Securing Big Data-Based Smart Applications Using Blockchain Technology

Rihab Benaich, Imane El Alaoui, and Youssef Gahi

Abstract Nowadays, Big Data is the most salient paradigm. It has become a game-changer in the current technology and has aroused various industries and research communities worldwide. Big Data-based applications have shown much potential in data-driven decisions that have changed all business operations, including cost optimization, online reputation, and customer loyalty. Nevertheless, Big Data also brings many security and privacy issues. For this reason, Big Data applications require accurate methods to ensure reliable data storage, sharing, and decision-making. Among these methods, the Blockchain is known as the clever combination of distributed exchange, consensus, and cryptography mechanisms. It has been brought to the forefront to resolve security challenges in various fields such as Healthcare, banking, smart cities, etc. This paper first presents the considerable difficulties faced when adopting Big Data that can be resolved using Blockchain. Afterward, we overview the Blockchain technology by projecting its components, workflows, classification, and related characteristics. Finally, we present the importance of combining Big Data and Blockchain through reviewing the novel implementations proposed by researchers in great domains.

Keywords Blockchain · Big data · Security · Healthcare · Banking · Game theory · IoT · Smart cities · VANETS

R. Benaich (✉) · I. E. Alaoui · Y. Gahi
Laboratoire Des Sciences de L'Ingénieur, Ecole Nationale Des Sciences Appliquées, Ibn Tofail University, Kenitra, Morocco
e-mail: rihab.benaich@uit.ac.ma

I. E. Alaoui
e-mail: Imane.el.alaoui@uit.ac.ma

Y. Gahi
e-mail: gahi.youssef@uit.ac.ma

Y. Baddi et al. (eds.), *Big Data Intelligence for Smart Applications*,
Studies in Computational Intelligence 994,
https://doi.org/10.1007/978-3-030-87954-9_11

1 Introduction

With the increase of people using the Internet, the amount of produced data is practically exploding by the day. This explosion opens up incredible processing opportunities for organizations, such as data sharing and data-driven decisions (Faroukhi et al. 2020) and proposing customizable services. However, this noteworthy and extensive usage also comes with the necessity of ensuring the security and protection of data storage and transmissions (Gahi et al. 2016).

It is worth noting that this vast data, also known as Big data, is, most of the time, very sensitive (Gahi et al. 2016) as it represents a direct projection of our day-to-day activities. Big data is closely related to our personal life as it mainly relies on our transactions, interactions, and pattern mining. As Big data becomes a grand reality, it has to be clearly defined. Since the 2000s, many characteristics have been associated with Big data (El Alaoui et al. 2019), such as Volume, Variety, Velocity, Value, Veracity, Variability, and Visualization. Also known as 7Vs (El Alaoui and Gahi 2019), see Fig. 1.

Because of these varied characteristics, ensuring security using traditional methods becomes a complex task (Hu et al. 2014). For example, many conventional security techniques cannot handle the scale and velocity required by Big Data-based applications.

To overcome the scalability and velocity aspects, distributed approaches should be used. These approaches are designed to run over large clusters, which are a set of interconnected machines. Big Data systems' common distribution aspect makes Blockchain technology an ultimate solution for Big Data security challenges. It allows coping with security issues related to sharing and storing Big Data by applying cryptography, decentralization, consensus principles, and ensuring transaction trust.

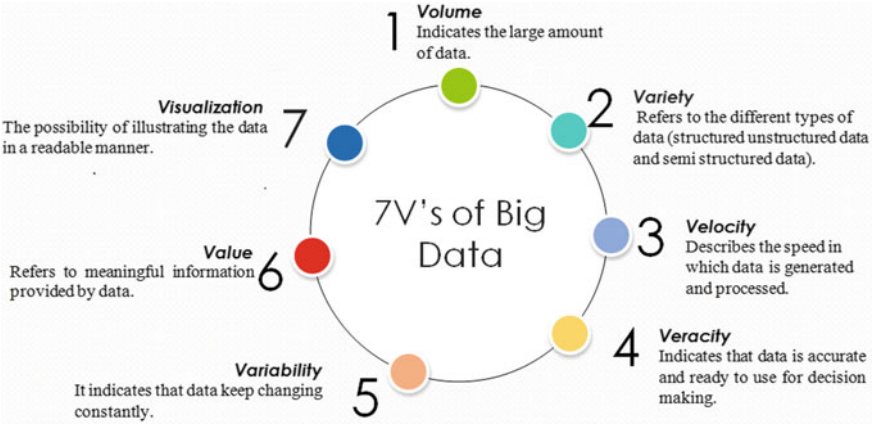

Fig. 1 The seven V's of big data

Moreover, today, Blockchain technology has become an ideal fit for Big Data problems through its high capacity to prevent the significant malicious activities caused by cybercriminals. The Blockchain not only restricts the intruders but also ensures access to multiple nodes in the network. This safeguarding feature of Blockchain strengthens and boosts the security level compared to many unique technologies used in various enterprises such as the cloud or virtual private network, or any other technology.

For this reason, many researchers have taken a keen interest in proposing attractive and diverse Blockchain models and frameworks to share and securely store Big Data in various sectors, namely Healthcare, Banking, IoT, gaming, and smart applications.

In this contribution, we focus on reviewing the recent studies that have tackled Big Data and Blockchain and their alliance. Furthermore, we discuss the advantage and the limits of the proposed models.

The following research questions will guide this contribution:

- What are the application domains of blockchain technology in the context of Big data?
- How does Blockchain technology address the challenge of Big data security?
- What are the significant challenges remaining in the usage of Blockchain as a secure solution for Big data intelligent applications?

This paper is organized as follows: in the second snd third sections, we present the research methodology, and we overview the concept of Blockchain technology by showing its components, workflows, and related characteristics. Afterward, in the fourth section, we highlight exciting studies based on Blockchain brought to various domains such as the healthcare field, smart cities, IoT, Vanets, 5G networks, Banking, and Game theory. Then in Sect. 5, we discuss the importance of using Blockchain in the context of Big Data. Finally, we conclude the paper and show some future works.

2 Research Methodology

To carry out this paper, we have used a systematic review methodology to identify and summarize the relevant research contributions that answer the above questions. According to (Tranfield et al. 2003) standards, we have conducted our review to develop a review plan regarding the methods to be used and the main research questions to be outlined and reports to be extracted.

This literature review highlighted the critical role of the outstanding Blockchain technology in coping with Big data challenges in numerous domains.

We took three stages to perform this study:

1. First of all, we researched and identified multiple primary papers related to our topic.
2. Secondly, we filtered the primary chosen papers.
3. Finally, we validated the filtered articles for data synthesis and analysis.

The primary studies were chosen based on the inclusion and exclusion criteria applied to the reviewed publications, the study's scope, and verifications done by subject and title and reading the abstract and findings.

This research method enables the investigation to be focused on papers published in renowned academic journals such as IEEE Xplore, Springer, Google Scholar, Wiley online library, ACM digital library, web of science, and so forth.

In our research approach, we covered various papers and magazines that were published in recent years. This allowed us to illustrate the progression of the thoughts. Therefore, we selected particular keywords such as "Big Data," "Blockchain," "Security," and "Blockchain applications" to assure a superior approach in literature research.

A considerable number of articles on the order of 300 were found through the selection of source research. A relevant and rigorous study of cross-referenced and redundant keyword combinations was used to refine this first bundle. A total of 180 articles were chosen, which were then subjected to a more in-depth study.

The filtered papers were carefully studied and assessed to ensure that they were relevant to the subject of our research. The last group consisted of 100 articles.

3 An Overview of Blockchain Technology

Blockchain is undoubtedly a brilliant invention, which has been introduced in 2008 by the pseudonym Satoshi Nakamoto (Nakamoto). It is a register where data is stored and kept by a decentralized system of computers. The Oxford dictionary is described as "A system in which a record of transactions made in Bitcoin or another cryptocurrency is maintained across several computers that are linked in a peer-to-peer network." This technology, based on trust, redefines our relationship to information and value transfers.

Blockchain architecture is composed of a set of functional components. The major ones are:

- **Nodes**: nodes are an essential component of Blockchain that allows access to data. They act as a communication point and perform multiple tasks such as ensuring the reliability of stored data (2021) in the Blockchain network.
- **Block**: it contains the block header and many transactions. The block header is hashed repeatedly to generate proof of work for mining rewards. Also, the block header serves to locate a specific block on an entire Blockchain.
- **Transactions**: defined as an exchange of assets, processed and managed under some specific rules of the entity service.
- **Smart contract**: is an autonomous agent unfalsifiable and permanent. Smart contracts are logical rules in coded text embedded in a file Blockchain to manage transactions.

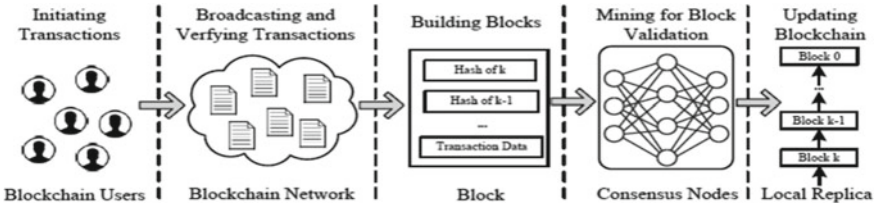

Fig. 2 Blockchain workflow

- **Merkle tree**: also known as a binary hash tree, it is a tree in which each leaf node is marked by the hash value of the block transaction data. Merkle trees are used to effectively and securely encode Blockchain data.

An overview of the Blockchain workflow is shown in Fig. 2.
The procedure shown in Fig. 2 works as follows:

- The first step consists of initiating and broadcasting a transaction to the distributed network via a node.
- After that, all Blockchain network nodes verify the transaction (the node that performs the broadcast).
- More than one node can bundle different subgroups of the newly verified transactions in their candidate constituencies and broadcast them over the entire network in the third step.
- In the fourth step, some nodes or the entire network nodes validate blocks by executing functions defined by consensus protocol.
- The last step consists of the attachment of the verified block to Blockchain, and finally, all nodes update their local replica.

The Blockchain can be classified into the following categories depending on its usage (Viriyasitavat and Hoonsopon 2019):

- **Public**: public Blockchains are decentralized and visible by anyone, and they don't have a specific owner, for example (Bitcoin, Ethereum).
- **Private or Permissioned**: they require some rules and privileges to manage and control who can read and write to the Blockchain.
- **Hybrid or Consortium**: these Blockchains are public only to a specific group. The consensus process is controlled by known, privileged servers using all parties' rules.

It is also important to mention that Blockchain technology is acknowledged with these main characteristics (Yaga et al. 2018):

- **Decentralized**: a key Blockchain feature. It means that Blockchain does not have to depend on location nodes anymore. Data can be recorded, stored, and updated for distribution.
- **Immutable**: every time a record of transactions is added, it's permanent and cannot be changed.

- **Consensus**: Trust verification, Blockchain provides some protocols such as Proof of Work, Proof of stake, and Byzantine Fault Tolerance to verify each block separately.
- **Transparent**: all transactions are visible to all the existing nodes in the network.
- **Anonymity and identity**: it is one of the major features of public Blockchain. It means that there's no need for a central organization to ensure privacy. Furthermore, a user can avoid exposure by obtaining different identities.
- **Tamper-proof ledger**: the usage of cryptography ensures the security of all transactions done in the network. They cannot be modified or altered until all nodes are compromised.

Combining these inherent attributes has made Blockchain a viable solution for Big data issues such as security. As a result, researchers have proposed attractive models based on big data and Blockchain to ensure the security of Big Data-based applications.

The following section presents this noteworthy healthcare, banking, smart cities, IoT, VANETs, and 5G.

4 Blockchain in the Service of Big Data

Researchers use Blockchain for different security aims in the context of Big Data. In this section, we first emphasize the security aims by reviewing recent researches in various domains. Then, we compare these studies by showing their benefits and limits.

4.1 Healthcare Field

Healthcare is a critical and essential domain that has attracted massive attention from researchers. As a result, several studies have proposed and enhanced Blockchain models to cope with several healthcare challenges in the context of Big Data.

Blockchain technology redefines the data modeling and management embedded in many health care systems (Liang et al. 2017b). This is mainly due to its adaptability and differentiation capabilities, security, and sharing of information and medical services unprecedentedly. One of the most crucial usages of Big Data and Blockchain is securing and sharing personal health data obtained from wearable devices such as smartwatches and activity trackers via the creation of a user-centric health data sharing based on decentralized and permissioned Blockchain. This solution offers data security via the deployment of a channel formation scheme. Furthermore, it ensures the enhancement of identity management via the membership service provided by Blockchain (Liang et al. 2017b). It also allows the prevention of any alteration of the patients' health data through applying a proof of integrity and validation

that is continuously reachable from a cloud database then linked to the Blockchain network. Another potential of Blockchain that should not be overlooked is its capability to surmount vulnerabilities in the old sharing methods, such as the inappropriate access of data and the integration of malicious programs.

Also, Blockchain is used in pharmacy (Clauson et al. 2018) and helps to control and secure the medication return process. In this regard, the Blockchain has shown its potential toward the product (medicines) tracing by allowing manufacturers, distributors, and pharmacists to submit tracing information in a shared ledger, with automated verification of critical data. Besides, Blockchain has demonstrated a potential toward detection and response and enables public and private actors to report and identify medicines suspected of imitation, illegal, or dangerous. Therefore, the safety of medications increases without disclosing sensitive information.

It is also important to mention that the interest in the Blockchain has drastically increased during the pandemic of COVID-19 (Abd-alrazaq et al. 2021). According to the European Parliamentary Research Service, the Blockchain was considered among the ten powerful technologies which have reduced COVID-19 challenges (Mihalis Kritikos 2020). It is used to track public health information, particularly outbreaks of infectious diseases. Combining both Blockchain and Big Data enables real-time data processing, further preventing the spread of the disease to epidemic levels. It also allows obtaining more accurate reporting and effective responses. However, these effective responses are insufficient without safeguarding the patient's information, which requires a high level of security. Hence, by health information exchange based on the Blockchain, the privacy of the patient is ensured. The aspect of securing each patient's medical data is obtained by deploying an electronic medical record Blockchain that uses on and off-chain storage verification records (Ahir et al. 2020).

4.2 Banking Field

Blockchain is widely used in the banking sector and has many advantages in the Big Data context. This innovation reduces the risk of fraud and prevents some potential scams in all banking environments, such as financial and online markets (Hassani et al. 2018) via monitoring and recording every change of the data within the blocks for every transaction in real-time. Furthermore, blockchain technology is considered a trusted network for the digital banking sector due to its encryption capabilities and public key infrastructure. Additionally, the distributed ledger "Blockchain" offers another aspect that should not be neglected, relying on the difficulty of operating the newly added data by various units in real-time. This difficulty results in enhanced privacy of the massive data transiting.

Blockchain also addresses various issues such as operational risk and administrative costs. Furthermore, integrating Blockchain and Big Data into the financial sector optimizes multiple processes, such as creating new financial services and ensuring the integration of the performance and profitability of actors (microcredit, micro-payment transactions almost free of charge, and so forth. These benefits

could be obtained through using a cost-cutting technique, "transaction commitment," that drastically decreases transaction time and storage for small amounts of money (Rezaeibagha and Mu 2019).

Furthermore, Blockchain and Big Data allow enhancing the security of banking transactions highly. They would generate new revenue models, capacity gains, cost reduction of millions, and significant losses across the industry. It is also worth noting that many analysts predict positive growth of the level of trust of customers due to the combination of Blockchain and Big Data. This combination will change customers and banks, insurers, and other financial institutions by identifying suspicious transactions by tracking customer transactions and activities in real-time.

4.3 Smart Applications

In recent years, the notion of "smart city" has emerged as a new paradigm to enhance citizen's daily life, by providing personalized and adapted services. This paradigm is based on personal data that are closely related to our daily basis. However, they are several challenges related to security and storage in smart city systems. For this reason, the idea of securing the data aroused the interest of many researchers.

Many studies have tackled the smart city field by proposing new methods and practices. The major one is the combination of Blockchain and Big Data. This combination provides secure communication between physical devices in a heterogeneous environment through real-time data monitoring. It also provides the quality of service (QoS) by minimizing traffic rate fluctuations and the diversity of new devices (Alam). Furthermore, Blockchains can be used to create a smart city system that enables devices to securely and reliably transfer currency and data between all smart city devices.

Also, the combination of Big Data and Blockchain is widely used in smart transportation, especially in the automotive industry. This combination has revolutionized intelligent transportation systems (ITS) by building safe, dependant and self-sufficient ITS ecosystems to provide services such as remote software-based vehicle operation (Deepa et al. 2021).In the same context of smart transportation security, it is noteworthy that (Wang and Qu 2019) have created a secure critical management architecture based on Blockchain technology to provide network security. This solution is based on using security managers to record vehicle departure data, encapsulate the blocks to transfer keys, and then deploy rekeying to the cars within the secured domain. This framework provides a key management system for critical transfers among security administrators in a heterogeneous vehicle network. Other researchers have proposed an interesting system that ensure the security of vehicles using Blockchain (Hîrţan et al. 2020). This system is based on an offline Blockchain storage system. All confidential data gathered from users is kept and then shared with the help of unique encryption keys applicable to a specific vehicle cluster. The

solution comprises two applications, the first one is installed on the client's smartphone, and the second is installed on the server. It has shown the capacity to ensure privacy policies sent to the client's application to offer the precise transit routes.

Furthermore, the combination of Big Data and Blockchain allows managing the security issues such as protocol vulnerabilities, privacy, eavesdropping, and attacks on internet-connected devices and vehicle communication networks (Ismagilova et al. 2020).

4.4 Game Theory

Today, most gaming platforms are hosted on centralized servers, and transactions are frequently conducted on mobile phones or desktop computers that lack adequate security (Jacob 2020). In fact, because of this centralization aspect, servers could be easily hacked. Furthermore, this lack of server security could lead to enormous damages such as data loss or interruption of services. Therefore, many studies have tackled the combination of Blockchain and Big Data in gaming to cover these risks. This combination has brought the potential to change the vulnerability of centralized servers. Hackers will be unable to damage a decentralized Blockchain network since (i) there is no server to destroy, (ii) nodes share the maintenance of distributed databases, (iii) each node has complete information in the database. As a result, players can safely store digital collectibles purchased in Blockchain-based games in their crypto wallets (Aran Davies 2020).

Furthermore, due to the decentralization aspect of Blockchain, trust is established between all nodes; players are unable to modify the data. This means that trust is built between all players in the industry, from developers to actors (Gainsbury and Blaszczynski 2017). Also, the Blockchain permit gamers to make their payments in a secure way using digital crypto coins.

Besides the security aspect, the Blockchain also offers new features to the gaming world, such as digital trade assets between players (Casino et al. 2019). The use of intermediary sites obtains this ability of digital trading. In addition, to provide user-friendly functions and interfaces for running shared data (Robin8 2019).

Another aspect of Blockchain technology resides in solving data instability in peer-to-peer games by providing data authentication and permanent storage via a new proof-of-play consensus methodology underpins this serverless turn-based strategy game (Wu et al. 2020). This solution is based on three key steps: matchmaking, gaming session, and gameplay.

4.5 Internet of Things

The Internet of things (IoT) refers to the billions of physical devices deployed worldwide connected to the Internet (Atzori et al. 2010). These devices collect and

share massive data. Thus, they allow conducting data analysis for better and faster decision-making. However, there are many security issues, such as data leaks and vulnerabilities concerning physical equipment; hackers can easily alter that.

For this reason, Blockchain was introduced to ensure privacy and to face challenges related to IoT security. Using Blockchain as the basis of devices reduces the possibility of hacking by decreasing malicious programs and spyware (Banerjee et al. 2018). In this context, the Blockchain was introduced to trace the history of the firmware. If a corrupted firmware is identified, it will be driven to revert to its prior version. Also, the Blockchain was brought to the front line to maintain the Refence Integrity Metric (RIM) of the datasets by storing membership informations such as the address, owner, and sharing policies.

Furthermore, Blockchain technology allows protecting IoT networks against potential entry points of intruders without the need of a central authority. Therefore, Blockchain is considered a leading solution for organizations involved in auditing, tracing a supply chain, or securing connected streetlights in intelligent cities due to the encryption capability of Blockchain, its dispersed storage, and its irrefutable, tamper-proof record (Plummer and Writer 2018).

Besides the security aspect, Blockchain creates a marketplace that allows data providers (IoT sensors) and data consumers (IoT application owners) to trade in real-time via smart contracts and makes value from collected data (Jain 2021).

Another breakthrough of blockchain use in IoT is embodied in a reliable and flexible solution designed for IoT services, specifically the adoption of drones in various IoT scenarios, namely agriculture, delivery, and so forth. In fact, (Liang et al. 2017b) have proposed a solution that consists of using a public Blockchain and the standard cloud server in which the hashed data records collected from drones are anchored to the Blockchain network. After that, a Blockchain receipt for each data record stored in the cloud is generated to reduce the burden of moving drones with limited battery and processing capability while gaining improved data security assurance. In addition, this usage provides trust, data capability integrity audit, data convenience, and scalability.

4.6 Big Data and Blockchain in the Service of VANETs

VANETs are one of the most brilliant inventions of the century. However, they still include many challenges that could be resolved with Blockchain. Such as data management and security in a vehicular environment. For example, the usage of Blockchain inVANETs enables essential functionalities such as ensuring traffic safety and parking space management through promoting decentralized system management (Peng et al. 2020).

Furthermore, Blockchain played a pivotal role in resolving critical message dissemination issues in VANETs via storing the history of vehicle trust level in Blockchain and event messages (Shrestha et al. 2020). Thus, when a car experiences an event, such as an accident, the event message with various parameters is

broadcasted to nearby vehicles in the Blockchain network. Also, Blockchain allows the preservation of the circulation of untrustworthy information sent by malicious vehicles. In addition to these advantages, using Blockchain and Big Data's combination as a security feature of the VANET network allows strengthening network security and facilitating data transmission. Lu et al. (2018) have implemented a Blockchain-based anonymous reputation system that builds a trust paradigm for VANETs. This paradigm enables the transparency of certificates and revocations to be efficiently accomplished through proofs of presence and absence based on enhanced Blockchain technology and conditional anonymity via public keys. This level of protection is guaranteed via a trust-based Blockchain design that effectively mitigates multiple network attacks such as (DoS attack and Sybil attack) (Khan et al. 2019). These attacks become almost non-existent since the potential of altering centralized data storage has become remarkably decreased through the application of consortium Blockchain-based data sharing for VANETS (Zhang and Chen 2019).In other words, given the tamper resistance property of Blockchain, the operation of data sharing become more controlled. The deployment of consortium Blockchain resides in sending each vehicle data to the nearest roadside units (RSU). Then, the RSUs function as pre-selected nodes that construct blocks based on data from the connected vehicles. At this stage, the Blockchain that contains the vehicle data is created by obtaining consensus among RSUs. Consequently, the RSUs control the data sharing process via intelligent contract technology provided by the distributed ledger technology.

However, despite the critical role of Blockchain in VANETs, research is still lacking, mainly concerning real-time and scalability. Moreover, the mobility of VANETs creates difficulty in proof of work in the Blockchain because of the continuous dynamicity of Blockchain nodes (Kim 2019).

4.7 The Fifth Generation Based Applications (5G)

The novel features of 5G allow supporting new business models and services, including mobile operators, businesses, call providers, government executives, and infrastructure providers. However, they also face many challenges, such as data mobility and network privacy (Nguyen et al. 2019).

For this reason, several studies have proposed various Blockchain models to address 5G challenges in the context of Big Data. Moreover, blockchain characteristics offer promising solutions to 5G's challenges. One of these essential Blockchain features is decentralization, which allows for banishing external authorities' requirements in the 5G ecosystem.

In other words, decentralization of 5G networks enables avoiding single-point failures and ensuring data availability (Nguyen et al. 2019). It is also important to note that Blockchain offers high security for 5G networks due to smart contracts that support the services of 5G, such as preservation of 5G resources against alteration and data authentification (Nguyen et al. 2019).

Furthermore, Blockchain provides immutability to the 5G network performs multiple tasks such as (usage information for billing, resource utilization, and trend analysis).

Besides the security and immutability aspects, the Blockchain allows to ensures transparency of the 5G services, enabling service providers and users to fully access, verify, and monitor transaction activities over the network with equal rights (Nguyen et al. 2019). Furthermore, as mentioned above, Blockchain reveals a sequence of security features that remove the need for centralized network infrastructure or third-party authorities and decrease the single point failure. Furthermore, the security of device-to-device (D2D) communication may be achieved by constructing a peer-to-peer network using Blockchain. This converts each D2D device as a Blockchain node to maintain a ledge replica of validating and monitoring transactions to improve the system's transparency and dependability (Nguyen et al. 2019).

All the above studies have designed exciting solutions based on Blockchain and Big data to face many fields. The table below summarizes these solutions by presenting the significant covered challenges and showing their limits and advantages (Table 1).

5 Discussion

Through this paper, we highlighted how security challenges related to Big Data could be resolved with Blockchain. Researchers have proposed Blockchain models and frameworks to face many security challenges in diverse Healthcare, banking, IoT, and so forth.

By reviewing these studies, we have noticed that Blockchain played a pivotal role in securing and preserving data privacy through its decentralization, tamper-proof, and immutability attributes. Moreover, far beyond, the combination of Big Data and Blockchain has assured the security of the data by creating partitioning models based on smart contracts. Accordingly, the data sharing models are generated without the need for a reliable third party.

Besides the security aspect provided by Blockchain, many outstanding advantages are emerging in the implementation of Blockchain claimed by several studies. For instance, transparency and a remarkable reduction of the costs of data storage. This significant minimization of the cost is achieved via the Blockchain storage capacity, which allows storing a large amount of data for long periods instead of using shared data holding platforms, which require many resources. In this process, Blockchain relies on the automatic performance of transactions that are supervised by smart contracts.

Moreover, due to the transparency and decentralization aspects of Blockchain technology, users can look through the record of all transactions.

Despite the advances brought by the usage of both Blockchain and Big Data. A lot of issues still unreliable and require more research. Scalability is an example of these problems; it is known as a critical requirement of Big Data. Companies and businesses

Table 1 Summary of blockchain and big data applications

Field	Reference	The faced challenges	Used methods	Advantages	Limits
Healthcare	Liang et al. (2017b)	The privacy issues and vulnerabilities existing in personal health data and storing system	Implementing an access control scheme by utilizing the Hyperledger Fabric membership service component, and a tree-based data processing	– Preservation of the integrity of health data within each record – Protection and Validation of personal health data	Difficulty in combining both personal health data and medical data
	Wang et al. (2018)	The challenge of sharing medical data	A medical data sharing platform based on permissioned Blockchains Smart contracts - Concurrent Byzantine Fault Tolerance-Consensus mechanism	– The use of encryption technology ensures that users have complete autonomy in their medical data – Smart contract technology is used to enable users to set up different access permissions of medical data	– High cost of centralized data storage – Lack of security and protection of personal privacy

(continued)

Table 1 (continued)

Field	Reference	The faced challenges	Used methods	Advantages	Limits
	Ahir et al. (2020)	Virus detection during a pandemic	– Mathematical modeling – The QR Code system – Artificial intelligence	– Deep learning to detect the symptoms of the virus – AI signals robots to keep social distancing – Blockchain technology to maintain patient records – Big Data to detect the spread of the virus	The high cost of the material
	Clauson et al. (2018)	The challenge of managing supply chain complexity in the healthcare field	Recognizing stakeholders engaged in finding solutions using Blockchain for the health supply chain	– Preserving integrity of the health supply chain – Medical devices become secured – Increasing the function of the Internet of health things(IoHT)	– Proof of concept – Requirement of a pilot phase

(continued)

Table 1 (continued)

Field	Reference	The faced challenges	Used methods	Advantages	Limits
Banking	Hassani et al. (2018)	Expanding the research and development of Blockchain in the banking field	– Blockchain-based KYC (know your customer) process – Smart contracts – Filtering technique and signal extraction	– The use of Blockchain prevents the piracy of historical information of banking data	– The high cost of developing a Blockchain-enabled system – Problem of currency stability
	Bandara et al. (2018)	The challenge of managing high transactions in the current public Blockchains	– Each node in the Blockchain runs its own Cassandra node – All services are built as docker containers and deployed via Kubernetes	– High scalability, high availability, and full-text search features It makes Big Data more secure, structured and allows data analytics to be more easily performed on big data – Support to store large data payloads with Cassandra	The challenge to introduce the designed application for an online 3D printing marketplace
Smart field	Biswas and Muthukkumarasamy (2016)	Overcoming security challenges in smart cities	They are integrating blockchain technology with smart devices	– Improved reliability – Better fault tolerance capability – Faster and efficient operation – Scalability	Interoperability issue

(continued)

Table 1 (continued)

Field	Reference	The faced challenges	Used methods	Advantages	Limits
	Liang et al. (2017a)	Securing drone communications during data collection and transmission	Compiling hash data records collected from drones to Blockchain network then creating a Blockchain receipt Containing each data recorded is stored in the cloud	– Enhanced security of the data – Provisioning data integrity and cloud auditing	Nodes don't require any permission to participate
	Xu et al. (2018)	The lack of cooperation between edge devices to share voluminous data	– Consensus mechanism none as Proof-of-Collaboration (PoC) – Filter algorithm for transaction offloading (FTF)	– Reducing storage resources occupied by the Blockchain	Framework in a green and efficient manner is still an open issue
	Ismagilova et al. (2020)	Security in smart cities	– Fog computing characteristics – Protocol of privacy-preserving authentication (PPA) – Fully Privacy-Preserving and Revocable Identity-BasedBroadcastEncryption (FPPRIB) – Piracy Zones	– Assuring data privacy – The content of the data is not revealed – The increased privacy will encourage the adoption of smart cities	Smart cities require the employment of many emergent technologies (IoT, sensors, GPS) remain notable threats related to security
Game theory	School of Computer Science and Electronic Engineering, University of Essex, U.K et Dey (2018)	Securing attacks in Blockchain using game theory	Using intelligent software agents to monitor the activity of stakeholders in the Blockchain networks to detect any sort of attack	Prevention of attacks and risk of alteration	The need for the application layer of the network to make a decision

(continued)

Table 1 (continued)

Field	Reference	The faced challenges	Used methods	Advantages	Limits
	Gainsbury and Blaszczynski (2017)	The necessity of a third party to prevent frauds to ensure the security of the game	– Using cryptocurrencies – Smart contracts – Blockchain-based domain name system	– Players can operate outside of regulatory jurisdictions – Transactions are verified	The issue of classification of Bitcoin as currency, money, or an item
	Wu et al. (2020)	The vulnerability remaining in a single cluster	– Proof-of-Play consensus model – Proof-of-work – Proof of stake – P2P turn-based strategy game	– The joining players have equal roles – Global synchronization – Offers gaming sessions	The difficulty to manage a large number of players
Internet of things	(Alam)	Introducing a new Blockchain architecture with Big data analytics to increase connectivity performance throughout the smart cities	The Integration system for smart objects using cloud and Blockchains on the IoT (IoT nodes, P2P network)	– The physical devices are allowed to communicate securely with other physical devices in heterogeneous environments – Big Data analytics perform an increasingly significant role in strategic planning	– Discovering the IoT nodes is a challenge across all smart devices – The privacy issue – Scalability issues

(continued)

Table 1 (continued)

Field	Reference	The faced challenges	Used methods	Advantages	Limits
	Dorri et al. (2017)	Providing security for IoT devices in a smart Home	BC-based smart home framework	Protection against DDOS and linking attacks	– Low scalability – Low latency and high requirement of resources
	Liang et al. (2017a)	The challenge of ensuring data resilience in cloud-based IoT applications	The Blockchain collects data from drones and commands from control systems – Cloud server – Cloud database	– Ensuring the security of drones – the solution offers the chance to store a large amount of data	– Scalability issue – The cloud operating systems are vulnerable
Vehicular ad hoc networks (Vanets)	Shrestha et al. (2020)	The problem of security in traditional VANETs	Building a local Blockchain for exchanging real-world event messages across vehicles within a country's borders	– The messages of vanets become secured – Vehicles ensure a secured and distributed database	– Scalability issues – Storage and message overhead
	Lu et al. (2018)	Preserving vehicles from attacks	– Using a privacy-preserving trust model forVANETs known as blockchain-based anonymous reputation system (BARS) – Public keys	– Prevention of vehicles from fraudulent messages – The solution provides an efficient and robust trust model for Vanets	Scalability (the solution can't support a large number of vehicles)

(continued)

Table 1 (continued)

Field	Reference	The faced challenges	Used methods	Advantages	Limits
	Tan and Chung (2020)	The problem of data interference and the lack of security caused by the large group of vehicles in heterogeneous VANETS	– Securing authentification – Key management – Edge computing infrastructure – Using Consortium Blockchain	– Real-time arrangement – Security and resistance against attacks	Cost of computation and communication
	Iqbal et al. (2020)	The traditional cloud data centers are inadequate for vehicles. A tremendous amount of bandwidth is consumed, which negatively influences delay-sensitive applications	– Edge computing – Fog computing – Using distributed ledger-based decentralized trust management scheme to solve the offloading issue in VANETs – Consensus-based on techniques such as Byzantine fault tolerance and Raft	– Blockchain-based social reputation framework at roadside unites permits the decision model to choose from a pool of trusted vehicles for any arriving assignments. As a result, the level of privacy expands	Overloaded fog resources
The fifth generation (5G)	Chaer et al. (2019)	How can Blockchain offers opportunities for the fifth generation(5G)	– Smart contract with the Service Level Agreements (SLA) – Home server subscriber (HSS) – Dynamic Spectrum Sharing Decentralized application (DApp)-DLT	– Network slicing – 5G infrastructure sharing – International roaming – 5G infrastructure crowdsourcing	– Scalability issues – Interoperability issues – Smart contracts – Data privacy complexity – Standardization and Regulations – Transaction and Cloud Infrastructure Costs

(continued)

Table 1 (continued)

Field	Reference	The faced challenges	Used methods	Advantages	Limits
	Nguyen et al. (2019)	Challenges in 5G network (risk of data interoperability)	Using the Blockchain Framework to store immutable ledgers	– Customized and advanced user-centric value – D2D communication – Software-defined networking(SDN)	– Scalability – latency problem

cannot efficiently handle and process the exploding volume of data without using data scaling techniques. Moreover, the lack of standardization is another drawback of combining Blockchain and Big Data. In other words, standardization denotes a minimal degree of interoperability, such as (different Blockchain networks and different consensus models).

The interoperability problem leads to increased resources used in transactions generated by users in various blockchain networks. Also, without standardization, sharing data and value between participants in divergent Blockchain networks becomes complex or almost impossible. Furthermore, another limit of the adoption of Blockchain in the service of Big Data is redundancy. This limit highlights the fundamental issue of decentralization attributes brought by Blockchain. In other words, redundancy means that every node in the Blockchain network must traverse and process each intermediate node independently (Mudrakola 2018) to target the objective node in the network. As a result, the redundancy inherent in Blockchain technology has an impact on its efficiency.

Additionally, Blockchain faces another limit related to signature verification. The employment of this signature verification in Blockchain technology means that each transaction made in the Blockchain network should be verified and signed digitally using private or public keys. Consequently, the computation of this signature verification process becomes highly complex and takes a long time. This limit hampers the alliance of Blockchain with Big Data.

6 Conclusion

The technology of Blockchain is considered a hot topic given its robust features and good advantages. Today, Blockchain's importance has extended far beyond the financial sector and has shown notable improvements, especially toward security concerns. On the other side, Big Data must be handled securely and accurately to avoid any interruption or loss of the data. Therefore, Blockchain has been brought to the front line in Big Data contexts. In this paper, we have presented the alliance of Blockchain with Big Data to cope with Big Data storing and sharing security issues. For this, we have overviewed the multiple advantages and limits of different Blockchain researches that have tackled Big data issues in many sectors ranging from health and technology (IoT, 5G, VANETs, Smart field) to entertainment concerns.

Despite the advancement brought by Blockchain technology, there are still plenty of challenging issues concerning the use of Blockchain in the context of big data, such as interoperability and scalability. However, distributed ledger technology is still maturing. Moreover, it can be linked to new-age technologies, such as artificial intelligence and IoT, to build platforms and infrastructures that ensure advanced data privacy and security.

As future work, we propose a Blockchain model that copes with unsolved Big data security issues such as scalability and standardization.

References

A.A. Abd-alrazaq, M. Alajlani, D. Alhuwail, A. Erbad, A. Giannicchi, Z. Shah, M. Hamdi, M. Househ, Blockchain technologies to mitigate COVID-19 challenges: a scoping review. Comput. Methods Programs Biomed. Updat. **1** (2021). https://doi.org/10.1016/j.cmpbup.2020.100001

S. Ahir, D. Telavane, R. Thomas, The impact of artificial intelligence, blockchain, big data and evolving technologies in coronavirus disease—2019 (COVID-19) curtailment, in *2020 International Conference on Smart Electronics and Communication (ICOSEC).* (IEEE, Trichy, India, 2020), pp. 113–120

T. Alam, Blockchain-based big data analytics approach for smart cities. Kansai University **62**(9), 17

A. Davies, How Blockchain Could Redefine the Gaming Industry, in DevTeam. Space (2020). https://www.devteam.space/blog/how-blockchain-could-redefine-the-gaming-industry/. Accessed 18 Mar 2021

L. Atzori, A. Iera, G. Morabito, The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010). https://doi.org/10.1016/j.comnet.2010.05.010

E. Bandara, W.K. Ng, K. De Zoysa, N. Fernando, S. Tharaka, P. Maurakirinathan, N. Jayasuriya, Mystiko—blockchain meets big data, in *2018 IEEE International Conference on Big Data (Big Data).* (IEEE, Seattle, WA, USA, 2018), pp. 3024–3032

M. Banerjee, J. Lee, K.-K.R. Choo, A blockchain future for Internet of things security: a position paper. Digit. Commun. Netw. **4**(3), 149–160 (2018). https://doi.org/10.1016/j.dcan.2017.10.006

K. Biswas, V. Muthukkumarasamy, Securing smart cities using blockchain technology, in *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS).* (IEEE, Sydney, Australia, 2016), pp. 1392–1393

F. Casino, T.K. Dasaklis, C. Patsakis, A systematic literature review of blockchain-based applications: current status, classification and open issues. Telemat. Inform. **36**, 55–81 (2019). https://doi.org/10.1016/j.tele.2018.11.006

A. Chaer, K. Salah, C. Lima, P.P. Ray, T. Sheltami, Blockchain for 5G: Opportunities and Challenges, in *2019 IEEE Globecom Workshops (GC Wkshps).* (IEEE, Waikoloa, HI, USA, 2019), pp. 1–6

K.A. Clauson, E.A. Breeden, C. Davidson, T.K. Mackey, Leveraging blockchain technology to enhance supply chain management in healthcare: an exploration of challenges and opportunities in the health supply chain. BHTY (2018). https://doi.org/10.30953/bhty.v1.20

N. Deepa, Q.-V. Pham, D.C. Nguyen, S. Bhattacharya, B. Prabadevi, T.R. Gadekallu, P.K.R. Maddikunta, F. Fang, P.N. Pathirana, A survey on blockchain for big data: approaches, opportunities, and future directions (2021). arXiv:200900858 [cs]

A. Dorri, S.S. Kanhere, R. Jurdak, P. Gauravaram, Blockchain for IoT security and privacy: the case study of a smart home, in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops).* (IEEE, Kona, HI, 2017), pp. 618–623

I. El Alaoui, Y. Gahi, The impact of big data quality on sentiment analysis approaches. Procedia Comput. Sci. **160**, 803–810 (2019). https://doi.org/10.1016/j.procs.2019.11.007

I. El Alaoui, Y. Gahi, R. Messoussi, Big data quality metrics for sentiment analysis approaches, in *Proceedings of the 2019 International Conference on Big Data Engineering* (Association for Computing Machinery, New York, NY, USA, 2019), pp. 36–43

A.Z. Faroukhi, I. El Alaoui, Y. Gahi, A. Amine, Big data monetization throughout big data value chain: a comprehensive review. J. Big Data **7**(1), 3 (2020). https://doi.org/10.1186/s40537-019-0281-5

Y. Gahi, M. Guennoun, H.T. Mouftah, Big data analytics: security and privacy challenges, in *2016 IEEE Symposium on Computers and Communication (ISCC).* (IEEE, Messina, Italy, 2016), pp. 952–957

S.M. Gainsbury, A. Blaszczynski, How blockchain and cryptocurrency technology could revolutionize online gambling. Gaming Law Rev. **21**(7), 482–492 (2017). https://doi.org/10.1089/glr2.2017.2174

H. Hassani, X. Huang, E. Silva, Banking with blockchain-ed big data. J. Manag. Anal. **5**(4), 256–275 (2018). https://doi.org/10.1080/23270012.2018.1528900

L.-A. Hîrţan, C. Dobre, H. González-Vélez, Blockchain-based reputation for intelligent transportation systems. Sensors **20**(3), 791 (2020). https://doi.org/10.3390/s20030791

H. Hu, Y. Wen, T.-S. Chua, X. Li, Toward scalable systems for big data analytics: a technology tutorial. IEEE Access **2**, 652–687 (2014). https://doi.org/10.1109/ACCESS.2014.2332453

S. Iqbal, A. Malik, A.U. Rahman, A. Waqar, Blockchain-based reputation management for task offloading in micro-level vehicular fog network. IEEE Access **8**, 52968–52980 (2020). https://doi.org/10.1109/ACCESS.2020.2979248

E. Ismagilova, L. Hughes, N.P. Rana, Y.K. Dwivedi, Security, privacy and risks within smart cities: literature review and development of a smart city interaction framework. Inf. Syst. Front. (2020). https://doi.org/10.1007/s10796-020-10044-1

N. Jacob, How blockchain is making digital gaming better, in *Blockchain Pulse: IBM Blockchain Blog* (2020). https://www.ibm.com/blogs/blockchain/2020/02/how-blockchain-is-making-digital-gaming-better/. Accessed 22 Mar 2021

S. Jain, Can blockchain accelerate internet of things (IoT) adoption, in *Deloitte Switzerland* (2021). https://www2.deloitte.com/ch/en/pages/innovation/articles/blockchain-accelerate-iot-adoption.html. Accessed 25 Mar 2021

A.S. Khan, K. Balan, Y. Javed, S. Tarmizi, J. Abdullah, Secure trust-based blockchain architecture to prevent attacks in VANET. Sensors **19**(22), 4954 (2019). https://doi.org/10.3390/s19224954

S. Kim, Impacts of mobility on performance of blockchain in VANET. IEEE Access **7**, 68646–68655 (2019). https://doi.org/10.1109/ACCESS.2019.2918411

X. Liang, J. Zhao, S. Shetty, D. Li, Towards data assurance and resilience in IoT using Blockchain, in *MILCOM 2017—2017 IEEE Military Communications Conference (MILCOM)*. (IEEE, Baltimore, MD, 2017a) pp 261–266

X. Liang, J. Zhao, S. Shetty, J. Liu, D. Li, Integrating Blockchain for data sharing and collaboration in mobile healthcare applications, in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. (IEEE, Montreal, QC, 2017b), pp. 1–5

Z. Lu, W. Liu, Q. Wang, G. Qu, Z. Liu, A privacy-preserving trust model based on blockchain for VANETs. IEEE Access **6**, 45655–45664 (2018). https://doi.org/10.1109/ACCESS.2018.2864189

M. Kritikos, Ten technologies to fight coronavirus **28** (2020)

S. Mudrakola, Blockchain limitations: this revolutionary technology isn't perfect—and here's why, in *TechGenix* (2018). https://techgenix.com/blockchain-limitations/. Accessed 28 Mar 2021

S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system **9**

D.C. Nguyen, P.N. Pathirana, M. Ding, A. Seneviratne, Blockchain for 5G and beyond networks: a state of the art survey (2019). arXiv:191205062 [cs, eess, math]

C. Peng, C. Wu, L. Gao, J. Zhang, K.-L. Alvin Yau, Y. Ji, Blockchain for vehicular internet of things: recent advances and open issues. Sensors (Basel) **20**(18) (2020). https://doi.org/10.3390/s20185079

L. Plummer, T. Writer, Blockchain, IoT & security, in *Intel* (2018). Accessed 21 Mar 2021

F. Rezaeibagha, Y. Mu, Efficient micropayment of cryptocurrency from blockchains. Comput. J. **62**(4), 507–517 (2019). https://doi.org/10.1093/comjnl/bxy105

Robin8, Why is trading digital assets via Blockchain the best solution? in *Medium* (2019). https://medium.com/@robin8/why-is-trading-digital-assets-via-blockchain-the-best-solution-6897db75faff. Accessed 22 Mar 2021

School of Computer Science and Electronic Engineering, University of Essex, U.K., Dey S, A proof of work: securing majority-attack in blockchain using machine learning and algorithmic game theory. IJWMT **8**(5), 1–9 (2018). https://doi.org/10.5815/ijwmt.2018.05.01

R. Shrestha, R. Bajracharya, A.P. Shrestha, S.Y. Nam, A new type of blockchain for secure message exchange in VANET. Digit. Commun. Netw. **6**(2), 177–186 (2020). https://doi.org/10.1016/j.dcan.2019.04.003

H. Tan, I. Chung, Secure authentication and key management with blockchain in VANETs. IEEE Access **8**, 2482–2498 (2020). https://doi.org/10.1109/ACCESS.2019.2962387

D. Tranfield, D. Denyer, P. Smart, Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. **14**(3), 207–222 (2003). https://doi.org/10.1111/1467-8551.00375

W. Viriyasitavat, D. Hoonsopon, Blockchain characteristics and consensus in modern business processes. J. Ind. Inf. Integr. **13**, 32–39 (2019). https://doi.org/10.1016/j.jii.2018.07.004

R. Wang, W.-T. Tsai, J. He, C. Liu, Q. Li, E. Deng, A medical data sharing platform based on permissioned blockchains, in *Proceedings of the 2018 International Conference on Blockchain Technology and Application—ICBTA 2018*. (ACM Press, Xi'an, China, 2018), pp. 12–16

S. Wang, X. Qu, Blockchain applications in shipping, transportation, logistics, and supply chain (2019), pp. 225–231

F. Wu, H.Y. Yuen, H.C.B. Chan, V.C.M. Leung, W. Cai, Infinity battle: a glance at how blockchain techniques serve in a serverless gaming system, in *Proceedings of the 28th ACM International Conference on Multimedia*. (ACM, Seattle WA USA, 2020), pp. 4559–4561

C. Xu et al. Making big data open in edges: a resource-efficient blockchain-based approach. IEEE Trans. Parallel Distrib. Syst. **30**(4), 870–882 (2018)

D. Yaga, P. Mell, N. Roby, K. Scarfone, Blockchain technology overview. (National Institute of Standards and Technology, Gaithersburg, MD, 2018)

X. Zhang, X. Chen, Data security sharing and storage based on a consortium blockchain in a vehicular ad-hoc network. IEEE Access **7**, 58241–58254 (2019). https://doi.org/10.1109/ACCESS.2018.2890736