or not, the smart contract will be recorded on the blockchain for auditing purposes. Specifically, the process of acquiring data by users will be divided into three stages. Specifically, the process of acquiring data by users will be divided into three stages. In the first stage, the user will generate a unique ID and register that ID into the blockchain network. Specifically, the process of acquiring

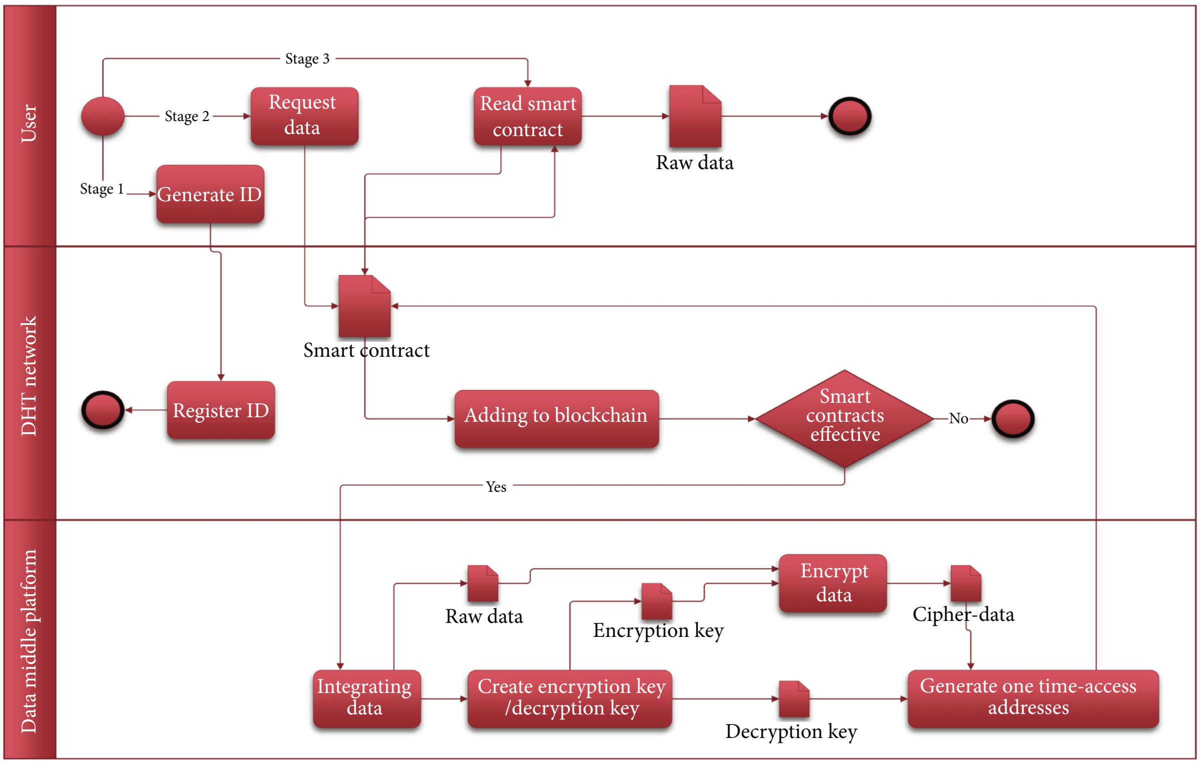


Figure 8: The process of data publishing from The DMP.

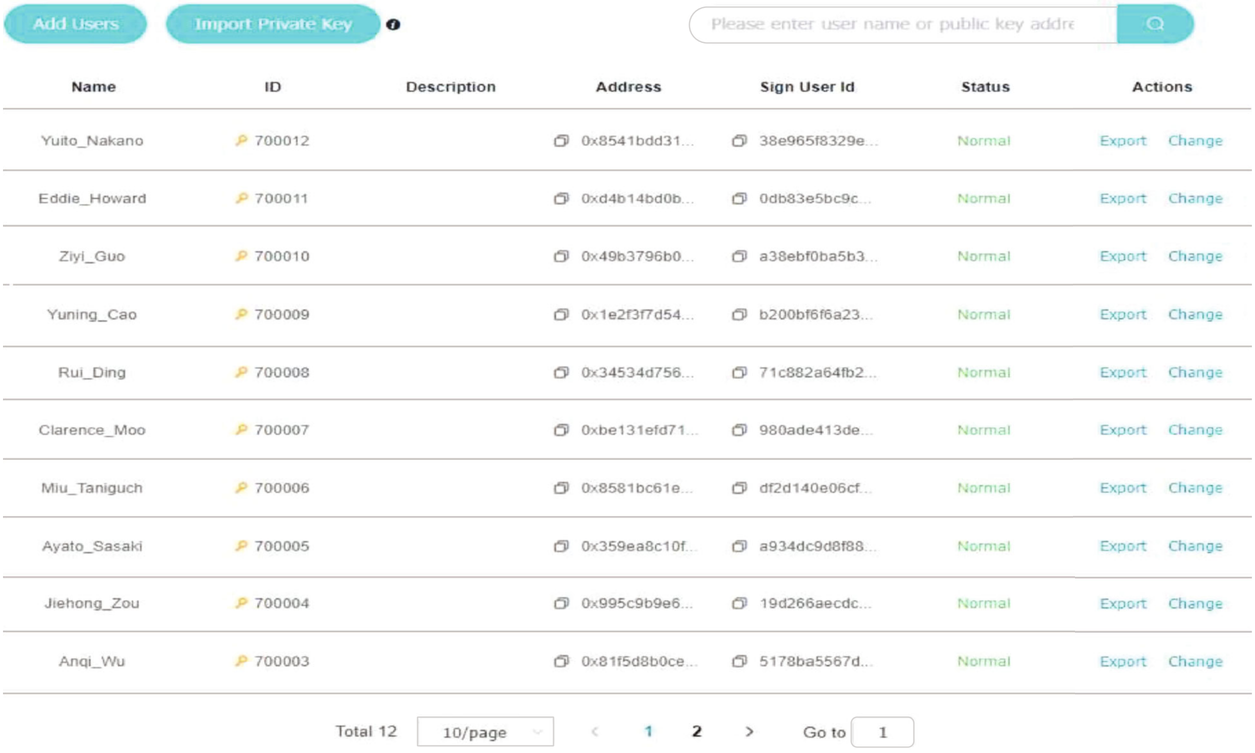
data by the user will be divided into three stages. In the first stage, the user will generate a unique ID and register that ID into the blockchain network. When the user makes a data request, the DHT network will initiate a smart contract. If the smart contract is approved, the platform will proceed to prepare the raw data and then, using the generated encryption key for DMP (EKeyDMP), encrypt them to cipher data. The encrypted data which are the cipher data will be paired with the decryption key for DMP (DKeyDMP) to generate a one-time-access address that will be written back to the smart contract. When the user reads the smart contract to get the access address, it can read the original data.

3.5. Implementation of the Test System. The researchers decided to use a workstation as the underlying physical server for the deployment of the test system. The workstation they selected had a total of 24 cores by two E5-2420 CPUs and 32 gigabytes of RAM. With the use of virtualization technology, a FISCO-BCOS blockchain with four nodes and the Webase management platform were developed to fulfil the solution's prerequisite for a blockchain. Due to the fact that a data middle is a large and complex system in a real industrial environment, the researchers decided to use two distinct kinds of databases, a K-V type of Redis and a traditional SQL type of MariaDB, as the data sources to host SQL type data and K-V type data in the data middle. This decision was made to make the simulation as easy as possible. Program languages such as python and java will be used to write the connection code, and an improved version of solidity that is adapted to FISCO-BCOS will be used to develop blockchain contracts.

4. Results

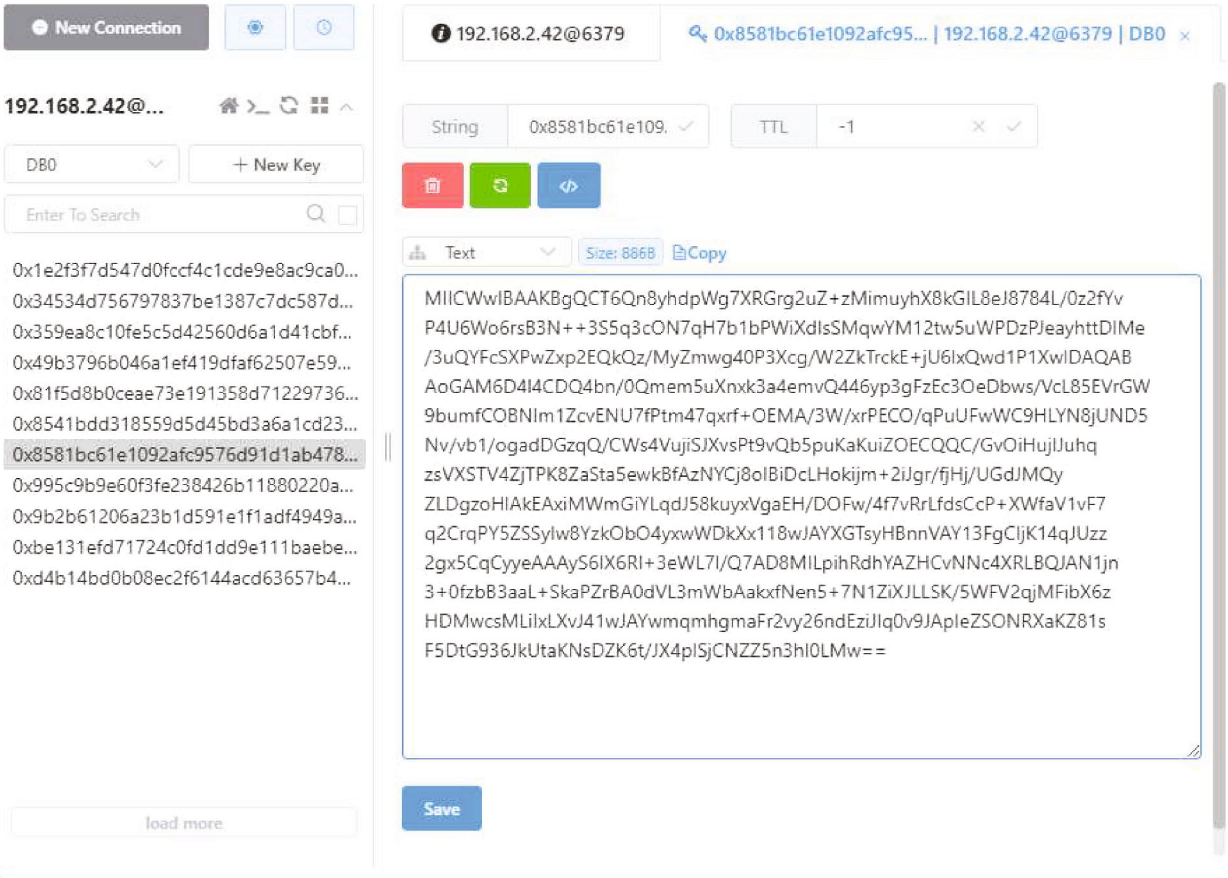
The results of the test system indicated that the framework was reasonable and could achieve the desired goals. During the tests, the research team used the default console command in FISCO-BCOS blockchain platform to recreate the procedure shown in Figure 6 in order to produce a total of 11 end users. Figure 9 shows these eleven users' information on the blockchain platform. Additionally, the RSA-1024 technique was used in order to produce the encryption and decryption keys for the users. The addresses of the user in the blockchain platform and the decryption keys were merged into a set of key-value pairs, and then those pairs were saved in a database of the K-V type that was chosen. Figure 10 is the screenshot from the K-V Database client.

The research team then encrypts the simulated privacy data in accordance with the processes shown in Figure 7, using the encryption key that is held by the user. Figure 11 indicated that this is accomplished by invoking a contract on the blockchain platform, which ultimately results in the completion of the address-cipher text upload. The Data Middle Platform will extract and decode the address using the decryption key that is already stored in the K-V database. The address will then be entered straight into the database. As evidence that it is possible, the cipher text as well as the encryption key and the decryption key are both shown on the screen alongside the display of the cipher text for purposes of presentation. This stage, of course, is done out in an automated fashion inside the Data Middle Platform in the actual industry, and it should not seem to be observable at any point in the process.



Name	ID	Description	Address	Sign User Id	Status	Actions
Yuito_Nakano	700012		0x8541bdd31...	38e965f8329e...	Normal	Export Change
Eddie_Howard	700011		0xd4b14bd0b...	0db83e5bc9c...	Normal	Export Change
Ziyi_Guo	700010		0x49b3796b0...	a38ebf0ba5b3...	Normal	Export Change
Yuning_Cao	700009		0x1e2f3f7d54...	b200bf6f6a23...	Normal	Export Change
Rui_Ding	700008		0x34534d756...	71c882a64fb2...	Normal	Export Change
Clarence_Moo	700007		0xbe131efd71...	980ade413de...	Normal	Export Change
Miu_Taniguch	700006		0x8581bc61e...	df2d140e06cf...	Normal	Export Change
Ayato_Sasaki	700005		0x359ea8c10f...	a934dc9d8f88...	Normal	Export Change
Jiehong_Zou	700004		0x995c9b9e6...	19d266aecdc...	Normal	Export Change
Anqi_Wu	700003		0x81f5d8b0ce...	5178ba5567d...	Normal	Export Change

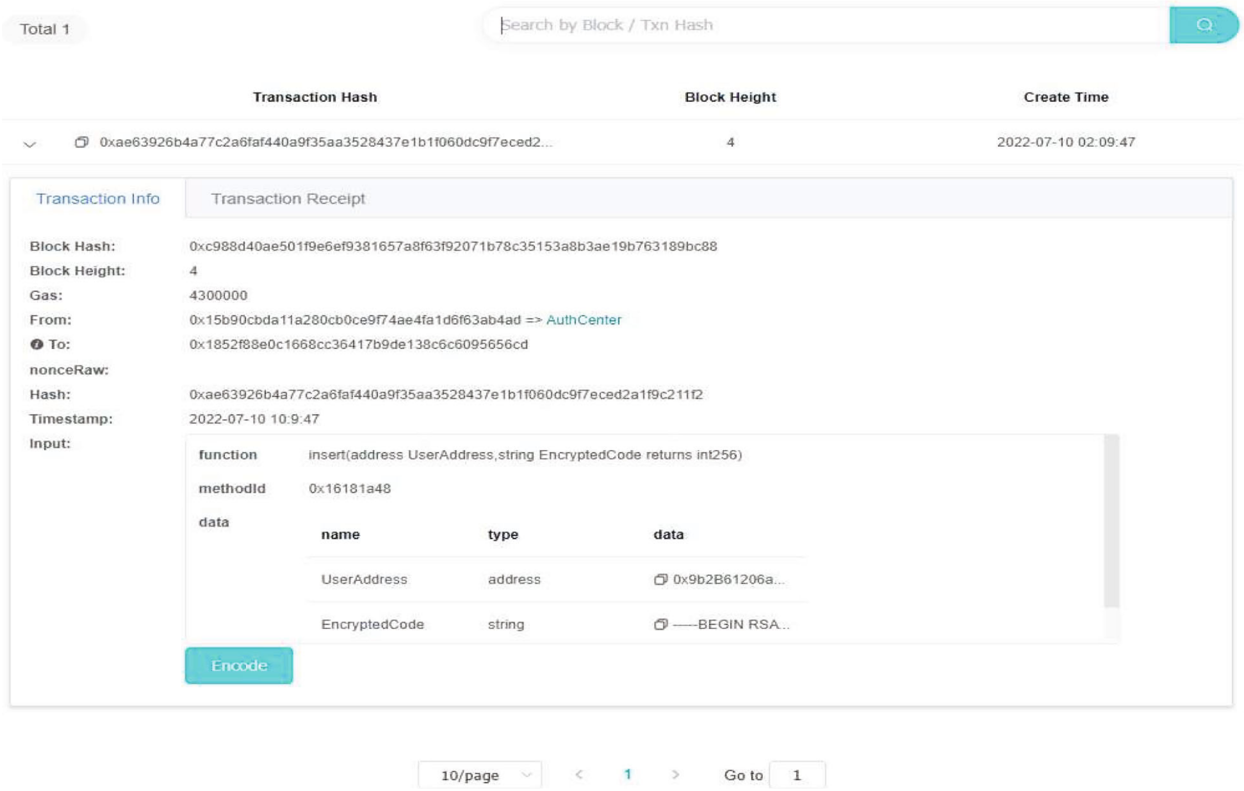
Figure 9: The end users in blockchain platform.



0x1e2f3f7d547d0fccf4c1cde9e8ac9ca0...
0x34534d756797837be1387c7dc587d...
0x359ea8c10fe5c5d42560d6a1d41cbf...
0x49b3796b046a1ef419dfaf62507e59...
0x81f5d8b0ceae73e191358d71229736...
0x8541bdd318559d5d45bd3a6a1cd23...
0x8581bc61e1092afc9576d91d1ab478...
0x995c9b9e60f3fe238426b11880220a...
0x9b2b61206a23b1d591e1f1adf4949a...
0xbe131efd71724c0fd1dd9e111baebe...
0xd4b14bd0b08ec2f6144acd63657b4...

```
MIICWwIBAAKBgQCT6Qn8yhdpWg7XRGrg2uZ+zMimuyhX8kGIL8eJ8784L/0z2fYv
P4U6Wo6rsB3N++3S5q3cON7qH7b1bPWiXdIsSMqwYM12tw5uWPDzPJeayhttDIMe
/3uQYFcSXPwZxp2EQkQz/MyZmwg40P3Xcg/W2ZkTrckE+jU6IxQwd1P1XwIDAQAB
AoGAM6D4I4CDQ4bn/0Qmem5uXnxk3a4emvQ446yp3gFzEc3OeDbws/VcL85EVrGW
9bumfCOBNIm1ZcvENU7fPtm47qxrf+OEMA/3W/xrPECO/qPuUFwWC9HLYN8jUND5
Nv/vb1/ogadDGzqQ/CWs4VujiSJXvsPt9vQb5puKaKuiZOECQQC/GvOiHujIJuhq
zsVXSTV4ZjTPK8ZaSta5ewkBfAzNYCj8oIBiDcLHokijm+2iJgr/fjHj/UGdJMQy
ZLDgzoHIAkEAxiMWmGiYLqdJ58kuyxVgaEH/DOFw/4f7vRrLfdsCcP+XWfaV1vF7
q2CrqPY5ZSSylw8YzkObO4yxwWDkXx118wJAYXGTsyHBnnVAY13FgCIjK14qJUzz
2gx5CqCyyeAAAyS6IX6RI+3eWL7I/Q7AD8MILpihRdhYAZHCvNNc4XRLBQJAN1jn
3+0fzbB3aaL+SkaPZrBA0dVL3mWbAakxfNen5+7N1ZiXJLLSK/5WFV2qjMFibX6z
HDMwcsMLiIxLXvJ41wJAYwmqmhgmaFr2vy26ndEziJIq0v9JApIeZSONRXaKZ81s
F5DtG936JkUtaKNsDZK6t/JX4pISjCNZZ5n3hI0LMw==
```

Figure 10: The address-decrypted code table.



Total 1

Search by Block / Txn Hash

Transaction Hash	Block Height	Create Time
0xae63926b4a77c2a6faf440a9f35aa3528437e1b1f060dc9f7eced2...	4	2022-07-10 02:09:47

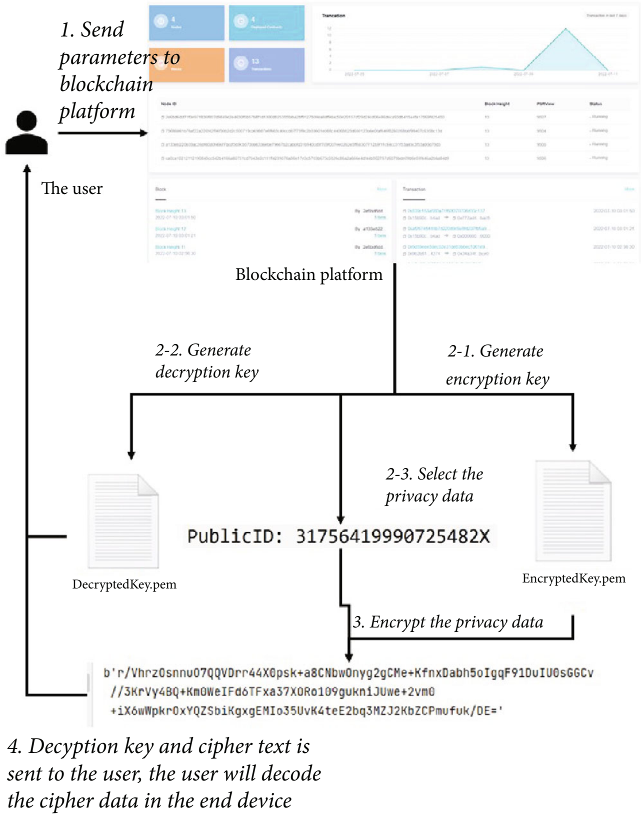
Transaction Info | Transaction Receipt

Block Hash:	0xc988d40ae501f9e6ef9381657a8f63f92071b78c35153a8b3ae19b763189bc88
Block Height:	4
Gas:	4300000
From:	0x15b90cbda11a280cb0ce9f74ae4fa1d6f63ab4ad => AuthCenter
To:	0x1852f88e0c1668cc36417b9de138c6c6095656cd
nonceRaw:	
Hash:	0xae63926b4a77c2a6faf440a9f35aa3528437e1b1f060dc9f7eced2a1f9c211f2
Timestamp:	2022-07-10 10:9:47
Input:	

function	insert(address UserAddress,string EncryptedCode returns int256)
methodId	0x16181a48
data	

name	type	data
UserAddress	address	0x9b2B61206a...
EncryptedCode	string	-----BEGIN RSA...

Figure 11: The cipher text adding to the blockchain.



1. Send parameters to blockchain platform
The user
Blockchain platform
2-2. Generate decryption key
2-1. Generate encryption key
2-3. Select the privacy data
PublicID: 317564199907254 82X
DecryptedKey.pem
EncryptedKey.pem
3. Encrypt the privacy data
4. Decyption key and cipher text is sent to the user, the user will decode the cipher data in the end device

```
b'r/Vhrz0snnu07QQVDrr44X0psk+a8CNbw0nyg2gCMe+KfnxDabh5oIgqF91DuIU0sGGCv
 //3KrVy4BQ+Km0WeIFd6TFxa37X0Ro109gukniJUwe+2vm0
 +iX6wWpkr0xYQZSbiKgxgEMIo35UvK4teE2bq3MZJ2KbZCPmufuk/DE='
```

Figure 12: The processes of sending privacy.

Figure 12 describes how to call the privacy data from the system, which the procedure has shown in Figure 8. Once the user has finished registering, he can inquire about privacy by filling in the relevant parameters on the blockchain platform. The platform will transfer these parameters to the DMP system to generate the SQL query statements or No-SQL query statements. This statement will be invoked by AuthCenter, the central user, to query the privacy in the data warehouse of the Data Middle Platform. Once the privacy is queried, it will be encrypted by using a one-time encryption key. In the test system, the user requires to provide the parameters, and it creates a SQL query statement. The reason for this is that the back-end database is a SQL type database. However, in actuality, the data warehouse does not necessarily have to be of the SQL type; hence, it is necessary to modify it to the particular circumstances. The end user will get an encrypted version of the decryption key that has been Base64-encrypted. Using the decryption key in conjunction with the cipher text allows the end user to get the data that they need.

5. Comparison and Discussion

In the related work section, we summarize the three levels of security and privacy issues faced by the DMP, and the BSPPF effectively addresses exactly the data security dimension. It is an aspect that has received less attention in the previous researches [8, 9, 30, 31]. Looking at the entire Data Middle Platform structure, if a malicious user chooses to attack at the stage when data is transferred into the DMP or the DMP releases data for business use, it is very vulnerable to a breach of sensitive data. And gaps in management boundaries make it difficult to audit when the breach occurs. The tamper-evident and traceability nature of the blockchain provides a good channel for data transmission. The flow of data is transparent and visible, which is very helpful for auditing [32].

The use of smart contracts in the BSPPF is also very subtle when it comes to data publishing. Typically, a smart contract includes a number of defined states, transformation rules, trigger situations, and response actions that are linked to the blockchain data in the form of computer code once signed by all parties. After propagation through a peer-to-peer computer network and validation by the nodes, it will be recorded in a distributed ledger at each node, where the blockchain can monitor the status of the entire smart contract in real time and activate and execute the contract after validating external data sources to confirm that specific triggering conditions are met. The definition of the contract and the description of the relational data are critical to the overall system, and the design must be clear so that the representation of the contract may satisfy the expression of the open environment's read-and-write rules. For example, if the private data of students in a class can only be accessed by their tutor, the smart contract condition could be set to allow only that tutor to execute it will no longer be allowed. If someone else invokes the contract to request access, the contract will no longer be legal, and therefore will not be executed. The smart contract rules work in conjunction with the blockchain to enable access control and thus the protection of private data.

While smart contracts solve the problem of access control for the BSPPF, an important feature of the blockchain, namely, information sharing and transparency, makes all transaction records visible to everyone in the DHT network. While this ensures the stability of the blockchain, it also creates the challenge of protecting privacy. For this reason, we have applied cryptography to try to solve this problem by encrypting the data coming into the DMP and the data published by the DMP, respectively. In the BSPPF, the keys are not the traditional public key and private key, because neither is allowed to be published. Each side that transfers the data only has one key for encryption or decryption.

However, this approach must assume that the channel over which the key is transmitted must be trusted. If the key is compromised, it is still possible to cause a phishing attack or other forgery attack. Therefore, whether there is a better way to implement the protection of private information in the blockchain will be a key question for the next phase of research.

6. Conclusion and Future Work

In this study, we review security and privacy concerns in DMP and then research the literature to see how blockchain can help with security and privacy protection. Then, we discuss a blockchain-based privacy and security prevention framework (BSPPF) in DMP. The most crucial features in the BSPPF pay much more attention to security and privacy, especially the hazards that a DMP confronts when data enters the DMP and when the DMP releases the data in the framework. Smart contract technology, along with encryption technology, assures the data's secrecy, validity, and availability. However, the framework is currently in the prototype stage, and the application of relevant cryptographic techniques needs to be further explored and refined.

In the next phase, the team's efforts will be directed toward testing and refining the model in more realistic settings. To be more specific, the model will be used in an Industry 4.0 scenario and its applicability in the Industry 4.0 environment explored. According to the results of the preliminary testing, there is a good chance that the model will be successfully implemented. However, given the diversity and richness of the actual industrial data, many minor issues will need to be resolved before they can be applied to the real industrial environment effectively.

Data Availability

Data will be available with the corresponding author on a request basis.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," *Bitcoin. Bitcoin.*, vol. 2018, 2018https://bitcoin.org/bitcoin.pdf.

[2] E. Bandara, W. K. Ng, K. De Zoysa et al., "Mystiko—Blockchain meets big data," in *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, 2018.

[3] Z. Zheng, S. Xie, H. N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: a survey," *International Journal of Web and Grid Services*, vol. 14, no. 4, p. 352, 2018.

[4] R. Shrestha and S. Y. Nam, "Regional blockchain for vehicular networks to prevent 51% attacks," *IEEE Access*, vol. 7, pp. 95033–95045, 2019.

[5] M. Saad, J. Spaulding, L. Njilla, C. A. Kamhoua, D. H. Nyang, and A. Mohaisen, "Overview of Attack Surfaces in Blockchain," in *Blockchain for Distributed Systems Security*, p. 352, Wiley-IEEE Computer Society Press, 2019.

[6] U. Bodkhe, S. Tanwar, K. Parekh et al., "Blockchain for industry 4.0: a comprehensive review," *IEEE Access*, vol. 8, pp. 79764–79800, 2020.

[7] L.-Y. Huang, J.-F. Cai, T.-C. Lee, and M.-H. Weng, "A study on the development trends of the energy system with blockchain technology using patent analysis," *Sustainability*, vol. 12, no. 5, p. 2005, 2020.

[8] C. Chuqiao and S. B. Goyal, "Data security and privacy-preserving framework using machine learning and blockchain in big-data to data middle platform in the era of IR 4.0," in *Recent Trends in Intensive Computing, Volume 39: Recent Trends in Intensive Computing:145–52*, IOS Press, 2021.

[9] C. Zhang and L. Hou, "Data middle platform construction: the strategy and practice of National Bureau of Statistics of China," *Statistical Journal of the IAOS*, vol. 36, no. 4, pp. 979–986, 2020.

[10] Z. H. Deng, *Big Data Innovation: The Way of Data Middle Platform on Alibaba Cloud*, Publishing House of Electronics Industry, Beijing, China P.R.C, 2018.

[11] The Data Technology and Product Department, *The Road to Big Data: Alibaba's Big Data Practice*, Publishing House of Electronics Industry, Beijing, China P.R.C, 2017.

[12] M. Su, X. Jia, D. Xiaomeng, and T. Gao, "Research on the recent development and future trends of data mid-end technology," 2019, https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2020&filename=KYXH201905012.

[13] H. Zhong, *The Way of Enterprise IT Architecture Transformation: Alibaba's Data Middle Platform Strategic Thinking and Architecture Practice*, China Machine Press, 2017.

[14] iResearch, "2021 China's data middle platform industry report_views_insights_iResearch," 2021, https://www.iresearchchina.com/content/details7_66750.html.

[15] F. Chang, J. Dean, S. Ghemawat et al., "Bigtable: A Distributed Storage System for Structured Data," in *OSDI '06: 7th SYMPOSIUM on OPERATING SYSTEMS DESIGN & IMPLEMENTATION, 26:1–26*, Association for Computing Machinery, 2008.

[16] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," in *OSDI '04: 6th SYMPOSIUM on OPERATING SYSTEMS DESIGN & IMPLEMENTATION*, pp. 137–150, Association for Computing Machinery, 2004, https://www.usenix.org/legacy/event/osdi04/tech/full_papers/dean/dean_html/.

[17] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google File System," in *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles (SOSP '03)*, Association for Computing Machinery, New York, NY, United States, 2003.

[18] B. Fang, J. I. A. Yan, L. I. Aiping, and J. I. A. N. G. Rong, "Privacy preservation in big data: a survey," *Big Data Research*, vol. 2, no. 1, pp. 1–18, 2015, https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2016&filename=DSJU201601001&v=LMMdV02dFjf3gB5xRXCP0b%25mmd2FauhrBw4OlLAphLPIl1sbZR1x8vozNB9DondjvnCrf.

[19] Y. Fan, "Big data security and privacy protection," *ZTE Technology Journal*, vol. 2016, no. 2, pp. 1–8, 2016.

[20] CAICT Security Research Institute, *Big Data Security White Paper*, CAICT, Beijing, China, 2018.

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

[22] M. J. M. Chowdhury, A. Colman, M. A. Kabir, J. Han, and P. Sarda, "Blockchain as a notarization service for data sharing with personal data store," in *2018 17th IEEE international conference on trust, security and privacy in computing and communications/12th IEEE international conference on big data science and engineering (TrustCom/BigDataSE)*, New York, NY, USA, 2018.

[23] O. B. Mora, R. Rivera, V. M. Larios, J. Raul Beltrán-Ramírez, R. Maciel, and A. Ochoa, "A use case in cybersecurity based in blockchain to deal with the security and privacy of citizens and smart cities cyberinfrastructures," *IEEE Xplore.*, vol. 2018, 2018.

[24] J. Yang, J. Wen, B. Jiang, and H. Wang, "Blockchain-based sharing and tamper-proof framework of big data networking," *IEEE Network*, vol. 34, no. 4, pp. 62–67, 2020.

[25] L. Tan, N. Shi, C. Yang, and Y. Keping, "A lockchain-based access control framework for cyber-physical-social system big data," *IEEE Access*, vol. 8, pp. 77215–77226, 2020.

[26] J. Li, W. Jigang, G. Jiang, and T. Srikanthan, "Blockchain-based public auditing for big data in cloud storage," *Information Processing & Management*, vol. 57, no. 6, article 102382, 2020.

[27] J. Stodt and C. Reich, *Data confidentiality in P2P communication and smart contracts of blockchain in Industry 4.0*, Computer Science & Information Technology, 2020.

[28] C. Zhang, Y. Li, W. Sun, and S. Guan, "Blockchain based big data security protection scheme," in *2020 IEEE 5th information technology and mechatronics engineering conference (ITOEC)*, Chongqing, China, 2020.

[29] X. Yu, Z. Shu, Q. Li, and J. Huang, "BC-BLPM: a multi-level security access control model based on blockchain technology," *China Communications*, vol. 18, no. 2, pp. 110–135, 2021.

[30] Z. Haotian, L. Tao, and Y. Song, *Design and implementation of data middle platform*, Association for Computing Machinery, Chongqing, China, 2021.

[31] G. Lyu, P. Liu, L. Yiming, T. Wang, and X. Kang, "A data middle platform architecture based on microservice serving power grid business," in *2021 8th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2021 7th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, pp. 219–224, 2021.

[32] T. Ji and S. B. Goyal, "Anti-counterfeiting and traceability mechanism based on Blockchain," in *Recent Trends in Intensive Computing, Volume 39: Recent Trends in Intensive Computing:134–44*, IOS Press, 2021.