Advances in Science and Technology Research Journal

*Advances in Science and Technology Research Journal* 2024, 18(4), 163–176
https://doi.org/10.12913/22998624/188862
ISSN 2299-8624, License CC-BY 4.0

Received: 2024.05.18
Accepted: 2024.06.08
Published: 2024.06.20

# Secure Big Data Model Based on Blockchain Technology

Ahmad Alshammari[1*]

[1] Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Kingdom of Saudi Arabia

E-mail: ahmad.almkhaidsh@nbu.edu.sa

## ABSTRACT

Blockchain has been growing rapidly in the cryptocurrency age and is one of the best information technologies that provide security and privacy to the data of people in crypto economy. In most cases, tampering with data and problems regarding data authentication tend to occur when data is shared and stored on centralized servers. With the assistance of blockchain technology, big data can be managed and saved in the cloud, and the technologies that enhance security by keeping out pernicious users could be used. Therefore, this paper has two aims: to discover the advantages and disadvantages of existing security big data models and to develop a secure big data model based on blockchain technology. The design science method is used for the purposes of this study. The developed secure big data model consists of three main processes: dataset storage and encryption, verification and consensus, and access control mechanism. The finding of this study discovered that the developed secure big data model offers a mix of both traditional and modern security measures which helps domain practitioners understand the security concepts of the blockchain along with big data as well.

**Keywords:** big data, blockchain, design science method, secure big data.

## INTRODUCTION

In recent years, big data has become a key component of both science and business societies [1, 2]. Over the past few years, a variety of new technologies and applications have emerged, including social networks, smartphones, and other smart devices. Throughout the world, billions of people are generating vast amounts of data every single day thank to the widespread adoption of these new technologies and applications [3]. There has been a boom in data use in various applications as a result of technological advancements such as cloud computing, IoT, and AI. By using data effectively, corporations and organizations are able to obtain more benefits [4]. As a result of the growing concern for data fair trading, it has become increasingly important to form a bridge between the generation of data and its use. The data stores gathered by data curators or agents are the primary data sources for data trading platforms [5]. There are several reasons why individual data providers struggle to provide direct data to end users. This is because their data volumes are commonly too small to meet their clients' needs. Industrial firms have recently used a variety of big data trading platforms, and researchers have also developed many trading methods [6–16]. In the majority of existing solutions, however, data is traded via centralized third-party trading platforms that have the disadvantages of single points of failure, opaque transactions, and uncontrollability. Dynamic price negotiation requirements are difficult to meet with existing data trading pricing models. User data may also be covertly analyzed and processed by dishonest data platforms. A set of accountability protocols have been proposed by [6, 7] to keep track of bookkeeping and to hold buyers and sellers accountable if they are dishonest. In these studies, the data provider must trust the broker (an agent who provides shopping services on behalf of the data provider). In order for the data broker to tamper with

or resell the data, both the source data and the purchase data budget must be hosted by them.

A distributed ledger has multiple benefits, similar to blockchains [17], which are peer-to-peer networks with a shared database. This is considered to be one of the major contributions of bitcoin [18]. To solve the problem of valid trading blocs not being added to the blockchain as blocks, the solution proposed by bitcoin is for most mining nodes to agree on a consensus to join. Decentralization, trust, tamper-resistance, anti-counterfeiting, and traceability are some of the main characteristics of blockchain technology. This technology provides a wide range of value-enhanced applications. Many researchers and developers have become interested in this concept in the past couple of years. Besides cryptocurrency, blockchain has been recognized to have a significant impact in other areas, especially crowdsourcing [18, 19] the Internet of Things [20–22, 25], supply chains [23], medicine [24], [25], and energy trading [15], all of which have rapidly been developing in recent years.

Furthermore, as blockchain technology developed, several problems were revealed. In order for the blockchain system to be secure and to improve the usability of the applications, the design of the consensus protocol must consider the performance of the most popular consensus protocol [26]. Across blockchain networks, consensus protocols serve as an underlying mechanism for reaching agreement and validating transactions by allowing the networks to reach consensus. Using the consensus algorithm, all participants in a network would be assured that the current state of the blockchain is the same from one node to the next [26]. Each one of the consensus protocols has a set of strengths and weaknesses that make the protocol unique among the others. Proof of work (PoW) is one of the most popular consensus protocols [27]. It is a form of PoW that miners use to validate and secure network transactions by solving complex puzzles. There has been criticism about PoW because of its scalability issues for a long time, despite its success in securing blockchain networks [27]. In addition to increasing the number of transactions, the computational power used for each transaction is also increasing, resulting in long processing times and high transaction fees as the number of transactions increases [28]. The authors of this study believe that a consensus protocol that considers the performance of the most popular protocol will improve the security and usability of blockchain applications. There are several approaches to achieving consensus, which can be combined into a hybrid protocol to ensure a successful decision-making process. To propose an effective hybrid protocol, the present study incorporates the strengths of exiting protocols into the proposed one, while avoiding their weaknesses. Smart contracts are based on protocols and provide a user interface that helps the parties involved execute the contract more efficiently [21]. The notion of a "smart contract" was first proposed by the cryptographer Szabo in 1994 [29]. In terms of functionality, smart contracts are similar to the If-Then statements, which are used in many software programs [30]. A smart contract enforces the conditions of a contract when a specified condition is met, i.e., when a preset event occurs. Because of the absence of a stable operating environment, these types of systems have not been used widely [30]. It is now possible to apply smart contracts to an increasingly large number of different fields because of the development of blockchain technology. As a result of its tamper-proofness, decentralization, openness, and transparency, blockchain provides a great platform for executing smart contracts.

The proxy re-encryption relies on the concept of proxy re-encryption [31], which is based on a public key system and is a cryptographic system that has the function of secure conversion. Re-encryption is a process in which ciphertext is converted by a semi-trusted proxy for the purpose of proxy re-encryption. For the same plaintext, the delegator and delegate can encrypt the ciphertext based on their own private keys, and the delegator can then use their private keys to decrypt and convert that ciphertext based on their own private keys. That way, if the proxy has a proxy re-encryption key, it is able to authorize the delegator to re-encrypt the message, and the proxy does not have access to the corresponding plaintext in the ciphertext during this conversion process. In addition to its widespread use in the cloud computing field [32, 33], digital copyright [34], and many other fields, proxy re-encryption technology is found in many disciplines. Therefore, this study has two aims: i) to discover the advantages and disadvantages of the existing security big data models, and ii) to develop a secure model for big data using blockchain technology. To achieve these objectives, the design science

methodology is used in this study, and the following questions are answered:

- What are the existing blockchain security models proposed for big data?
- What is the methodology, aim, outcomes, benefits, and weaknesses of the existing security blockchain models proposed for big data?
- How can blockchain technology be used to secure big data?

The rest of the paper is organized as follows: Section 2 describes the related work. Next, Section 3 explains the methodology used. Then, Section 4 reports the findings and discusses the results. Finally, Section 5 presents the conclusions of the research and recommendations for future work.

## RELATED WORK

Since the beginning of the past decade, scholars have become interested in the use of blockchain-based data sharing models that consist of a variety of functions, including data storage, secure data sharing, and secure authority management. For example, the authors in [35] focused on integrating big data analytics into security information management environments to enhance the management of information security. A detection model was developed to improve the detection and response to threats, the response to incidents, and the communication and collaboration among parties. A limitation of the developed model is that it requires a great deal of sensitive information to manage and process big data analytics. A framework for securely sharing sensitive data was provided in [36] using the design science research method, Their framework focuses on the big data platform. The outputs of the framework are secure delivery, storage, sharing, and destruction of sensitive data, which are just a few of the features outlined in the paper. While the framework provides some levels of security for sharing information, it has a disadvantage: unauthorized parties can access or accidentally disclose shared information. In [37], an eHealth data privacy and security model was developed to secure and protect clinical data. It was designed based on encryptions and attribute-based access controls as a way to secure and protect clinical data. The developed model helps to provide hierarchical access to healthcare data through cryptographic tools to ensure a secure system. However, it focused on electronic health records only. The author in [38] focused on the characterization of phishing attacks using big data to mitigate the risk of cyberattacks by developing a big data framework architecture for cyberattack defenses. Developed frameworks were used to continuously process large quantities of diverse security data and ensure the continuous process of data analysis. Although, there are some limitations to this approach; for example, to extract meaningful insights from large volumes of security data, a specialist with expertise in the field of data science and statistics may need to be involved. In [39], the researchers proposed a model that focuses on automating the triage of cybersecurity data. The model also automates the generation of data triage automated models based on operation traces. It reduces the generation costs of these models by order of magnitude compared to existing methods. Using the data triage operations of analysts and analyzing the patterns derived from those operations, the researchers demonstrated that patterns can be gathered. However, the model has not been successfully implemented in the real world. By creating proactive cybersecurity systems, the authors in [40] concentrated on the development of big data infrastructure to prevent vulnerabilities. This developed system helps identify and retrieve data from public sources, transform (clean) and load it, cluster, and visualize it, and curate it. In spite of these advantages, it lacks implementation and is limited to ad hoc changes. In [41], the authors focused on the analysis of network traffic data to detect the network anomalies, by developing a detection metamodel that used metamodeling methodology. The developed metamodel has several advantages regarding scalability, flexibility, model selection capabilities, and the ability to integrate model structure across a range of datasets, which is more than adequate for extracting insights from large datasets. Owing to its complexity and scalability, the developed metamodel does not provide the same flexibility and extensibility as existing metamodels. A study carried out by [42] examined Rivest Cipher4 (RC4) security on Transport Layer Security. The researchers presented a new method for analyzing the security of RC4 based on the use of big data analysis. This method has some advantages, including the ability to obtain the key distribution from a large set of keys without the need of a computer program. However, any secure protocol that uses RC4 as a deterministic mechanism should be deprecated as soon as possible, as there is a lack of sufficient security margin in RC4 to provide adequate protection from attacks. The

researchers in [43] discussed the computational intelligence-based big data analysis approach for windows desktop users. They investigated how desktop users could identify malicious attacks in their untouched files by developing a digital threat detection approach based on computational intelligence. To this end, the authors applied methods such as computational intelligence and Emeditor. In this approach, only simulated data is available on the platform based on a few protocols, firewall rules, and IP addresses. A study conducted by [44] focused on the security of multimedia big data, developing a model containing multiple levels of intelligence to preserve the confidentiality of multimedia big data with multiple levels of encryption. Their developed model helps optimize resource distribution at the system level in multiple streams in the system. Furthermore, experiments have shown that the presented schemes are capable of accommodating both real-time and asynchronous applications. The authors in [45] investigated the way big data can be used to create phishing pages based on Enron e-mail information. The study found that phishers and hackers are able to create big data security threats through big data analysis by understanding the behavior of email users through big data analysis. Consequently, big data brings us much joy and creates many opportunities for new ideas and inventions. There are a few limitations with this software, for example, it does not cover the prevention of such big data security threats through email. In [46], the authors concentrated on the monitoring and analysis of network traffic to detect anomalies in network monitoring applications. They developed a big data analytics framework using off-the-shelf big data storage and processing engines to perform big data analytics. With the developed framework, users are able to process unstructured and structured heterogeneous data sources in both stream-based and batch-based manners. In [47], the authors developed an ecosystem for blockchain access control to manage and protect large data sets against breaches, particularly other types of data breaches. Through the use of blockchain technology, the developed systems provide data transparency and traceability within a network, which solves the challenges of access control presented by traditional and centralized methods. Its drawback, though, is that it does not seem to have been implemented in the real world. The authors in [48] developed a trusted system for storing, sharing, and analyzing big data with all participants across the whole big data world, using Physical Unclonable Functions and Trusted Security Modules. The framework helps trade private data and reduce the conflicts between a rise in privacy invasions and an increase in privacy loss in the future. Based on blockchain technology and deep learning techniques, [49] proposed a framework for protecting data in smart power networks. It was found that this method outperforms previous methods and it can avert data poisoning and inference attacks by manipulating the original datasets. However, the scalability and utility of the application have not been evaluated. The authors in [50] presented a unique approach to sharing data among smart cities, called "Privy-Sharing", which is based on blockchain technology. With the proposed model, smart cities can secure the integrity and privacy of users' data and protect the data from the attacks launched by insiders and outsiders. Users would also have the right to forget about their personal information. However, this model has some limitations, most notably that peers have to maintain multiple ledgers places a tremendous strain on networks. In [51], a new blockchain-based telehealth architecture was designed to provide secure access to data in such a way that patients' privacy could be kept unchanged. The system is designed so that most of common attacks on telemedicine systems, including impersonation, replay attacks, and data tampering, are mitigated. The only limitation of their model is the limited ability to store and retrieve the data. Researchers in [52] proposed blockchain-based hierarchical models to protect IoT data and interoperate blockchains in smart cities using Hyperledger Fabric and Ethermint technologies. The proposed model addressed the challenges related to data management, established data integrity, and enabled interoperability among smart cities. In [53], the author proposed a modular Hyperledger fabric that ensures big data integrity is secure and verifiable. In the proposed framework, consensus protocols are used to validate the accuracy of evidence of big data recorded during live HFM operations. It has not been implemented in the real world. The authors in [54] offered an industrial big data analytics framework to provide references for big data analytics that maintain privacy and security in the industrial sector. The offered framework assisted in the construction of an industrial big data analytics platform that preserves privacy and is secure. It is the first framework proposed for privacy-preserving and secure industrial big data analytics, but it lacks implementation. In [54], a framework for big data analytics

applicable to the industrial sector was proposed in order to provide reference material for big data analytics and, at the same time, maintain privacy and security for industrial data. A framework was provided to aid in creating a data analytics platform that not only can handle big data for industrial purposes, but also is both secure and privacy preserving. Nevertheless, it is a first attempt at securing and protecting industrial big data analytics. The authors in [55] proposed a blockchain-based decentralized data trading system that contains smart contracts for matching of data, pricing negotiation, and rewards assignment during the course of data trading. Data users will be able to generate revenues from applications using the proposed data trading system based on the quality of the data. Based on blockchain technology, [56] proposed a general solution for outsourcing services payments, which can protect against malicious participant cheating and ensure correct service enforcement. In [57], the authors discussed the concept of blockchain technology, as well as some of the most exciting research trends in this field. Furthermore, they investigated in depth how blockchain security can be adapted to the cloud computing environment and discussed the related secure solutions. The authors in [58] focused on IoT data integrity in smart cities and attempted to find an effective way to ensure that the infrastructure for C-ITS provides both privacy and security. To this end, a secure framework based on a privacy-preserving protocol was proposed. The proposed framework is able to securely communicate data and to protect original datasets from data poisoning attacks. However, it lacks evaluating scalability and utility with different real-world datasets.The blockchain has been the subject of intense discussion among researchers over the past ten years. These models offer innovative solutions to address data storage, secure data sharing, and authority management. By leveraging blockchain technology, these models provide enhanced data integrity, privacy, and access control. This makes them an ideal choice for organizations seeking reliable and efficient data sharing solutions.

## METHODOLOGY

The design science method is used in this study to achieve the research objectives stated in Introduction. This method emphasizes the practical importance of iteration, documentation, and validation of findings [59]. It is one of the most powerful frameworks that has ever been developed for exploring and evaluating innovative solutions to complex problems. By promoting collaboration, ethical considerations, and empirical evidence, the design science methodology ensures that innovative solutions are developed efficiently and effectively. It can be applied in various domains and disciplines, providing a versatile and valuable tool for innovation and problem-solving. The design science method adapted to the present paper involves the following five stages (Figure 1):

- identification stage – this stage was dedicated to identifying the common online databases. This resulted in the identification of six online databases: IEEE Xplore, Scopus, Web of Science, Springer Link, Science Direct, and Google Scholar;
- assigning search protocols stage – this stage assigned four search protocols: the keywords, the language, the type of articles, and the period of publication. The keywords were set to be "Secure big data" and "Blockchain", the language to be English, the type of articles to be journal articles and conference articles with a high-quality content, and finally the publication period to be 2015–2024;
- collecting data stage – data was collected from the six online databases mentioned above based on the search protocols identified. The search resulted in a total of 693 articles among which 450 articles were extracted from Google Scholar, 32from Science Direct, 27 from IEEE Xplore, 140 from Springer Link, 12 from Web of Science, and 32 from Scopus (see Figure 2);
- filtering data stage – the 693 articles collected in the previous stage were filtered based on the criteria adapted from [60, 61]. Those articles with readable results published in journals or conferences with an excellent result and good literature were included in this study. The articles published in books, book chapters, reports, or case studies were not included in this study. After screening and filtering the articles and removing the duplicated ones, 21 articles were selected for the purposes of this study. These 21 were mainly focused on the secure handling of big data with the use of blockchain technology (Table 1). Next stage will analyze the selected articles regarding their advantages and disadvantages;
- analyzing data stage – this stage analyzes the articles identified and selected in the previous stage. Table 1 demonstrates that these articles

focused on a wide range of sectors. Table 2 displays the comprehensive analysis of the filtered data. It highlights the advantages, disadvantages, finding, and the used techniques/methods.

- Developing Stage: In this stage, a secure big data model was developed based on blockchain technology. The model involves three main processes: dataset storage and encryption process, verification and consensus process, and access control mechanism (Figure 3). In the following, each process is described in detail.

Dataset storage and encryption The aim of this process is to store and encrypt the dataset on the blockchain. This process ensures that sensitive data is protected and accessible to authorized users only. By implementing encryption techniques, individuals can safeguard their information from unauthorized access, tampering, or disclosure. To store the dataset on the blockchain, it is broken down into smaller chunks, as shown in Figure 3. This is necessary to accommodate the limited storage capacity of the blockchain network. Breaking

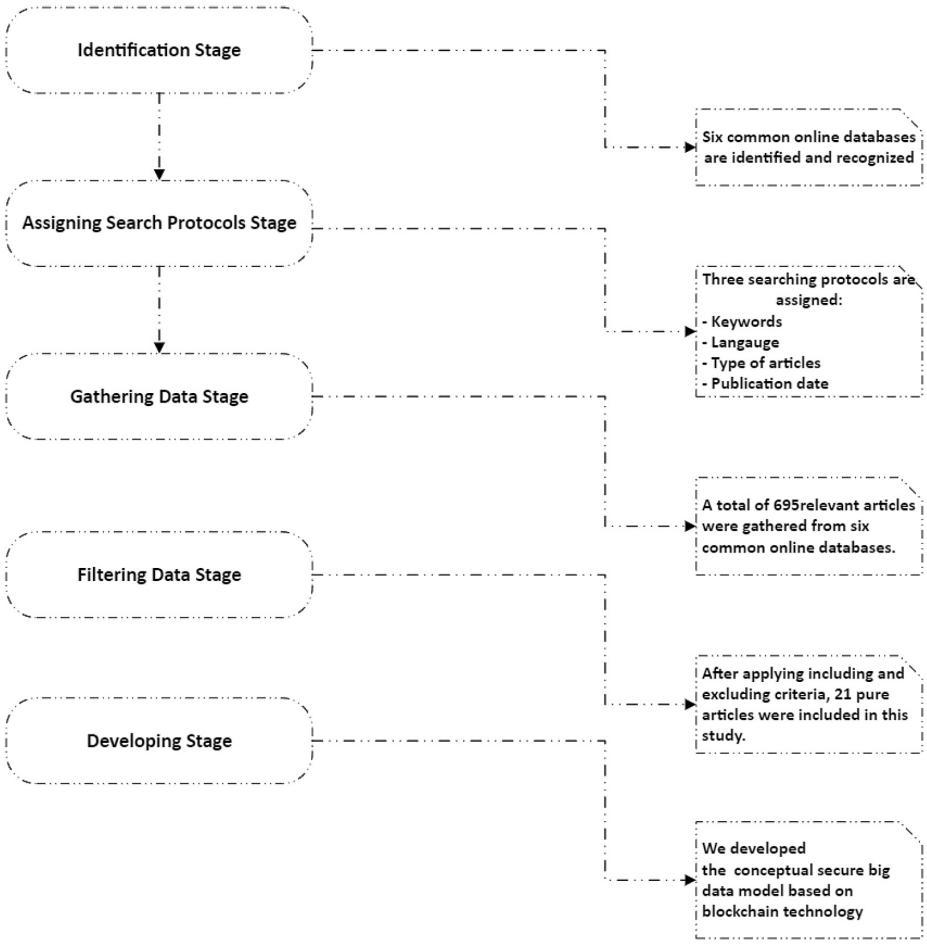


**Figure 1.** The adapted methodology [59]

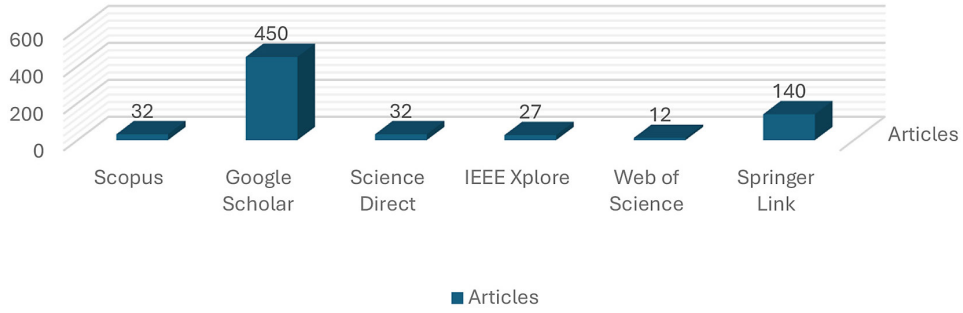


**Figure 2.** Gathered data from six common online database

**Table 1.** Articles selected to be analyzed in this study

| ID | Year | Ref. | Focus | Goal |
|---|---|---|---|---|
| 1 | 2015 | [35] | Integrating big data analytics into security information management environments | To enhance information security management capability |
| 2 | 2015 | [36] | Big data boards | To protect sensitive user data effectively |
| 3 | 2015 | [37] | Electronic health information | To secure and protect clinical data |
| 4 | 2016 | [38] | Big data classification of phishing | To mitigate the risk of cyberattacks |
| 5 | 2016 | [39] | Automating the triage of cybersecurity data | To reduce the generation costs of these models by orders of magnitude compared to existing methods |
| 6 | 2016 | [40] | The development of big data infrastructure | To avert vulnerabilities |
| 7 | 2016 | [41] | Examining the network stream of traffic data | To detect anomalies in network traffic in order to detect network accidents |
| 8 | 2016 | [42] | Examining RC4 security on Transport Layer Security | To acquire the key distribution from a large set of keys without a computer program |
| 9 | 2016 | [43] | Investigation of the users of Windows desktops | To investigate how desktop users can identify malicious attacks in their untouched files |
| 10 | 2016 | [44] | The security of multimedia big data | To preserve the confidentiality of multimedia big data with multiple levels of encryption |
| 11 | 2017 | [45] | Investigating how big data can be used to create phishing pages based on Enron E-mail information | To create phishing pages using the e-mail information obtained from Enron |
| 12 | 2017 | [46] | The monitoring and analysis of network traffic | To detect anomalies in network monitoring applications through monitoring the network traffic |
| 13 | 2018 | [47] | Developing an ecosystem for blockchain access control | To manage and protect large data sets against breaches, particularly other types of data breaches |
| 14 | 2018 | [48] | Blockchain technology and trusted security modules | To prevent the future loss of privacy by trading private data and reducing privacy invasions |
| 15 | 2019 | [49] | Intelligent energy networks | To protect data in Intelligent Energy Networks |
| 16 | 2020 | [50] | Smart cities that preserve privacy and protect data | To apply Privy-Sharing to sharing data among smart cities |
| 17 | 2021 | [51] | Smart healthcare | To provide secure access to data in a way that keeps patients' privacy unchanged |
| 18 | 2021 | [58] | Iot data integrity in smart cities | To ensure that the infrastructure for C-ITS provides both privacy and security |
| 19 | 2022 | [52] | Industrial iot | To protect IoT data and interoperate blockchains |
| 20 | 2023 | [54] | Industrial big data analytics | To ensure big data integrity is secure and verifiable |
| 21 | 2024 | [53] | Data-protected and privacy-preserving smart cities | To protect privacy and security of smart cities |

down the dataset into smaller units allows for efficient and secure storage on the blockchain. Each chunk is then encrypted using cryptographic algorithms. These algorithms are specially designed to encrypt data in a way that makes it unreadable to any unauthorized party. By encrypting the data, it becomes unintelligible to anyone who does not have the necessary decryption keys. The encrypted data is stored in blocks that are securely chained together to form a ledger. The ledger acts as a chronological record of all the encrypted data chunks added to the blockchain.

## Verification and consensus

This process plays a vital role in the functioning of blockchain technology. Once encrypted data is added to the blockchain, it is verified by a network of nodes. These nodes are spread out geographically, providing redundancy and decentralization. Consensus is reached among nodes through cryptographic algorithms, ensuring that the integrity of the data is maintained. Therefore, this process consists of five concepts:

- verification – verification is the procedure of ensuring that the data added to the blockchain is accurate and trustworthy. It involves checking the integrity of the data against predefined rules or algorithms. In this process, the data is verified to ensure that it has not been altered since it was originally generated;
- consensus – it refers to the procedure of reaching agreement among a network of nodes regarding the validity of a transaction or block. It involves nodes coming to a consensus on the current state of the blockchain

and which transactions should be added to the next block;
- a network of nodes – it is a fundamental component of blockchain technology. The nodes are distributed geographically, with each node holding a copy of the blockchain. The spread of nodes across different geographic locations adds redundancy to the network, making it more resilient to attacks. If one node goes offline or experiences technical difficulties, other nodes can step in and continue verifying and updating the blockchain.
- cryptographic algorithms – consensus is reached among nodes through cryptographic algorithms. These algorithms rely on mathematical principles to ensure the integrity and authenticity of the blockchain. By employing cryptographic techniques, nodes can validate transactions and blocks without relying on a central authority.
- integrity of data – the integrity and authenticity of the data stored on the blockchain are crucial. Through verification and consensus, blockchain technology ensures that data remains unadulterated and tamper-proof. The cryptographic algorithms used by nodes verify the data and ensure that it has not been altered in any way. This protects the integrity and security of sensitive information stored within the blockchain.

### Access control mechanism

This mechanism ensures that data access is only granted to authorized individuals, depending upon their authorization level. This process involves two concepts:
- managing access with smart contracts – access control in the secure big data model is managed using smart contracts. In the context of access control, smart contracts are employed to identify and impose the rules that govern access to data. By using smart contracts, the verification process could be automated, removing the demand for guide intervention. As a result, both time and human error decrease. Smart contracts are programmed to ensure that access is granted only to users who meet predefined criteria, such as authorization level, role, or specific permissions.
- compliance with predefined rules – the secure big data model ensures that access control is compliant with predefined rules. These rules can be defined and enforced by the organization responsible for managing the data. Smart contracts enforce these rules automatically, ensuring that access is granted or denied based on the predefined criteria. This compliance is important for guaranteeing the security and secrecy of the data. By implementing access control using smart contracts, corporations can guarantee that their data is protected and only accessed by authorized individuals. This helps to maintain data integrity and relieve the risk of unauthorized access or data breaches.

## RESULTS

This section discusses the findings of the study. The present research comprehensive analyzed the existing studies conducted on securing big data based on blockchain technology and then developed a security model for big data using blockchain technology. The following are the three research questions:

a) What are the existing blockchain security models proposed for big data? To answer this question, several blockchain security models previously proposed for big data were identified and analyzed in this paper. The identified models consider the security of big data from different perspectives, such as access control, encryption, authorization, authentication, detection, and prevention, as well as blockchain technology.

b) What is the methodology, aim, outcomes, benefits, and weaknesses of the existing security blockchain models proposed for big data? To tackle this question, this paper thoroughly analyzed all the 21 identified models and discovered that the existing security blockchain models proposed for big data have different outcomes, benefits, and weaknesses. A detailed description of the methodology, aim, outcomes, benefits, and weaknesses of the existing security blockchain models is given in Table 2. Despite the weaknesses of most of the proposed models resulting from the lack of implementation in the real world, they have offered several advantages and provided many clues for overcoming the security challenges of big data. Nevertheless, blockchain technology has proven to be one of the best technologies that can be used to secure sensitive data when it is being transferred in the field of big data.

**Table 2.** Comprehensive analysis of the filtered data

| ID | Year | Ref. | Advantages | Disadvantages | Findings/results | Techniques/ methods |
|---|---|---|---|---|---|---|
| 1 | 2015 | [35] | Improving the detection and response to threats, the response to incidents, and the communication and collaboration among parties | Requiring a great deal of sensitive information to manage and process big data analytics | Improving security information management | Big data analysis |
| 2 | 2015 | [36] | Securing delivery, storage, sharing, and destruction of sensitive data, which are just a few of the features outlined in the paper | Shared information can be accessed by unauthorized people or accidentally disclosed. | Number of secure features dealing with sensitive data delivery, storage, sharing, and destruction | Proxy re-encryption |
| 3 | 2015 | [37] | Ensuring that the electronic health records are trusted and reliable | Focusing on electronic health records only | Hierarchical access to healthcare data was provided through cryptographic tools to ensure a secure system | Access control and encryption |
| 4 | 2016 | [38] | Handling large quantities of diverse security data in a continuous manner and ensuring the continuous process of data analysis | Security data may require specialist analysis, such as data science and statistics, to extract meaningful insights | Big data framework architecture for cyberattack defenses | Real experiments and datasets |
| 5 | 2016 | [39] | Generating data triage automata from operation traces, greatly reducing production costs | Not being successfully implemented in the real world | Analyzing triage data and mining it to identify trends has been conducted by 69 researchers | Data triage operations |
| 6 | 2016 | [40] | Helping to identify and retrieve data from public sources, to transform (clean) and load it, to cluster and visualize it, and to curate it | Lacking implementation and being limited to ad hoc changes | Ensuring the cyber-security of business infrastructure by deploying a cybersecurity system that plays a proactive role as part of a proactive approach | Extracting information techniques |
| 7 | 2016 | [41] | Facilitating the obtainment of insights from large datasets with its scalability, flexibility, and model selection capabilities | Not providing the same flexibility and extensibility as existing metamodels | An anomaly detection system can be leveraged by network administrators to ensure a reliable and secure network infrastructure. | Meta-modeling methodology |
| 8 | 2016 | [42] | Determining the key distribution from a large set of keys using the distribution of the keys | Lacking sufficient security margin in RC4 to provide adequate protection from attacks | The security of the RC4 algorithm was analyzed by analyzing large amounts of data | Transfer Level Security |
| 9 | 2016 | [43] | Working successfully regarding the tasks defined | Only simulated data is available on the platform based on a few protocols, firewall rules, and IP addresses. | It introduced a number of ways in which big data security can be analyzed, including computational intelligence, in the context of Windows desktop environments, which is one way in which big data security can be analyzed | Computational intelligence and emeditor |
| 10 | 2016 | [44] | Optimizing resource distribution at the system level in multiple streams within the system | Being limited for multimedia big data security | The presented schemes were found capable of accommodating real-time applications and asynchronous applications | Multiple intelligence levels in encryption control |
| 11 | 2017 | [45] | The use of big data to create a great deal of joy and support new possibilities. | Not covering email as a method of preventing such threats to big data security | The study found that phishers and hackers can create big data security threats through big data analysis by understanding the behavior of email users, through big data analysis | Dataset for Enron emails: a case study |
| 12 | 2017 | [46] | Developing Big-DAMA that is capable of both stream and batch processing of unstructured and structured heterogeneous data sources | Being limited for anomaly detection | Users were found capable of processing unstructured and structured varied data resources in both stream-based and batch-based ways | Engines and storage for big data off the shelf |

| | | | | | | |
|---|---|---|---|---|---|---|
| 13 | 2018 | [47] | Providing data transparency and traceability within a network, which solves the challenges of access control presented by traditional and centralized methods | Not having been implemented in the real world | In contrast to conventional access control approaches, network-based access control results were found more transparent and traceable | Blockchain technology |
| 14 | 2018 | [48] | Facilitating the exchange of private data, reducing conflicts arising from invasions of privacy | Lacking execution in the real world | A significant step toward realizing the full potential of data-driven initiatives has been made by discovering how to collect and use big data trust-based | Physical unclonable function, and trusted security module |
| 15 | 2019 | [49] | Manipulating the original datasets to prevent data destroying and the inference attacks that are associated with them | Not evaluating the scalability and utility of the application | The method was found effective in outperforming previous methods and preventing data poisoning and inference attacks. It was because the method prevented the original datasets from being manipulated in a way that would allow data poisoning and inference attacks to occur | Blockchain technology and deep learning techniques |
| 16 | 2020 | [50] | Protecting data from external and internal attacks, and giving users the right to forget about their personal information | Placing a tremendous strain on networks as peers need to maintain multiple ledgers | It was found that “Private Sharing,” a blockchain-based solution developed by Airbnb to share sensitive data with citizens, can be used to share data we find it hard to share with others | Private data collection |
| 17 | 2021 | [51] | Mitigating most common attacks on telemedicine systems, including impersonation, replay attacks, and data tampering | Having limited data storage and retrieval | By leveraging the security, interoperability, and privacy-preserving features of blockchain technology, healthcare providers can provide secure and efficient telehealth services, benefiting both patients and healthcare systems | Design science approach |
| 18 | 2021 | [58] | Helping C-ITS participants to securely communicate data among themselves and to protect original datasets from data poisoning attacks | Lacking evaluating scalability and utility with different real-world datasets | A number of advantages were provided by the proposed framework both for blockchain- and non-blockchain-based methods | Hyperledger fabric and ethermint |
| 19 | 2022 | [52] | Guaranteeing data integrity enabling interoperability in smart cities | Interoperability in smart cities does not allow communication among services | Providing examples of how a consortium blockchain can be established using a trust-based model based on trust | Design science method |
| 20 | 2023 | [54] | Using compromise protocols to confirm the signing of evidence captured during HFM operations in real-time | Lacking real-world implementation | Proposing a model capable of protecting the privacy and security of players in the industrial sector from attacks relating to big data analysis | Analysis and DSR |
| 21 | 2024 | [53] | Aiding with establishing a secure and privacy-preserving big data analytics platform for industry | Lacking real-world implementation | An industrial framework for big data analytics should not only ensure privacy and security, but also include privacy protection | Private data collection |

c) How can blockchain technology be used to secure big data? To answer this question, the secure big data model based on blockchain technology was developed in this study. It explains how the blockchain technology concept can improve the security of big data. The developed model involves three main processes to overcome the security issues with the big data: dataset storage and encryption, verification and consensus, and access control mechanism. As a result, the data is stored and encrypted on the blockchain to preserve its security. It ensures

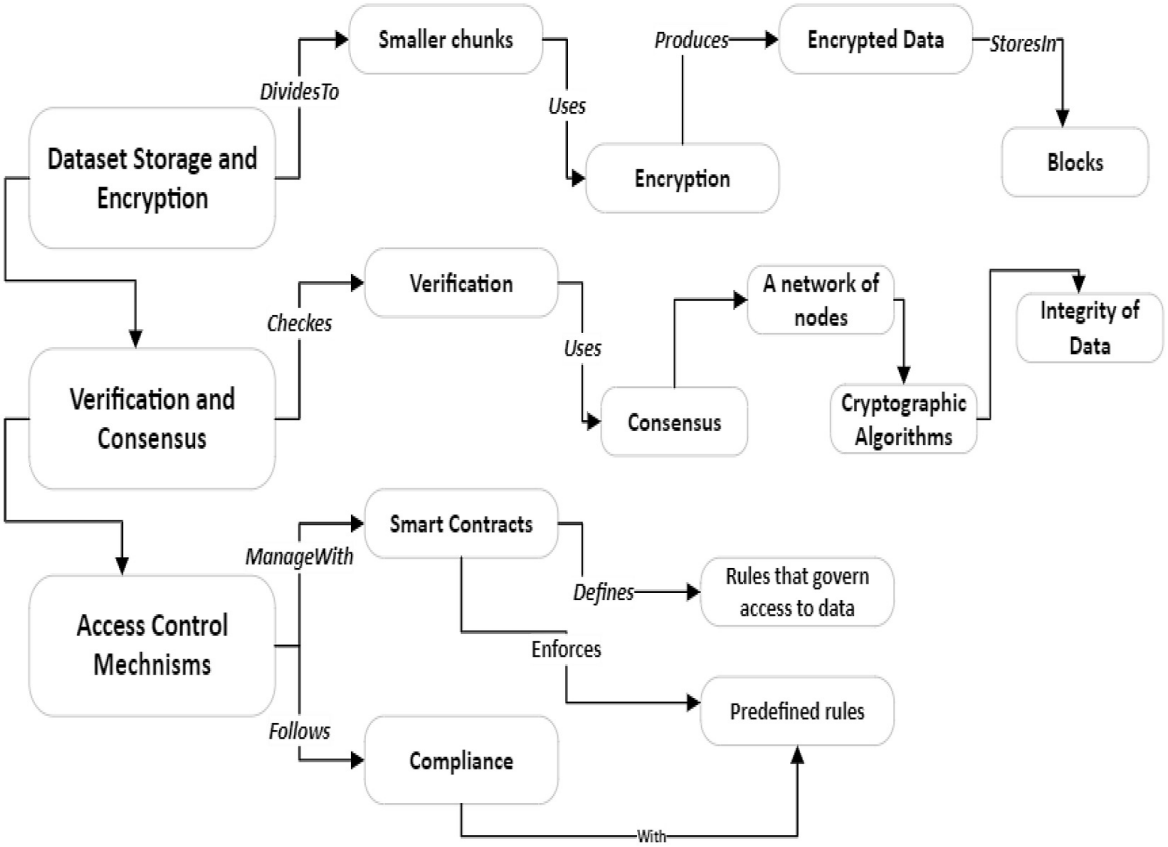


**Figure 3.** Secure big data model developed based on blockchain technology

the security of sensitive data and that only authorized users can access it. A person can encrypt their personal data using blockchain so that it is protected from being accessed, tampered with, or compromised by unauthorized individuals or groups. Whenever encrypted data is added to the blockchain, a network of nodes verifies that the data is legitimate. As a result of the geographical dispersion of these nodes, decentralization and redundancy are provided. Cryptographic algorithms are used to reach consensus among the nodes to ensure that the integrity of data is maintained throughout the network.

Therefore, the developed secure big data model based on blockchain technology provides a mix of both traditional and modern security measures. This helps domain practitioners understand the security concepts of the blockchain along with big data as well**.** In addition, organizations can use several significant advantages associated with using blockchain technology, including data privacy, security, integrity, access controls, auditability, provenance, and significant cost savings, as part of a developed model by combining it with big data. With the advent of blockchain technology, it was demonstrated that companies can securely manage, store, and share large amounts of information in an immutable and decentralized way, which, in turn, enables them to make informed business decisions and comply with regulatory requirements in a decentralized and immutable way.

## CONCLUSIONS

Blockchain technology has grown rapidly in the era of cryptocurrency and is one of the leading technologies that are aiding in ensuring that the data and personal information of people are kept secure and private in the crypto economy. As a general rule, when data are shared and stored on centralized servers, it is more likely that data will be tampered with and there will be problems concerning data authentication. Blockchain technology has the potential to help manage and store big data securely in the cloud and to enhance security by keeping out pernicious users. Accordingly, this paper discovered the advantages and disadvantages of the

existing security big data models and developed a secure big data model based on blockchain technology. To accomplish this task, the design science approach was used. As described above, the developed secure big data model is composed of three main processes, each of which is responsible for securing data: dataset storage and encryption, verification and consensus, and access control mechanism. This study found that the developed secure big data model combines traditional and modern security procedures in order to help domain practitioners understand both big data and blockchain security concepts in order to allow them to implement both. In future research, the developed model could be tested in real-world environments to verify its effectiveness and capabilities.

## Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2024-2990-03".

## REFERENCES

1. Wang R., Xu C., Ye F., Tang S., and Zhang X., "S-MBDA: A Blockchain-Based Architecture for Secure Storage and Sharing of Material Big-Data," IEEE Internet Things J., 2024.
2. Mythili M.E.J., Nareash Kumar K.C., Athanan M.M., and Rahman M.M. Secure Information Sharing For The Supply Chain Using Blockchain Technology, Adv. Sci. Technol. Res. J., 2024, 18(1).
3. Maheshwari S., Gautam P., and Jaggi C.K. Role of Big Data Analytics in supply chain management: current trends and future perspectives, Int. J. Prod. Res., 2021, 59(6): 1875–1900.
4. Sagiroglu S., and Sinanc D. International conference on collaboration technologies and systems (CTS), IEEE, 2013, 2013: 42–47.
5. Liu Z., Huang B., Li Y., Sun Q., Pedersen T.B., and Gao D.W. Pricing game and blockchain for electricity data trading in low-carbon smart energy systems, IEEE Trans. Ind. Informatics, 2024.
6. Jung T., Li X.Y., Huang W., Qiao Z., Qian J., Chen L., Han J., Houet J. Accounttrade: accountability against dishonest big data buyers and sellers, IEEE Trans. Inf. Forensics Secur., 2018, 14(1): 223–234.
7. T. Jung, Li X.Y., Huang W., Qian J., Chen L., Han J., Hou J., Su C.Accounttrade: Accountable protocols for big data trading against dishonest consumers, in IEEE INFOCOM 2017-IEEE Conference on Computer Communications, IEEE, 2017, 1–9.
8. Dai W., Dai C., Choo K.-K.R., Cui C., Zou D., Jin H. SDTE: A secure blockchain-based data trading ecosystem, IEEE Trans. Inf. Forensics Secur., 2019, 15: 725–737.
9. Chen J., Lv Z., Song H., Design of personnel big data management system based on blockchain, Futur. Gener. Comput. Syst., 2019, 101: 1122–1129.
10. Yue L., Junqin H., Shengzhi Q., and Ruijin W. Big data model of security sharing based on blockchain, in 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM), IEEE, 2017, 117–121.
11. Zyskind G., and Nathan O. Decentralizing privacy: Using blockchain to protect personal data, in 2015 IEEE security and privacy workshops, IEEE, 2015, 180–184.
12. Xiangqian D., Bing G., and Yan S. An Efficient and Secure Decentralizing Data Sharing Model [J], Chinese J. Comput., 2018, 41(5): 1021–1036.
13. Zheng Z., Xie S., Dai H., Chen X., Wang H. An overview of blockchain technology: Architecture, consensus, and future trends, in 2017 IEEE international congress on big data (BigData congress), Ieee, 2017, 557–564.
14. Taleb I., Serhani M.A., and Dssouli R. Big data quality assessment model for unstructured data, in 2018 International Conference on Innovations in Information Technology (IIT), IEEE, 2018, 69–74.
15. Jiang Y., Zhou K., Lu X., Yang S., Electricity trading pricing among prosumers with game theory-based model in energy blockchain environment, Appl. Energy, 2020, 271: 115239.
16. Guan Z., Shao X., Wan Z. Secure fair and efficient data trading without third party using blockchain, in 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), IEEE, 2018, 1395–1401.
17. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system bitcoin: a peer-to-peer electronic cash system, bitcoin.org. Dispon. en https//bitcoin. org/en/bitcoin-paper, 2009.
18. Yang M., Zhu T., Liang K., Zhou W., Deng R.H. A blockchain-based location privacy-preserving crowdsensing system, Futur. Gener. Comput. Syst., 2019, 94: 408–418.
19. Ma H., Huang E.X., Lam K.-Y. Blockchain-based mechanism for fine-grained authorization in data crowdsourcing, Futur. Gener. Comput. Syst., 2020, 106: 121–134.
20. Kshetri N. Can blockchain strengthen the internet of things?, IT Prof., 2017, 19(4): 68–72.
21. Christidis K., and Devetsikiotis M. Blockchains and

smart contracts for the internet of things, IEEE access, 2016, 4: 2292–2303,.

22. [22] Mazzei D., Baldi G., Fantoni G., Montelisciani G., Pitasi A., Ricci L., Rizzello L. A Blockchain Tokenizer for Industrial IOT trustless applications, Futur. Gener. Comput. Syst., 2020, 105: 432–445.
23. Kshetri N.and Loukoianova E. Blockchain adoption in supply chain networks in Asia, IT Prof., 2019, 21(1): 11–15.
24. Azaria A., Ekblaw A., Vieira T., and Lippman A. Medrec: Using blockchain for medical data access and permission management, in 2016 2nd international conference on open and big data (OBD), IEEE, 2016, 25–30.
25. Chen L., Lee W.-K., Chang C.-C., Choo K.-K.R., and Zhang N. Blockchain based searchable encryption for electronic health record sharing, Futur. Gener. Comput. Syst., 2019, 95: 420–429.
26. Xiao Y., Zhang N., Lou W., and Hou Y.T. A survey of distributed consensus protocols for blockchain networks, IEEE Commun. Surv. Tutorials, 2020, 22(2): 1432–1465.
27. Seang S., and Torre D. Proof of Work and Proof of Stake consensus protocols: a blockchain application for local complementary currencies, Fr. Univ. Cote d'Azur-GREDEG-CNRS. Str, 2018, 3(4).
28. Guo H., Liang H., Huang J., Ou W., Han W., Zhang Q., Zhang R. A framework for efficient cross-chain token transfers in blockchain networks, J. King Saud Univ. Inf. Sci., 2024, 101968.
29. Tengyun J. On the Security of Blockchain-based Applications. Swinburne University of Technology, 2024.
30. Abdellatif T., and Brousmiche K.-L. Formal verification of smart contracts based on users and blockchain behaviors models, in 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), IEEE, 2018, 1–5.
31. Blaze M., Bleumer G., and Strauss M. Divertible protocols and atomic proxy cryptography, in International conference on the theory and applications of cryptographic techniques, Springer, 1998, 127–144.
32. Tran D.H., Nguyen H.-L., Zha W., and Ng W.K. Towards security in sharing data on cloud-based social networks, in 2011 8th International Conference on Information, Communications & Signal Processing, IEEE, 2011, 1–5.
33. Qin Z., Wu S., and Xiong H. Strongly secure and cost-effective certificateless proxy re-encryption scheme for data sharing in cloud computing, in Big Data Computing and Communications: First International Conference, BigCom 2015, Taiyuan, China, August 1–3, 2015, Proceedings 1, Springer, 2015, 205–216.
34. Lee S., Park H., and Kim J. A secure and mutual-profitable DRM interoperability scheme, in The IEEE symposium on Computers and Communications, IEEE, 2010, 75–80.
35. Gottwalt F., and Karduck A.P. SIM in light of big data, in 2015 11th International Conference on Innovations in Information Technology (IIT), IEEE, 2015, 326–331.
36. Dong X., Li R., He H., Zhou W., Xue Z., and Wu H. Secure sensitive data sharing on a big data platform, Tsinghua Sci. Technol., 2015, 20(1): 72–80.
37. Soceanu A., Vasylenko M., Egner A., and Muntean T. Managing the privacy and security of ehealth data, in 2015 20th International Conference on control systems and computer science, IEEE, 2015, 439–446.
38. Las-Casas P.H.B., Dias V.S., Meira W., and Guedes D. A big data architecture for security data and its application to phishing characterization, in 2016 IEEE 2nd international conference on big data security on cloud (BigDataSecurity), IEEE international conference on high performance and smart computing (HPSC), and IEEE international conference on intelligent data and security (IDS), IEEE, 2016, 36–41.
39. Zhong C., Yen J., Liu P., and Erbacher R.F. Automate cybersecurity data triage by leveraging human analysts' cognitive process, in 2016 IEEE 2nd International Conference on big data security on cloud (BigDataSecurity), IEEE International Conference on high performance and smart computing (HPSC), and IEEE International Conference on intelligent data and security (IDS), IEEE, 2016, 357–363.
40. Chen H.-M., Kazman R., Monarch I., and Wang P. Predicting and fixing vulnerabilities before they occur: a big data approach, in Proceedings of the 2nd International Workshop on BIG Data Software Engineering, 2016, 72–75.
41. Yang B., and Zhang T. A scalable meta-model for big data security analyses, in 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), IEEE, 2016, 55–60.
42. Liu C., Cai Y., and Wang T. Security evaluation of rc4 using big data analytics, in 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), IEEE, 2016, 316–320.
43. Naik N., Jenkins P., Savage N., and Katos V., Big data security analysis approach using computational intelligence techniques in R for desktop users, in 2016 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2016, 1–8.
44. Xiao C., Wang L., Jie Z., and Chen T. A multi-level intelligent selective encryption control model for multimedia big data security in sensing system with resource constraints, in 2016 IEEE 3rd International Conference on Cyber Security and Cloud

Computing (CSCloud), IEEE, 2016, 148–153.

45. Zaki T., Uddin M.S., Hasan M.M., and Islam M.N. Security threats for big data: A study on enron e-mail dataset, in 2017 international conference on research and innovation in information systems (icriis), IEEE, 2017, 1–6.
46. Casas P., Soro F., Vanerio J., Settanni G., and D'Alconzo A. Network security and anomaly detection with Big-DAMA, a big data analytics framework, in 2017 IEEE 6th international conference on cloud networking (CloudNet), IEEE, 2017, 1–7.
47. Uchibeke U.U., Schneider K.A., Kassani S.H., and Deters R. Blockchain access control ecosystem for big data security, in 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), IEEE, 2018, 1373–1378.
48. Guan Z., Zhao Y., Li D., and Liu J. Tbdct: A framework of trusted big data collection and trade system based on blockchain and tsm, in 2018 IEEE International Conference on Smart Cloud (SmartCloud), IEEE, 2018, 77–83.
49. Keshk M., Turnbull B., Moustafa N., Vatsalan D., and Choo K.-K.R. A privacy-preserving-framework-based blockchain and deep learning for protecting smart power networks, IEEE Trans. Ind. Informatics, 2019, 16(8): 5110–5118.
50. Makhdoom I., Zhou I., Abolhasan M., Lipman J., and Ni W. PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities, Comput. Secur., 2020, 88: 101653.
51. Younis M., Lalouani W., Lasla N., Emokpae L., and Abdallah M. Blockchain-enabled and data-driven smart healthcare solution for secure and privacy-preserving data access, IEEE Syst. J., 2021, 16(3): 3746–3757.
52. Rahman M.S., Chamikara M.A.P., Khalil I., and Bouras A. Blockchain-of-blockchains: An interoperable blockchain platform for ensuring IoT data integrity in smart city, J. Ind. Inf. Integr., 2022, 30: 100408.
53. Juma M., Alattar F., and Touqan B. Securing big data integrity for industrial IoT in smart manufacturing based on the trusted consortium blockchain (TCB), IoT, 2023, 4(1): 27–55.
54. Liu L., Li J., Lv J., Wang J., Zhao S., and Lu Q. Privacy-Preserving and Secure Industrial Big Data Analytics: A Survey and the Research Framework, IEEE Internet Things J., 2024.
55. Hu D., Li Y., Pan L., Li M., and Zheng S. A blockchain-based trading system for big data, Comput. Networks, 2021, 191: 107994.
56. Zhang Y., Deng R.H., Liu X., and Zheng D. Blockchain based efficient and robust fair payment for outsourcing services in cloud computing, Inf. Sci. (Ny), 2018, 462: 262–277.
57. Park J.H., and Park J.H. Blockchain security in cloud computing: Use cases, challenges, and solutions, Symmetry (Basel)., 2017, 9(8): 164.
58. Kumar R., Kumar P., Tripathi R., Gupta G.P., Kumar N., and Hassan M.M. A privacy-preserving-based secure framework using blockchain-enabled deep-learning in cooperative intelligent transport system, IEEE Trans. Intell. Transp. Syst., 2021, 23(9): 16492–16503.
59. Peffers K. Tuunanen T., Rothenberger M.A., and Chatterjee S. A design science research methodology for information systems research, J. Manag. Inf. Syst., 2007, 24(3): 45–77.
60. Alotaibi F.M., Al-Dhaqm A., and Al-Otaibi Y.D. A Novel Forensic Readiness Framework Applicable to the Drone Forensics Field, Comput. Intell. Neurosci., 2022, 2022.
61. Al-Dhaqm A., Razak S., Othman S.H., Choo K.-K.R., Glisson W.B., Ali A., Abrar M. CDBFIP: Common database forensic investigation processes for Internet of Things, IEEE Access, 2017, 5: 24401–24416.