IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 April 2023, accepted 29 May 2023, date of publication 2 June 2023, date of current version 13 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3282322

RESEARCH ARTICLE

A Secure Big Data Storage Framework Based on Blockchain Consensus Mechanism With Flexible Finality

A. SASIKUMAR[1], LOGESH RAVI[2], KETAN KOTECHA[3,4], AJITH ABRAHAM[5,6], (Senior Member, IEEE), MALATHI DEVARAJAN[7], AND SUBRAMANIYASWAMY VAIRAVASUNDARAM[8]

[1]Department of Data Science and Business Systems, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India
[2]Centre for Advanced Data Science, Vellore Institute of Technology, Chennai 600127, India
[3]Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune 412115, India
[4]UCSI University, Kuala Lumpur 56000, Malaysia
[5]Faculty of Computing and Data Sciences, FLAME University, Lavale, Pune, Maharashtra 412115, India
[6]Innopolis University, Innopolis, 420500 Republic of Tatarstan, Russia
[7]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India
[8]School of Computing, SASTRA Deemed University, Thanjavur 613401, India

Corresponding authors: Ketan Kotecha (head@scaai.siu.edu.in) and Subramaniyaswamy Vairavasundaram (vsubramaniyaswamy@gmail.com)

This work was supported by the Analytical Center for the Government of the Russian Federation, in 1 November 2021, under Grant 70-2021-00143 and Grant IGK 000000D730321P5Q0002.

ABSTRACT Data security and integrity are becoming increasingly important as the volume of data being created and stored grows. A controlled third party that provides most of the existing big data security systems makes them susceptible to several security risks. By resolving current technology challenges, including scalability, non-tampering, trustworthiness, data governance, and transparency, blockchain technology plays a vital role and has a significant potential to safeguard personal information. Therefore, this work focuses on addressing real-time big data storage issues based on a transdisciplinary research approach. This study introduces a brand-new approach to big data storage security that leverages blockchain technology and applies highway protocol to generate new blocks. The proposed highway protocol works based on the flexible finality condition to overcome issues of baseline models. The highway allows blocks to run the consensus mechanism to configure security thresholds more freely. The proposed protocol also allows blocks with lower thresholds to reach finality more quickly than blocks requiring greater degrees of confidence. Therefore, the proposed big data framework can dynamically control data manipulation and continuously support individuals to participate in the data-sharing process. A comparison was performed with the number of data requests in terms of hit ratio, and a highway protocol provides better results than baseline models. The proposed model provides a data processing period of 13 to 30 ms and an energy consumption of 32 to 41 mJ.

INDEX TERMS Big data storage, blockchain, consensus mechanism, highway protocol, flexible finality.

I. INTRODUCTION

Web pages, social networking portals, wireless sensor networks, IoT devices, and other customized services collect huge amounts of data, referred to as big data. Big data is also a collection of complicated or large data sets that are challenging to edit using typical database maintenance software tools. In general, big data is classified into two categories such as Unstructured and structured data. Data is frequently extracted and stored on the various systems mentioned above.

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

The storage of big data is the difficult task as more individuals recognize the value of data-driven decisions [1]. The volume of big data has increased dramatically in the last two decades. In the year 2020, IDC anticipated that the volume of data might be more exceeding 40 ZETA-BYTES. According to newer IDC predictions, the volume may reach 75 zeta bytes by 2025 [2]. People and industries benefit from big data by recognizing emerging styles, characterizing developments, and making intelligent choices. Big data has raised concerns over privacy and security, particularly regarding data availability, anonymity, transparency, authenticity, auditing, and surveillance [3].

The exponential growth in global data traffic has piqued researchers' curiosity in big data. Big data is a new wave of advances that aims to analyze massive amounts of data and identify crucial characteristics. Human civilization has reached the era of big data due to the rise of social networking sites, e-commerce, communication devices, and the Internet of Things (IoT) [4]. Big data and IoT techniques are integrated to help industries to take intelligent decisions such as monitoring health care, smart manufacturing and smart energy systems. The volume of big data can be collected in petabytes, and shared computing capability is employed for accurate assessment, projection, and tactical decisions. Smartphone's and wearable technology track every movement and alteration in our activity. Physiological data is also captured and evaluated for a variety of reasons. The effect of data stores occurs when large amounts of data were held and managed by various organizations, including government agencies, corporate groups, academic institutes, and even individuals. In commerce, research, and government services, linking and exchanging the dispersed nodes of big data to gather important insights about goods and assistance has become an inevitable demand [5].

Data, not simply ability, is the most precious commodity in the twenty-first era. Examples of common big data exchange operations in essential services, industries, and studies are the following. Centers for Disease Control and Prevention (CDC) teamed up with Google to analyze 50 million Americans' most recent search histories. Google's huge search inputs accurately identified the H1N1 virus pandemic in the winter of 2009 weeks ahead of time, using the CDC's seasonal influenza data from the previous five years [6]. The Forecast research used tens of trillions of price records from numerous airlines to anticipate domestic trip rates with 75% accuracy. It enables travelers to save an average of more than $50 per ticket, resulting in considerable cost reductions and improved flight usage. From 2003 to 2013, researchers examined 3 billion base pairs in the human body for ten years, and it may take an additional decade to sequence 3 billion base pairings in the body. Nevertheless, the data exchange of genetic decipherment across institutions worldwide can now be accomplished in just 15 minutes. Consolidated big data sharing is used in even the most mundane services, such as Taobao, Weibo, and video sharing.

Despite the numerous benefits of big data, problems exist with every innovation, and big data is no exception. Security, transparency, information storage and analysis, power efficiency, real-time computing, and smart perception are some of the fundamental difficulties of big data. Some of the issues have been resolved due to current re-search projects. For example, the authors of [7] employed the network concept and a coalitional game to secure social network data. Huang et al. [8] developed reinforcement learning-based security-aware algorithms for the mobile edge computing infrastructure.

Public data exchange is a major requirement for China's implementation of information growth, just as it is for other nations. In the framework of the national informatization growth plan announced in July 2016, the executive secretariat of the state council stressed the significance of digital information. This was stated clearly in the "13th Five-Year" national informatization program simultaneously announced by the directorate general of the executive council of the communist party of china that competent management and supervision of data consolidation and distribution is required [9]. As a result, the state council released a strategy for converging and exchanging government data services in May 2017. According to the strategy, free sharing of government records can aid in developing and implementing policies and is one of the key linkages in the state's leading design.

On May 26, 2018, President Xi Jinping addressed a letter of congratulations to the china international big data industrial expo, which was held in Guiyang city, Guizhou County [10]. China places a high priority on expanding big data, according to the official notice, and should stick to the products by creating transparency, collaboration, creativity, and the theme of digitalization of anything. The expo's subject was "Integration of Knowledge" to support the big data sector's revolutionary growth. Even though there are several use cases and regulations to encourage the exchange of big data, the issue of "data silos" continues to be a major source of concern. The total sharing level is rather modest, given the vast amount of data in the hands of agencies and citizens. This stymies data value conversion and human social advancement. The lack of a visible, accessible, and trustworthy data-sharing ecosystem for all parties participating in data-sharing is the core cause of this phenomenon. As a result, agreeing on a fair allocation of data advantages is difficult, and data security needs to be assured. This, in turn, leads to three primary issues: data connectivity issues, data control issues, and data security issues.

The use of blockchain technology has opened up new possibilities for resolving the challenges mentioned earlier. Blockchain is a decentralized digital ledger that stores authorized data that has been hashed and protected. The data ledger is unchangeable; each error or update is traced back to the source. Decentralization, tamper-evident information, and information transparency are all fundamental features of blockchain. As a result, a blockchain system enables the creation of a transparent, open, safe, and reliable data exchange environment for connecting big data from many disciplines. Creating decentralized big data storage with consensus model is very difficult one. Since for every data transfer operation the system needs to verify its validation through PoW or PoS. We implemented a public consensus mechanism for big data storage to overcome the communication cost and energy consumption. This article suggests a large blockchain-based data-sharing approach in this context. The study looks into techniques for establishing secure, open, and effective big data connectivity, data permission administration, and data service personalization.

II. BACKGROUND

A. BIG DATA TECHNOLOGY

Big data technology is the tools and techniques used to process and analyze large amounts of data. This data can come from various sources, such as sensors, social media, and transactional databases, and itcan be structured or unstructured. Big data technology aims to extract valuable insights and knowledge from source generated data that can be used to improve decision-making and drive business value. Some key technology used in big data includes distributed computing, such as Apache Hadoop and Spark, as well as data visualization and machine learning tools.

The "5Vs" of big data are a set of characteristics that define the unique properties of big data. These characteristics are:

Volume: The amount of data generated by humans can be massive, with some estimates proposing that we produce around 2.5 quintillion bytes of data every day.

Variety: Big data can come in many different formats, including structured (e.g., relational databases) and unstructured data (e.g., text, audio, and video).

Velocity: The speed at which data is generated and processed can be extremely high, with some applications requiring real-time analysis at the edge.

Veracity: The quality and accuracy of big data can vary, requiring careful verification and cleansing before it can be used.

Value: The ultimate goal of big data is to extract valuable insights and knowledge that can drive business value.

B. SECURITY CONCERNS OF BIG DATA

The security and maintenance of big data exposes significant issues in real-time processing system. Its qualities posed numerous issues, particularly with the advent of IoT devices that generate huge amounts of data that must be processed to offer important insights and new solutions, such as open-source cloud computing ability and warehousing. As a result, big data security has become essential part of distributed computing system. Standard security solutions, such as firewalls and other peripheral security measures, are also ineffective in a large data environment. We highlight the following large information security challenges in this manuscript:

- Keeping counterfeit data from being generated and added by attackers.
- With large data, accessibility control strategies such as granularity network access (which permits multiple users varied accessibility) are no longer helpful.
- Spreading the storage and computation of massive data over several machines introduced a slew of security concerns. The detection of an attack, for example, could take a lengthy time.
- For any enterprise, cyber security accounting is a critical technical evaluation. It is, however, infrequently used in big data platforms. As a result, managing traceability, which is useful for detecting where intrusions originate, becomes more difficult.

As per the literature, big data storage has concerns with security, privacy, and efficiency. In distributed computing systems, blockchain provides a dependable solution to these issues. These benefits prompted the current research, which aims to use blockchain technology to alleviate edge computing restrictions. On the other hand, traditional blockchain implementations focus on enabling minimal data flow across distributed nodes (i.e., digital transactions). At the same time, edge computing is mostly used to process IoT data, which grows dramatically. As a result, direct implementation suffers from performance concerns when transmitting large amounts of data across a network. The incompatibility of these two technologies presents a slew of challenges during integration. This research focuses on tackling several critical issues, which are described as follows.

Blockchain technology is not intended for the transactions of large amounts of data. Its architecture requires that all nodes in the network share the same copy of the ledger information. Since transaction sizes are so small in crypto exchange, blockchain was born. In the 12 years until 2009, the Bitcoin network has grown to only 350 GB. As a result, even a home computer may join the blockchain network as a miner. The data size could reach hundreds of terabytes in the edge computing platform within hours or days. Finally, all nodes require unbounded physical storage to join the blockchain network, which is not feasible in an edge devices context. Our innovative architecture addresses this problem, which makes blockchain accessible for edge devices.

Security: Edge computing and blockchain should be interconnected for various factors, one of which is security. The key issues regarding the entire system's dependability are the security and authenticity of IoT data [11]. To preserve security, blockchain demands sharing real data, yet edge devices don't have the ability to handle network overflow. Therefore, dealing of data security and scale is crucial one.

System performance: Regarding blockchain and IoT networks, there are two issues to consider: block production time and data transfer time. In a classic blockchain, the first is a natural delay. Verification of Bitcoin transactions takes about 10 minutes. Due to the high amount of IoT data, broadcasting to all nodes may cause unanticipated system delays. This also increases the system latency and make IoT devices are unusable. To increase network dependability, IoT responses should be processed in milliseconds.

C. BLOCKCHAIN TECHNOLOGY

Protecting and processing huge amounts of data is a challenging task. Blockchain may be a great option for resolving many problems associated with big data processing and storage. Decentralization, authenticity, and integrity are all important aspects of blockchain related to bigdata storage.

Blockchain technology: A blockchain is a decentralized node that stores payments or occurrences in connected blocks using hash function. Only miners, which seem to be nodes with significant computing capabilities that authenticate new activities and register them on the database, modify the entries distributed and watched by all nodes in the network. The blockchain comprises connected blocks with time stamp and hash function.

Network of Blockchains: Blockchain is built on a peer-to-peer network as its foundation. Decentralized protocol rules, payment services, consensus algorithm, and block record administration are all followed by every peer in the system. There are three types of blockchains: public, private, and consortium. These classifications are distinguished depending on a block authorization level in terms of being a validator and access data in the network.

A permissionless blockchain lets anyone participate in the network as a validator or user without prior approval or permission. This means that anyone can join the network, read and write data to the blockchain, and validate transactions without requiring permission from a centralized authority or coordinator. Permissionless blockchains are frequently decentralized, which means they lack a centralized point of control or authority. The platform is rather managed by a decentralized network of nodes that collaborate to validate transactions and reach consensus on the state of the blockchain [12].

Permissioned blockchains include both private and consortium networks. A permissioned blockchain is one that needs users or validators to seek permission before they may join the network. This means that users must be granted access to the network by a central authority or coordinator, and they may be required to meet particular contracts or credentials. Permissioned blockchains are frequently used when it is necessary to retain control over who has access to the network and the data kept on the blockchain. This can be useful when the data is sensitive or confidential, or when particular norms or regulations must be followed.

In a permissioned blockchain, the central authority or coordinator has the ability to grant or deny network access, as well as define smart contracts for network participation. This gives the network more control, but it also means that the network is less decentralized than a permissionless blockchain. Hyperledger [13] and Corda [14] are two examples of permissioned blockchains. These networks are frequently utilized in enterprise settings where access to the network and data kept on the blockchain must be controlled. Permissioned blockchains have several advantages, including greater security and the ability to apply specific rules or regulations. They may, however, be perceived as less transparent and decentralized than permissionless blockchains.

The consensus techniques are used to ensure that the transaction data endpoints using smart contracts. All participating peers must agree upon the digitized ledger maintained in the public network. Various distributed ledger categories use different types of consensus mechanisms. For example, permission-less ledger employs a variety of consensus mechanisms, such as proof-of-work and proof-of-stake [15]. Permissioned ledge uses variants of the Byzantine Fault Tolerance (BFT) [16]. The practical BFT (PBFT) [17] and Redundant BFT (RBFT) [18] are two examples of BFT consensus mechanism.

The following stages are included in the blockchain consensus mechanism in general:

- Block suggestion, which entails the creation of cryptographic proofs and their attachment.
- Across-the-network blocks and transactions broadcasting/advertising.
- Check the transactions' authenticity and generate proofs to block verification.
- Block completion is achieved by consensus process of validated block.
- Bonus scheme introduced to encourage honest players and reward them with crypto tokens.

Blockchain innovation can be used for distributed big data storage. Blockchain's distributed structure makes it well-suited to preserving vast amounts of data safely and transparently [19]. The utilization of smart contracts to autonomously handle information storage and recovery is one future application of blockchain in big data storage. This could improve the efficiency and security of storing and accessing vast volumes of data. Furthermore, because blockchain is immutable, it can help protect the integrity of the data recorded.

D. CONTRIBUTION

The convergence of blockchain, IoT, and big data represents the basic techniques that help dynamic data transfer in industrial operations [20]. Combining these techniques also gives several benefits in addressing security issues such as visibility, confidentiality, guaranteeing rights of ownership, decentralization, and so on [21]. On the other hand, the combination of blockchain with IoT is still being researched. Throughout history, these research efforts have struggled to build a big data protection model based on machine learning. Additionally, blockchain is integrated with big data processing system's to avoid the risk of handling real-time data in the network. Instead, scientists developed a blockchain concept that combines a public ledger with a safe transaction mechanism in the industrial sector.

A safe big data storage system based on blockchain technology would store and protect enormous volumes of data by leveraging blockchain's decentralized and immutable characteristics. This might include using smart contracts to manage access to data and ensuring that only authorized parties can view or update information. Furthermore, cryptographic techniques like as hashing and digital signatures could be employed to further secure and assure the integrity of the data. The architecture could potentially include decentralized storage options, such as IPFS (Inter Planetary File System), to distribute data across numerous nodes, making it more resistant to assaults or failures.

The contributions of the big data storage framework are given as follows:

- Developing a secure data storage framework using the decentralized blockchain network. The proposed framework based on the blockchain architecture works on the highway protocol to create blocks in the network.
- Introducing a flexible finality condition-based trust ability validation model for new block creation in the big data storage environment to verify the user authentication.
- Employing a flexible finality condition that deals with verifying new block and protecting distributed data storage network.
- Evaluating the performance of the blockchain-based big data storage framework on metrics such as hit ratio and transaction time for file transfer operations and other baseline approaches.

III. RELATED WORK

IoT and edge computing-based industry is affected by security protection towards distributed file storage system. Existing research focuses on the third-party distributed file system for big data. These methods try to enhance the data encryption protocol, detection and management systems. Although part of the work has third party security process, there is still a gap between distributed file storage adoption and deployment for bigdata applications. The key performance to enhance security through blockchain is the current research of selecting multiple nodes and validating blocks through a consensus mechanism. This article integrates decentralized big data storage with the blockchain layer and realizes the consensus of each node analysis results through the highway protocol.

A. THE SECURITY DEMAND OF BIG DATA FRAMEWORKS

The researchers discussed the most current developments in privacy and security problems in the big data context. A complete review of the fundamental requirements for implementing the privacy and security architecture was discussed in this section. The security requirement of big data storage is illustrated in Fig. 1. The following criteria were analyzed based on big data storage security requirements.

1) USER ACCESS CONTROL

Access Control is the key part of big data in terms of privacy and security. Individuals and companies working with big data should ensure an access control policy in their organization. Access control policies provide how various nodes are interconnected to the network. A flawed access control strategy can enable hacker's unauthorized access to the data that is being kept and create privacy and security risks [22], [23]. Although access control rules have received much academic attention, there may still be problems.

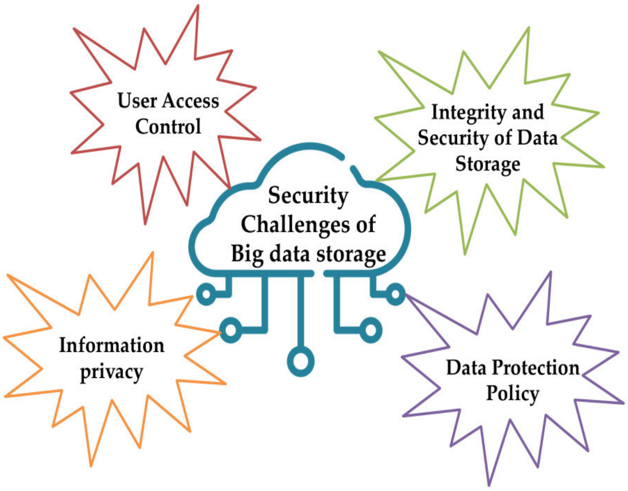

User Access Control
Integrity and Security of Data Storage
Security Challenges of Big data storage
Information privacy
Data Protection Policy

FIGURE 1. Real-time security requirements of big data storage.

2) INTEGRITY AND SECURITY OF DATA STORAGE

The data protection during the transmission stage is another way to ensure security of big data systems. This provides the assurance of data accuracy, consistency, and integrity. End users, software, hardware, and clients are the main offenders of data integrity problems [24]. Big data can occasionally be kept in secret locations, and when this happens, and then data integration becomes a problem. The sets of guidelines applicable to manage transmit, and access data while moving from one memory location to another are called policies in the context of big data. Attackers frequently change policies to acquire data illegally. It might be challenging for most data owners to check whether a policy is legitimate. In contrast to earlier research, which mainly concentrated on data protection than policy, finding a secure mechanism to prevent policies from being amended is crucial.

3) INFORMATION PRIVACY

Data privacy is not disclosing a person's personal information without the permission of that person's knowledge. Big data may contain sensitive or personal information about specific people. Thus, it's crucial to ensure that individuals' information isn't shared without their consent. Even getting a person's consent can only be done for very specific, important reasons. Therefore, datasets does not include any personal distinguishing information, such as name and date of birth, to guarantee the privacy of the data. A security method regarding the application of the Elliptic Curve Cryptography (ECC) technique was suggested by Gupta et al. [25]. Big data are alphabetically categorized into successive segments based on similar or identical IP data types. Because the ECC method uses a tiny key size and offers a high security level, its implementation is as simple as possible. This technique improves the reliability, accessibility, and security of data access.

4) DATA PROTECTION POLICY

Data protection rules for big data are intended to ensure that personal data is gathered, utilized, and kept in accordance with applicable data protection laws and regulations while respecting individuals' privacy rights. Some major parts of a large data protection policy may include:

- Personal data should only be collected, utilized, and stored for specified, clearly stated purposes.
- Only the bare minimum of personal data should be collected, utilized, and retained.
- Personal information should be maintained accurate, and up to date.
- Private data should be stored and transmitted securely, with suitable protocol and organizational safeguards in order to prevent unauthorized access, use, or disclosure.
- Personal data should only be kept for as long as it is needed for the purposes for which it was obtained or as long as required by law.
- Individuals shall have the right to access, rectify, erase, or restrict the processing of their personal data, as well as the right to object to such processing and to data portability.
- It is vital to note that data protection rules and regulations differ by country, thus it is critical to ensure that a data protection policy complies with all applicable legal requirements.

A conceptual framework is proposed by Chung et al. [26] to adopt the protected pattern-enabled data sensitivity interface to enhance the confidentiality of sensitive data in healthcare environments. This method uses machine learning to safeguard information by collecting the same features in health information, such as data frequency and various symbol patterns. The framework also uses a Hadoop system to carry out operations like identifying and encrypting sensitive data. Tests on this framework have shown that it responds quickly and secures critical data at a high level. Privacy safeguards could be used only when transmitting data across the healthcare industry. Another security approach developed by Zeyad et al. [27] involves breaking big data into manageable parts and distributing them among several cloud regions. The arrangement and recovery of the data constitute first of two processes in this solution. In the second stage, data received from the cloud are verified using a secure hash method.

According to Zhang et al. [28], cloud storage should be used to ensure big data security. The cloud platform for this service assures the confidentiality of user data. However, compared to the other alternatives mentioned above, this one leaks at a rate of 1/3, which is higher. The findings also indicate that this approach is appropriate for cloud-based data storage. Fan et al. also put forth a key hierarchy management-based approach to big data security [29]. This scheme has three levels of keys: lower, middle, and upper. The method uses keys to provide security; the middle key encrypts the lower key, which encrypts the middle key. This system sends encrypted data from the client to the cloud. However, this method has some flaws, improving key distribution and transfer, cutting down on encryption time, and improving server-side encryption and decryption behavior.

B. BLOCKCHAIN FOR BIG DATA STORAGE

Several researchers have implemented blockchain-based health data access frameworks in the medical industry, which enables that consumers must have data access to their medical details. Blockchain technology allows secure digital health record sharing in which patients are the proprietors. The authors of [30] employed that the blockchain solely holds metadata for medical fields. As a result, there is no need for massive blockchain architecture to hold all of the medical files.

Human actions on social networking sites create massive amounts of data. Depending on third parties to safeguard sensitive personal information carries considerable risk. As a result, users must be able to monitor and regulate their online activity. According to Ushare [31], blockchain can be utilized as an authorization tool in social platforms. Ushare is a blockchain-based social media platform that gives users control over their data access. Each post has its personal certificate authority that stores the user details and encrypts data before it is published. The blockchain is used to keep track of transactions involving users' transactions. Even though Ushare, as a consumer blockchain platform, may give end-users control over their data, the challenging solution is to execute the large number of blockchain transactions and protecting the full content of the user's data.

G S Aujla et al. proposed a unique format that allows the secure storage of big data in the cloud network [32]. A client-trusted party auditor, cloud service provider (CSP), data service provider (DSP), and Kerberos server are several entities that make up this solution. The DSP uses the attribute-based encryption algorithm to partition and encrypt data into blocks. After that, the CSP receives the encrypted data and stores it at the public while keeping the corresponding tag in the private cloud. This technology can survive numerous attacks and aids in preserving data integrity. Muqaddas et al. [33] provided the Blockchain-based architecture for data sharing, transportation of digital assets, assurance of data quality and authenticity, and a reliable commercial platform for the owner. To solve the swelling issue at the proprietor end, this architecture uses decentralized storage via IPFS storage. In this architecture, a customer review-based mechanism has been implemented to enable users to comments on the data, ensuring data validity. Various smart contracts, including smart ownership contracts and smart access contracts, are being created. This solution performs well in terms of encrypting the transactions with the lowest amount of computing effort.

Hao et al. [34] suggested the novel architecture to address scalability and security issues in an untrusted environment. The proposed design employs blockchain technology to guard against metadata tampering. The fundamental concept behind this design is to leverage blockchain to distribute the name node server, which will safeguard metadata and control user access. The results of the models demonstrate that the system prevent despiteful access and meta-data tampering. A big data management solution based on blockchain technology was proposed by Chen et al. [35]. The central database will store the non-core data, while the blockchain will store the core data and the hash value of the non-core data. This model can address the issue of data redundancy and storage capacity shortage. Blockchain technology is used to increase the security of big data storage because of its decentralized nature and non-tampering ability.

An information management platform created by Chung et al. [36] uses blockchain to record transactions containing data item names, IP addresses, port numbers, and the signature of the information creator. Transactions can be checked by name, and owner-ship can be confirmed to find the desired data item. In contrast to earlier studies, this work develops a complete approach for data security and privacy in a big data context using a smart contract and blockchain in addition to access monitoring and data storage. It is important to note that the usage of a blockchain to address security and privacy issues in a big data context has already been shown in several real-world situations. Fig. 2 provides the use cases of blockchain for big data storage applications.

The main goal of blockchain is to protect and safeguard personal data collected by various entities, such as businesses, organizations, and institutions. Each member in this scenario communicates their encrypted information with a blockchain platform that has been provided with a smart contract that contains transaction regulations. This approach is only applicable to scenarios involving smart transportation. Asaithambi [37] presented a blockchain and trustworthy security model-based SDN data acquisition and marketing system. They merged various technologies, including the control plane, which employs security as data identification, and the SDN model, which guarantees the data-collecting process's trustworthiness.

The developed model employed blockchain to streamline accounting and marketing operations, allowing them to trade-sensitive data and mitigate attacks. In edge devices, trust-building issues among endpoints share massive data. However, their approach needs to consider end-to-end data security. The blockchain-based architecture for co-operative edges is described by Lang [38]. The proposed model enable trusted massive data exchange and ensure optimized resource consumption in end node. Furthermore, the research suggested an offloading mechanism for achieving optimal computational power. Intelligent ecosystems generate a lot of data, usually personal and sensitive information that needs to be secured and protected.

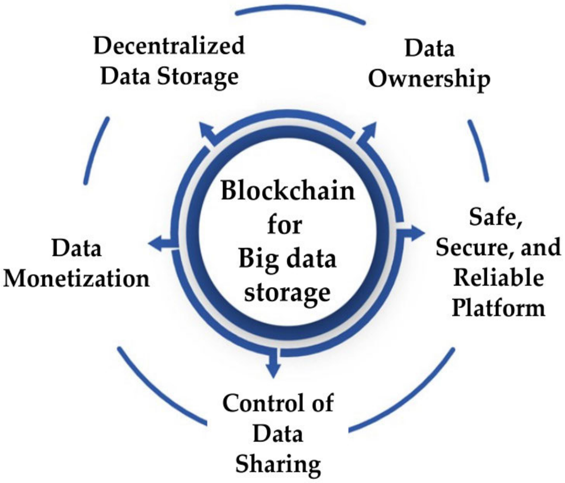

Decentralized Data Storage
Data Ownership
Blockchain for Big data storage
Data Monetization
Safe, Secure, and Reliable Platform
Control of Data Sharing

FIGURE 2. The use cases of blockchain for bigdata storage and management.

The study in [39] presented a strategy for distributed big data audits in pervasive computing based on blockchain to increase stable operation to engage in smart city building. Another study by Sasikumar et al. [40] builds a smart city by combining big data, the power Internet, Sensors, and ethereum. They take advantage of blockchain capabilities consistent with the power internet's architecture. As a result, the problem of costly centralized system management in massive data centers is resolved. A contract is self-executing computer software that is recorded in the chain and runs based on assessing particular conditions. By continuously following default commands, smart contracts on blockchains provide a novel answer to trust challenges with big data. Even though the work provides redundant and distributed memory, it has a high cost, limited recovering capacity, and an increased price of IoT equipment servicing.

The authors of [41] proposed a large blockchain-based information system that used smart contracts to secure huge data exchanges. Access and identification are essential for resolving big data security and privacy issues. Es-Samaali et al. [42] employed a huge data access control system using blockchains. The presented model use smart contracts to code access control policies to validate authorization for huge data access requests. The proposed system implemented with blockchain to enforce access regulations in distributed environment with no centralized authority.

The authors presented a distributed large data auditing strategy for smart buildings relying on blockchain technology in [43]. Their purpose is to improve the reliability and dependability of accounting schemes by eliminating centralized third-party monitoring system. The presented system performance of cost to consumers is high. This method adds extra charges to the ledger navigation during the auditing process. The developed model advocated storing data integrity validation tags on blockchain to minimize the communications and computational costs generated by the authenticity validation process.

In [44] authors developed record chain network for edge computing environment to process the big data. The blockchain technology integrated with bigdata to address the storage utilization issues. An on-chain and off-chain integrated storage model based on blockchain technology to ensure the consistency of bigdata proposed in [45]. The proposed model implemented for spatio-temporal data application. The medical data management using blockchain based framework is implemented in [46]. The authors uses the keyword-searchable encryption model to access the health care data. Authors in [47] presented a big data security framework by incorporating fragmentation and novel access control mechanism. In this method the data fragments are stored in the distributed architecture. The proposed model provides an optimal security solution to the users.

Generally, implementing blockchain mechanism in the context of big data monitoring and protection are novel and time-consuming. While blockchain has been used for information sharing and access management, it has been limited to certain areas and uses in big data. The proposed approach must also be linked with network access, data protection at rest, in transit, and monitoring to increase big data protection. Earlier studies, however, have found that this connectivity is restricted. This research proposes a wide and complete blockchain-based method for organizing and safeguarding massive data to overcome these shortcomings.

IV. PROPOSED BLOCKCHAIN BASED BIG DATA STORAGE FRAMEWORK

In this part, the proposed big data storage framework will be described. The proposed architecture is explained first. Then, the proposed framework workflow discussed. Finally, we describe the novel consensus mechanism of the proposed framework. Fig.3 elaborates on the proposed big data storage architecture of the developed framework. The function and responsibilities of components in the implemented blockchain-based decentralized storage framework are described in the following section.

A. BIG DATA STORAGE NETWORK

The framework of the proposed big data storage chain is described in this section. Due to the computing devices, multi-processing sensor head nodes collect user data and broadcast to the distributed chain. By removing data from the computing network, the proposed framework addresses the challenge of processing large amounts of data in distributed environments. A big data storage chain uses a shared distributed storage node for this purpose. The big data storage chain is the conventional blockchain network initiative to extend its scope beyond cryptocurrency transactions to big data processing. The generated hash function is shared across blockchain network nodes while the actual data is kept in the distributed edge node network.

In contrast to the standard blockchain, a big data storage chain may store multimedia files, GPS information, virtual assets, specifications of sensors, and any other data format that can be secured using a distributed network. As long as the edge processing nodes can adapt it, we can add any number of users to the proposed framework. Because the security performance of the big data storage is unaffected by data size, each node has reliable communication to the distributed storage, which acts as the decentralized network platform, thanks to the proposed blockchain consensus mechanism. Lightweight storage verification and collective consensus model implemented by blockchain eliminate the complex computations and shared data exchange. Big data storage architecture can readily support low-cost edge computing devices thanks to the proposed flexible finality consensus mechanism, which qualifies it for IoT applications.

To prevent single-point failure (SPF) in big data shared storage, edge nodes may turn into data nodes. Big data storage uses a new consensus mechanism for creating a record as single-unit data in the distributed network since the data differ from those in conventional blockchains. In this distributed network, the records combine the hash value, time stamp, and real data. These records are validated through the proposed consensus mechanism, independent of block creation in the distributed network. In comparison to conventional blockchains, the proposed big data storage framework offers the following three primary advantages in relation to the data distribution technique.

Scalability: IoT data's varying sizes make them non-uniform. The hash is never changed, and its size is inconsequential. This size advantage greatly reduces the load on the nodes. Therefore, network coverage can be expanded to any big data application, starting with high-performance servers and ending with embedded circuits with low performance. Since edge computing substantially benefits from the big data storage chain.

Utilization: The crucial component for edge node is storage consumption. The node might only be able to store important data for a short period because of the massive data they get from the outside. Thanks to shared storage, long-term data can be kept indefinitely without draining the resources of edge nodes.

Efficiency: Regarding network throughput, the distribution simply records has an advantage over exchanging actual data. The new block generation is also created based on flexible finality consensus thanks to record approval unanimity. As a result, each record could be validated in just a few milliseconds, providing the blockchain's fast throughput. The subsequent subsections explain the big data storage chains distinctive features more thoroughly.

B. DISTRIBUTED BIG DATA STORAGE IMPLEMENTATION

The deployment of shared storage constitutes the most difficult part of our proposed architecture. The reliability, recovery, and efficiency of shared storage were important considerations as we created a big data framework. First, access to the shared storage must continuously be possible via the hardware memory unit and uninterrupted network connection. The data coming from the user is unpredictable, and computing nodes are dynamic. Therefore, blockchain-based big data framework needs shared storage to be accessible constantly to maintain the network. Second, in the event of a system failure, shared storage should be capable of self-recovery. Since the real data are stored in a single location, as was previously said, data loss may cause a network failure. Data read/write performance is the final crucial aspect of shared storage. The location, bandwidth, and database performance are only a few of the elements that are relevant here. In this study, a shared storage system was created employing distributed service providers, taking into account all the variables mentioned above.

Large volumes of data would be divided into smaller bits and spread across numerous nodes or storage devices in a distributed big data storage network. This can be accomplished through the use of several technologies such as blockchain based decentralized network. A blockchain distributed file system is a prominent storage system that is commonly used in big data processing applications. Distributed file system stores data on decentralized hardware in a cluster and employs replication to assure data dependability and availability through consensus mechanism. Since the blockchain based file storage system can store and retrieve any quantity of data from anywhere. A distributed storage system can improve the storage infrastructure's overall scalability, availability, and fault-tolerance by distributing data among numerous nodes.

We present a new big data security architecture based on blockchain that operates in a distributed. In our proposed architecture blockchain layer placed between user and storage to ensure data security. Fig.3 depicts how it is distributed in each user's main node. We have several distributed file system to store and share the data in the network. The actual data are stored in storage nodes throughout the decentralized file system. The nodes in the network are generated and validated through the proposed highway protocol. We also propose using flexible finality condition to get benefits such as the cheap and efficient use of consensus mechanism channels with high latency and low bandwidth. Our suggested framework has three components: a user interface, blockchain, and a distributed storage component.

C. BIG DATA STORAGE BLOCK GENERATION AND VERIFICATION

The smallest data stored in the proposed framework is a block record. Each block includes a timestamp, collective signature, and the hash produced by the data. The block formation and highway protocol techniques are carried over to the big data storage framework. As seen in Fig. 3, the proposed framework uses a new consensus mechanism-based record approval model. The flexible finality-based block verification approach covered in the next section is the main driving force behind the proposed architecture. To shift all calculations to the genesis node in contrast to the conventional method, highway protocol uses a simplified model with flexible finality condition-based confirmation from the nodes.

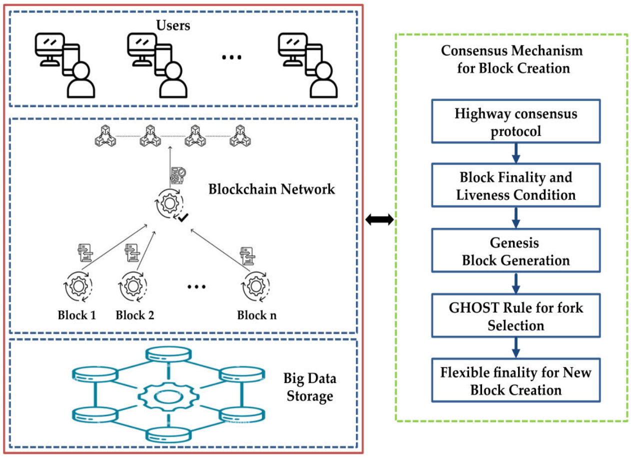

Users
...
Blockchain Network
...
Block 1
Block 2
Block n
Big Data Storage
Consensus Mechanism for Block Creation
Highway consensus protocol
Block Finality and Liveness Condition
Genesis Block Generation
GHOST Rule for fork Selection
Flexible finality for New Block Creation

FIGURE 3. The blockchain-based big data storage architecture.

V. THE CONSENSUS MECHANISM FOR THE BLOCK CREATION IN BIG-DATA STORAGE FRAMEWORK

To resolve the security and privacy problems of the big data storage unit, we proposed the blockchain-based distributed ledger framework. We construct a distributed ledge for big data applications by developing a highway consensus protocol with flexible finality.

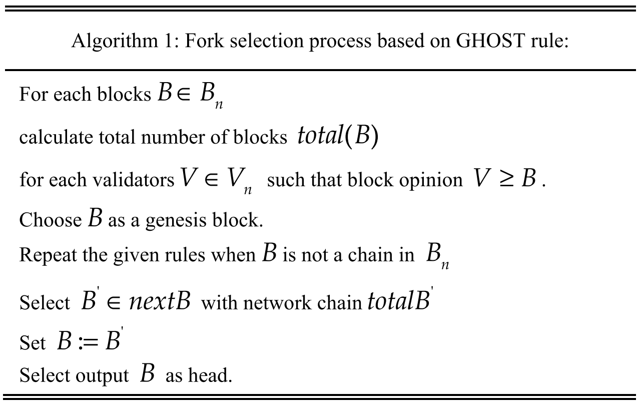
A. A HIGHWAY CONSENSUS PROTOCOL

In this manuscript, we presented a highway consensus mechanism to provide flexible finality based on the various threshold conditions to validators convince that a given block is validated. The finality of a block can only progress for a particular validator until several validators deliberately violate the protocol, which conceptually correlates to the continuously rising number of approvals for a block in a PoW scenario. However, in contrast to PoW, the confidence levels for a particular block on the highway can be easily translated into the number of misbehaving validators required to invert such a block. This can be ensured using the following conditions:

Finality Condition: Once the genuine validator achieves finality with assurance threshold $t \geq f$ for a given legitimate block B, then no genuine validator will ever achieve finality with assurance threshold t for a block participating with B. In threshold condition, the assessment results defined that the impossibility to guarantee the liveness of assurance threshold $n/3$ and greater, in real-time the more number of validators will not get away from the consensus mechanism most of the conditions. In such cases, assurance threshold can be achieved without finality conditions, which create new block during these times, are almost impractical to revert.

The confidence threshold values are generating through a limit value on the amount of genuine validators that guarantee the protocol to select blocks. We consider a limit value that

TABLE 1. **Consensus algorithm.**



```
Algorithm 1: Fork selection process based on GHOST rule:
For each blocks $B \in B_n$
calculate total number of blocks $total(B)$
for each validators $V \in V_n$ such that block opinion $V \geq B$.
Choose $B$ as a genesis block.
Repeat the given rules when $B$ is not a chain in $B_n$
Select $B' \in nextB$ with network chain $totalB'$
Set $B := B'$
Select output $B$ as head.
```

incorporate with the conventional n $\geq$ 3 f +1 limit value combine with notion of crashing faults which is represents by c. The combination of limits and faults interrupt the consensus process, which is not as much as the Byzantine crashing faults model.

Block guarantee condition: For block guarantee process, every confidence threshold value should be in the range of $0 \leq t < n/3$, when f $\leq$ t and c $< \frac{n-3t}{2}$. The confidence threshold value used to generate the chain of blocks based on the genuine validator with confidence t values.

The genesis block generation: The genesis block of blockchain network is represents as G. We need to develop smart contract which generate genesis block. Exclude genesis block, all other blocks are considered as valid block B which indicates to its parent block B. Expect parent block, other valid blocks are denoted by $nextB$ and $prevB$.

The parent block B is validated through $prevB$ and for $nextB$ block the parent will be validating block. Additionally, we consider the height of genesis block is 0 and donated by $H(G)$. For parent and all other block, the height is represents as $H(B) = 1 + H(prevB)$. Consider the following example, we consider B_1 successor B_2, then height of $H(B_2) \leq H(B_1)$, by this way block can reach parent link. GHOST based voting rule for selecting genesis block: The GHOST (Greedy Heaviest Observed Sub-Tree) rule implements to select fork in decentralized network. The following rules are used as a virtual voting process in the blockchain network to guarantee the genesis block.

The rule for fork section: A main building block of blockchain client network is to select fork. In this proposed blockchain model, we consider a group of blocks B_n, which do not need to create a single chain. These set of blocks create a tree in order to choose a single tip from the network, which selected as the head of blockchain. In case of PoW (proof of work) consensus mechanism longest chain of tree consider as head, because which consume more energy than other blocks. In our proposed model, we introduce GHOST rule for fork selection, which is validated by set of blocks in the blockchain network. Sometimes, this process is not possible in case of absence of additional information of validator blocks. We include the set of blocks B_n with opinions of each validators (V_n). The complete steps of GHOST rules are described in the algorithm 1.

Virtual voting based on tree chain: the algorithm 1 already presented the rules for head leaf selection. For virtual voting, each blocks carries a virtual vote, which is denoted as u the overall virtual vote is calculated by the following expression.

$$VirtualVote(u) = GHOST(B_{n_u}, V_{n_u})$$

Flexible finality Condition: The main advantage of our proposed consensus mechanism is flexible finality condition. Here we introduce rules to compute finality of each blocks in the blockchain network. We consider $A(q, k)$ is the submit values of the block B, such that σ is a nested sequence in the chain $(Z_0, Z_1, \ldots, Z_k)$ when $(Z_0 \supseteq Z_1 \supseteq \ldots \supseteq Z_k)$. The rules for flexible finality are given by:

1. Agreement: $VirtualVote(u) \geq B_k$ for all $u \in Z_0$.
2. Genuinity: $E(\sigma) \cap S(Z_0) = 0$, where E represents proven equivocators and S denoted the sender unit.
3. Convexity: $u_0, u_2 \in C_i$ which express $u_1 \in C_i$ includes all $u_0 \leq u_0 \leq u_2$ in order that $S(u_1) = S(u_2) \leq S(u_3)$
4. Solidity: $S(D_u \cap C_i') \geq q$ for all $u_1 \in C_{i+1}$. where D represents down-set and $C_i' = u \in C_i \cup$ $which\ implies\ u \in C_{i+1}\ such\ that\ S(u) = S(u')$.

In our propose protocol, the submit denote uninterrupted blocks generated by a chain of validator that virtual vote of the block$B$. A submit values are used to finalize blocks based on the finality rules. The following finality condition is used to convince the validator in the blockchain and also block do not retracted.

$$Final(B, \sigma, t) = \begin{cases} 1, & if\ (q, k) - for\ block\ (2q - n)(1 - 2^{-k}) > t \\ 0, & otherwise \end{cases}$$

Finally, based on the aforementioned expression the validator V as considering block B will be finalized with confidence threshold t. Hence, we prove that the submit values consists of block height and unit size, therefore a new unit do not virtual vote against them (q, k).

B. SECURITY ANALYSIS OF PROPOSED CONSENSUS PROTOCOL

The following security analysis was performed to ensure the correctness and liveness of block creation. From the analysis results the proposed consensus protocol ensures the liveness of the block in the bigdata storage framework.

1) CORRECTNESS OF PROPOSED CONSENSUS MODEL

Let assume $r \geq 1$ then the sequence $(C_0, C_1, .., C_r)$ is a set of (q, r) for block Bensure the correctness of the new block creation. The sequence element C_0 ensure the convexity for all element in the sequence. Hence the subset of C_0 ensure the honesty and unanimity properties for the sequence of i.

2) OPTIMALITY OF PROPOSED CONSENSUS PROTOCOL

An optimality of highway propocol is proceeding by the down set sequence D_i. The down set ensure the convexity requirement for block such that $D_i \subseteq C_{i\text{and}} S(D_i) \subseteq S(C_i)$ where the base value $i = 0$.

3) LIVENESS OF PROPOSED CONSENSUS MODEL

Assume that $t \geq f$ be an integer number in the block and the affecting node is c $< \frac{n-3t}{2}$. Then the genuine validator move to the state of $Final(B, \sigma, t)$ for any block of B.

This can be ensured by the following assumption. Let assume the set H of all validators with byzantine and crashing nodes. Let n and t are integer values and we consider c $\leq \frac{n-3t-1}{2}$, then the validators guarantee the following condition to ensure the liveness of the node.

$$|H| = \frac{n+t+1}{2}$$

VI. RESULT EVALUATION AND DISCUSSION

The investigations on the big data framework's performance that were carried out in a dynamic and heterogeneous computing environment are presented in this section. The experiments were run on the intel i5 processor CPU, 16 GB RAM, 512 GB external storage and network bandwidth of 10 GB with truffle client ethereum environment. Numerous tests were conducted to determine how well the proposed methodology worked using various metrics. Three different blockchain platforms were used for the evaluations to compare them. As the most recent sample and parent source, PoW [7] was chosen. PoS [28] was also utilized as a benchmark for subsequent high-throughput research deployment on blockchain computing systems. Evaluations are typically carried out on one of the public test networks. One of the secure networks to use shared storage was the big data framework, which was introduced. A virtual evaluation was created in this work employing hardware resource, including an Intel core i5 CPU with 8 GB of RAM and 1 TB of storage to ensure a fair comparison. Multiple instances of the blocks were used to mimic the test network, and the default PoW and PoS setups were used, just like in the proposed system. Table 1 compares the performance of the suggested framework with the benchmark works.

We compare the effectiveness of the proposed approach to that of current techniques like a PoW and a PoS to examine the data transaction rate. Due to its lengthy computation process, PoW has a relatively low transaction rate. Based on computer power, blocks are confirmed in a PoW system. Similarly, PoS require additional transaction time because stake procedures confirm the block. A highway protocol based on a flexible finality mechanism is presented to shorten transaction times.

Similarly, our proposed algorithm was confirmed using validators' finality conditions, which took less time to verify. As a result, our improved highway protocol boosts the transaction rate compared to the other two mechanisms. The transaction rate of proposed model is 2.2 times higher than the PoS and 12 times higher than the PoW consensus model. The reduced energy usage is the key driver behind introducing a flexible finality-based consensus mechanism for big data because it will both replace malicious nodes using genesis blocks and verify the blocks under the concept of a flexible finality condition-based consensus method. The detailed performance comparison of proposed consensus mechanism with base line model is described in Table 2. The main advantage of our proposed model is decentralized public control model with highest number of transactions per time. Since our model will enhance the data transfer rate compared with existing works.

TABLE 2. Performance comparison of various consensus mechanisms.

Consensus Algorithm	PoW	PoS	Highway Protocol
Block creation	Computing power	Stake	Validator weight
Security issues	Constant power	Inactive nodes	Destruction of the Validator
Energy Consumption	Very high	High	Low
Transaction per second	7-30	30-175	100-2500
Reliability	High	Low	Low
Block Identity Control	Public	Public	Public

We compared our proposed model to the following existing works models in order to evaluate our highway procedure using baseline model.

- Random caching (RC), According to the random caching model on edge nodes, nodes should select their cloud server uploads at random [48].
- First input, first output (FIFO): The edge devices transfer the desired data in chronologically to the cloud network. The device with the longest stay will be replaced by the new one if the cloud network caches more than that amount of data.
- Least recently used (LRU): In the event that the cloud network caches more content than needed, the data that hasn't been accessed in the most recent time frame is always chosen to be replaced with fresh data.
- BECS (Blockchain enabled edge computing system): the edge nodes use a distributed selection of contents before uploading them to the cloud server.

The effectiveness of searching is described using the hit ratio. Searching is more efficient the higher the hit ratio. In Fig. 4 and 5, proportion of hot data request (PHDR) represents, respectively, 30% and 70% of all queries. Examining these two figures shows that our hit ratio is better while PHDR is low. This is because the content in the cache is regularly transferred and restored, and our system removes data from the cache that is never needed or won't be accessible for a long time. Therefore, when PHDR is low, the proposed model's hit ratio is at its greatest. However, our scheme has a higher

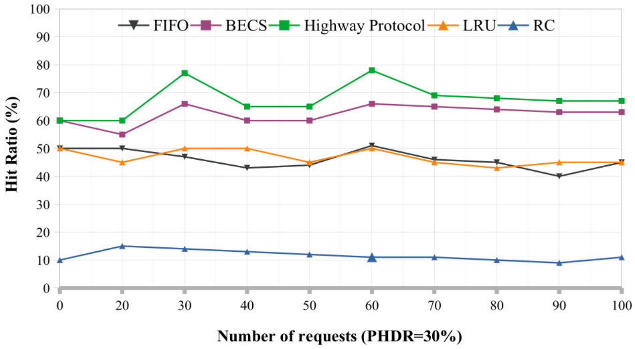


FIGURE 4. Comparison of proposed hit ratio (PHDR = 30%) with existing models.

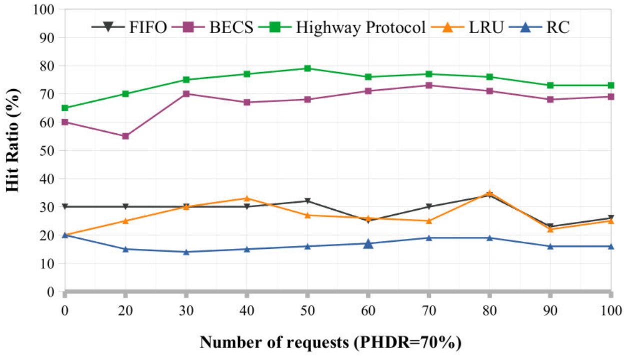


FIGURE 5. Comparison of proposed hit ratio (PHDR = 70%) with existing models.

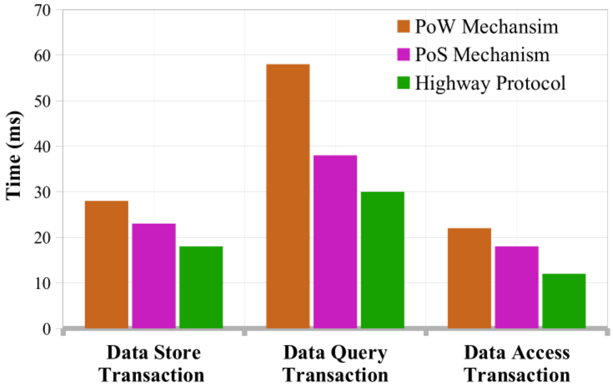


FIGURE 6. End-to-end delay comparison of highway protocol with baseline models.

hit ratio than others when PHDR is high. This is due to the pre-stored hot data being unmodified in the cache, and the higher the hit ratio, the more requests there are for hot data.

For performance evaluation, the baseline strategy, which structures the overlay network like a crypto currency, is contrasted with the proposed way. It is important to note that every node on the network has control over the coin's blockchain. Unlike the proposed technique, this restricts blockchain maintenance to a small number of compute nodes. These nodes check each new block's transactions.

The proposed framework takes less time to handle packets than the baseline, as seen in Fig. 6. The worst-case extra time required by the baseline technique for the query-based store transaction is 33 ms. This is primarily caused by having encryption and hashing functions quite similar.

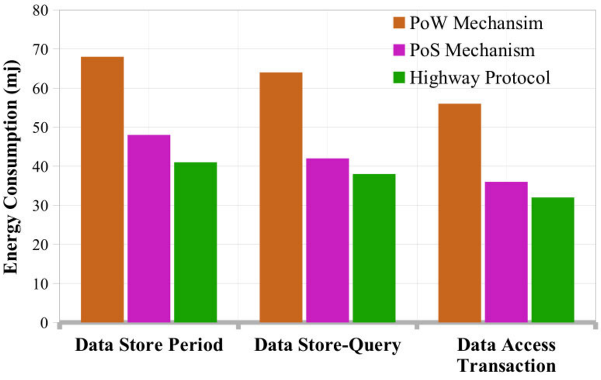


FIGURE 7. End-to-end delay comparison of highway protocol with baseline models.

The Fig. 7 demonstrates how much energy is saved by the suggested structure for performing transactions. The various performance evaluations of the blockchain-based big data storage system created for distributed platforms. All performance metrics prove that the highway protocol consensus mechanism boosts data transaction rates per second while using less energy. The suggested consensus process is best suited for distributed storage environments so that data can be stored securely while using fewer energy resources.

VII. CONCLUSION

The security and privacy of big data storage network is affected by third parties control architecture and it's required-more attention. We discussed significant issues with the big data storage platforms and suggested an effective way to solve them by implementing a safe framework. To ensuresecure and safe data storage in big data contexts, our main goal is to build a reliable and complete architecture based on a blockchain. We set up a virtual blockchain on the distributed storage network to test the effectiveness of the proposed approach. We combined the blockchain with the adaptable finality consensus mechanism to achieve decentralized data storage. We used a consensus mechanism based on the highway protocol for building blocks in the big data storage architecture. We improved the security and privacy of massive data storage by using adjustable finality conditions to secure the security of new blocks. To demonstrate the high scalability and mobility of our framework and reduce traffic overheads, we employed baseline models to assess the security requirements of the new user device. We conducted a performance of our suggested framework to confirm its effectiveness and identify areas that require improvement in subsequent studies. In future we plan to develop energy-efficient consensus mechanism for edge computing data transfer in the context of big data.

REFERENCES

[1] C. Cichy and S. Rass, "An overview of data quality frameworks," *IEEE Access*, vol. 7, pp. 24634–24648, 2019.

[2] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," *IDC iView, IDC Analyze Future*, vol. 2007, pp. 1–16, Jan. 2012.

[3] J. Hu and A. V. Vasilakos, "Energy big data analytics and security: Challenges and opportunities," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2423–2436, Sep. 2016.

[4] R. Kitchin, "Big data, new epistemologies and paradigm shifts," *Big Data Soc.*, vol. 1, no. 1, Apr. 2014, Art. no. 205395171452848.

[5] M. Anshari and S. A. Lim, "E-government withBig data enabledthrough smartphone for public services: Possibilities and challenges," *Int. J. Public Admin.*, vol. 40, no. 13, pp. 1143–1158, Nov. 2017.

[6] W. K. Caldwell, G. Fairchild and S. Y. Del Valle, "Surveilling influenza incidence with centers for disease control and prevention web traffic data: Demonstration using a novel dataset," *J. Med. Internet Res.*, vol. 22, no. 7, 2020.

[7] Z. Su and Q. Xu, "Security-aware resource allocation for mobile social big data: A matching-coalitional game solution," *IEEE Trans. Big Data*, vol. 7, no. 4, pp. 632–642, Oct. 2021.

[8] B.-Q. Huang, G.-Y. Cao, and M. Guo, "Reinforcement learning neural network to the problem of autonomous mobile robot obstacle avoidance," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2005, pp. 85–89.

[9] Y. Hong, "Reading the 13th five-year plan: Reflections on China's ICT policy," *Int. J. Commun.*, vol. 11, pp. 1755–1774, Jan. 2017.

[10] Y. Shi, Z. Shan, J. Li, and Y. Fang, "How China deals with big data," *Ann. Data Sci.*, vol. 4, no. 4, pp. 433–440, Dec. 2017.

[11] I. Ashraf, Y. Park, S. Hur, S. W. Kim, R. Alroobaea, Y. B. Zikria, and S. Nosheen, "A survey on cyber security threats in IoT-enabled maritime industry," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 2, pp. 2677–2690, Feb. 2023.

[12] A. Sasikumar, N. Senthilkumar, V. Subramaniyaswamy, K. Kotecha, V. Indragandhi, and L. Ravi, "An efficient, provably-secure DAG based consensus mechanism for industrial Internet of Things," *Int. J. Interact. Design Manuf.*, vol. 2022, pp. 1–11, May 2022.

[13] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. DeCaro, and J. Yellick, "Hyperledger fabric: A distributed operating system for permissioned blockchains," in *Proc. 13th EuroSys Conf.*, 2018, pp. 1–15.

[14] D. B. Gajić, V. B. Petrović, N. Horvat, D. Dragan, A. Stanisavljević, V. Katić, and J. Popović, "A distributed ledger-based automated marketplace for the decentralized trading of renewable energy in smart grids," *Energies*, vol. 15, no. 6, p. 2121, Mar. 2022.

[15] T. Duong, A. Chepurnoy, L. Fan, and H.-S. Zhou, "TwinsCoin: A cryptocurrency via proof-of-work and proof-of-stake," in *Proc. 2nd ACM Workshop Blockchains, Cryptocurrencies, Contracts*, May 2018, pp. 1–13.

[16] K. Driscoll, B. Hall, H. Sivencrona, and P. Zumsteg, "Byzantine fault tolerance, from theory to reality," in *Proc. Int. Conf. Comput. Saf., Rel., Secur.*, 2013, pp. 235–248.

[17] M. Castro and B. Liskov, "Practical Byzantine fault tolerance and proactive recovery," *ACM Trans. Comput. Syst.*, vol. 20, no. 4, pp. 398–461, Nov. 2002.

[18] P. Aublin, S. B. Mokhtar, and V. Quéma, "RBFT: Redundant Byzantine fault tolerance," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst.*, Jul. 2013, pp. 297–306.

[19] S. Vairavasundaram, K. Kotecha, L. Ravi, G. Selvachandran, and A. Abraham, "Blockchain-based trust mechanism for digital twin empowered industrial Internet of Things," *Future Gener. Comput. Syst.*, vol. 141, pp. 16–27, Apr. 2023.

[20] J. Li, X. Zhang, and W. Shi, "Blockchain application analysis based on IoT data flow," *Electronics*, vol. 11, no. 23, p. 3907, Nov. 2022.

[21] A. Sasikumar, B. Karthikeyan, S. Arunkumar, P. Saravanan, V. Subramaniyaswamy, and L. Ravi, "Blockchain-based decentralized user authentication scheme for letter of guarantee in financial contract management," *Malaysian J. Comput. Sci.*, vol. 2022, pp. 62–73, Mar. 2022.

[22] M. Wazid, A. K. Das, R. Hussain, G. Succi, and J. J. P. C. Rodrigues, "Authentication in cloud-driven IoT-based big data environment: Survey and outlook," *J. Syst. Archit.*, vol. 97, pp. 185–196, Aug. 2019.

[23] C. Hu, W. Li, X. Cheng, J. Yu, S. Wang, and R. Bie, "A secure and verifiable access control scheme for big data storage in clouds," *IEEE Trans. Big Data*, vol. 4, no. 3, pp. 341–355, Sep. 2018.

[24] Y. E. Oktian, S.-G. Lee, and B.-G. Lee, "Blockchain-based continued integrity service for IoT big data management: A comprehensive design," *Electronics*, vol. 9, no. 9, p. 1434, Sep. 2020.

[25] S. Gupta, S. Vashisht, D. Singh, and P. Kushwaha, "Enhancing big data security using elliptic curve cryptography," in *Proc. Int. Conf. Autom., Comput. Technol. Manage. (ICACTM)*, Apr. 2019, pp. 348–351.

[26] Y. C. Yau, P. Khethavath, and J. A. Figueroa, "Secure pattern-based data sensitivity framework for big data in healthcare," in *Proc. IEEE Int. Conf. Big Data, Cloud Comput., Data Sci. Eng. (BCD)*, May 2019, pp. 65–70.

[27] Z. A. Al-Odat, E. M. Al-Qtiemat, and S. U. Khan, "A big data storage scheme based on distributed storage locations and multiple authorizations," in *Proc. IEEE 5th Int. Conf. Big Data Secur. Cloud Int. Conf. High Perform. Smart Comput., (HPSC) IEEE Int. Conf. Intell. Data Secur. (IDS)*, May 2019, pp. 13–18.

[28] Y. Zhang, R. H. Deng, D. Zheng, J. Li, P. Wu, and J. Cao, "Efficient and robust certificateless signature for data crowdsensing in cloud-assisted industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5099–5108, Sep. 2019.

[29] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "MedBlock: Efficient and secure medical data sharing via blockchain," *J. Med. Syst.*, vol. 42, no. 8, pp. 1–11, Aug. 2018.

[30] A. A. Vărzaru, "Assessing digital transformation of cost accounting tools in healthcare," *Int. J. Environ. Res. Public Health*, vol. 19, no. 23, p. 15572, Nov. 2022.

[31] A. Chakravorty and C. Rong, "Ushare: User controlled social media based on blockchain," in *Proc. 11th Int. Conf. Ubiquitous Inf. Manage. Commun.*, Jan. 2017, pp. 1–6.

[32] S. Garg, A. Singh, K. Kaur, G. S. Aujla, S. Batra, N. Kumar, and M. S. Obaidat, "Edge computing-based security framework for big data analytics in VANETs," *IEEE Netw.*, vol. 33, no. 2, pp. 72–81, Mar. 2019.

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

[34] K. Hao, J. Xin, Z. Wang, K. Cao, and G. Wang, "Blockchain-based outsourced storage schema in untrusted environment," *IEEE Access*, vol. 7, pp. 122707–122721, 2019.

[35] J. Chen, Z. Lv, and H. Song, "Design of personnel big data management system based on blockchain," *Future Gener. Comput. Syst.*, vol. 101, pp. 1122–1129, Dec. 2019.

[36] K. Chung and H. Jung, "Knowledge-based block chain networks for health log data management mobile service," *Pers. Ubiquitous Comput.*, vol. 30, pp. 1–9, Apr. 2019.

[37] S. Asaithambi, L. Ravi, H. Kotb, A. H. Milyani, A. A. Azhari, S. Nallusamy, V. Varadarajan, and S. Vairavasundaram, "An energy-efficient and blockchain-integrated software defined network for the industrial Internet of Things," *Sensors*, vol. 22, no. 20, p. 7917, Oct. 2022.

[38] P. Lang, D. Tian, X. Duan, J. Zhou, Z. Sheng, and V. C. M. Leung, "Cooperative computation offloading in blockchain-based vehicular edge computing networks," *IEEE Trans. Intell. Vehicles*, vol. 7, no. 3, pp. 783–798, Sep. 2022.

[39] R. Zheng, J. Jiang, X. Hao, W. Ren, F. Xiong, and Y. Ren, "BcBIM: A blockchain-based big data model for BIM modification audit and provenance in mobile cloud," *Math. Problems Eng.*, vol. 2019, pp. 1–13, Mar. 2019.

[40] A. Sasikumar, L. Ravi, K. Kotecha, J. R. Saini, V. Varadarajan, and V. Subramaniyaswamy, "Sustainable smart industry: A secure and energy efficient consensus mechanism for artificial intelligence enabled industrial Internet of Things," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, Jun. 2022.

[41] B. Kirpes and C. Becker, "Processing electric vehicle charging transactions in a blockchain-based information system," in *Proc. Digital Disruption Americas Conf. Inf. Syst. (AMCIS)*. New Orleans, LA, USA: Association for Information Systems, Hyatt Regency, 2018.

[42] H. Es-Samaali, A. Outchakoucht, and J. P. Leroy, "A blockchain-based access control for big data," *Int. J. Comput. Netw. Commun. Secur.*, vol. 5, no. 7, p. 137, 2017.

[43] D. Minoli, K. Sohraby, and B. Occhiogrosso, "IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 269–283, Feb. 2017.

[44] K. Tulkinbekov and D. Kim, "Blockchain-enabled approach for big data processing in edge computing," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18473–18486, Oct. 2022.

[45] Y. Ren, D. Huang, W. Wang, and X. Yu, "BSMD: A blockchain-based secure storage mechanism for big spatio-temporal data," *Future Gener. Comput. Syst.*, vol. 138, pp. 328–338, Jan. 2023.

[46] C. Li, M. Dong, J. Li, G. Xu, X. Chen, W. Liu, and K. Ota, "Efficient medical big data management with keyword-searchable encryption in healthchain," *IEEE Syst. J.*, vol. 16, no. 4, pp. 5521–5532, Dec. 2022.

[47] H. E. Alhazmi, F. E. Eassa, and S. M. Sandokji, "Towards big data security framework by leveraging fragmentation and blockchain technology," *IEEE Access*, vol. 10, pp. 10768–10782, 2022.

[48] S. Li, J. Xu, M. van der Schaar, and W. Li, "Trend-aware video caching through online learning," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2503–2516, Dec. 2016.

A. SASIKUMAR received the B.E. and M.E. degrees from Anna University, in 2011 and 2013, respectively, and the Ph.D. degree from SASTRA Deemed University, in 2020. He is currently an Assistant Professor with the Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Chennai. India. He has published more than 25 journal articles. His research interests include blockchain, analog VLSI, digital VLSI, and swarm intelligence.

LOGESH RAVI is currently with the Centre for Advanced Data Science, Vellore Institute of Technology, Chennai, India. He has published more than 100 papers in reputed international journals and conferences. His research interests include artificial intelligence, recommender systems, big data, information retrieval, fintech, and social computing. He is listed and ranked in prestigious Top 2% Scientists Worldwide by Stanford University and Elsevier B.V.

KETAN KOTECHA is currently an Administrator and a Teacher with the Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India. He has expertise and experience in cutting-edge research and projects in AI and deep learning for the last 25 years. He has published more than 200 papers widely in several excellent peer-reviewed journals on various topics ranging from cutting-edge AI, education policies, teaching-learning practices, and AI for all. He has published three patents and delivered keynote speeches at various national and international forums, including the Machine Intelligence Laboratory, USA, IIT Bombay under the World Bank Project, and the International Indian Science Festival organized by the Department of Science and Technology, Government of India. His research interests include artificial intelligence, computer algorithms, machine learning, and deep learning. He was a recipient of the two SPARC projects in AI worth INR 166 lakhs from MHRD, Government of India, in collaboration with Arizona State University, USA, and The University of Queensland, Australia. He was also a recipient of numerous prestigious awards, such as Erasmus+ Faculty Mobility Grant to Poland, the DUO–India Professors Fellowship for research in responsible AI in collaboration with Brunel University, U.K., the LEAP Grant by Cambridge University, U.K., the UKIERI Grant by Aston University, U.K., and a Grant from the Royal Academy of Engineering, U.K., under Newton Bhabha Fund. He is an Associate Editor of IEEE Access.

AJITH ABRAHAM (Senior Member, IEEE) received the M.Sc. degree from Nanyang Technological University, Singapore, in 1998, and the Ph.D. degree in computer science from Monash University, Melbourne, Australia, in 2001. He is currently a Professor in artificial intelligence at Innopolis University, Russia, and the Yayasan Tun Ismail Mohamed Ali Professorial Chair in artificial intelligence with UCSI, Malaysia. He is also the Director of the Machine Intelligence Research Laboratories (MIR Laboratories), a Not-for-Profit Scientific Network for Innovation and Research Excellence Connecting Industry and Academia. The Network with HQ in Seattle, USA, currently has over 1,500 scientific members from over 105 countries. As an investigator/a co-investigator, he has won research grants worth more than U.S. $100 million. Currently, he holds two university professorial appointments. He works in a multi-disciplinary environment. He has authored/coauthored more than 1,400 research publications out of which there are more than 100 books covering various aspects of computer science. One of his books was translated into Japanese and a few other articles were translated into Russian and Chinese. He has more than 46,000 academic citations (H-index of more than 102 as Per Google Scholar). He has given over 150 plenary lectures and conference tutorials (in more than 20 countries). He was the Chair of IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing (which has over 200 members), from 2008 to 2021, and served as a Distinguished Lecturer for IEEE Computer Society representing Europe, from 2011 to 2013. He was the Editor-in-Chief of *Engineering Applications of Artificial Intelligence* (EAAI), from 2016 to 2021, and serves/served on the editorial board for over 15 international journals indexed by Thomson ISI.

MALATHI DEVARAJAN received the B.Tech. degree in information technology and the M.Tech. degree in computational biology from Pondicherry University, India, and the Ph.D. degree in computer science and engineering (cybersecurity) from SASTRA, India. She is currently with the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India. Her research interests include cybersecurity, blockchain, network security, the IoT, and cloud computing.

SUBRAMANIYASWAMY VAIRAVASUNDARAM received the Ph.D. degree from Anna University, in 2013. After his Ph.D. degree, he continued the extension work with the Department of Science and Technology support as a Young Scientist Award holder. He is currently a Professor with the School of Computing, SASTRA Deemed University, Thanjavur, India. With the experience of more than 18 years as an academician and researcher, he has contributed more than 175 papers and chapters for many high-quality technology journals and books that are being edited by internationally acclaimed professors and professionals. He has received government funded and consultancy projects from DST-SERB, ICSSR–IMPRESS, MHRD, TVS MOTORS, and SERB–MATRICS. His research interests include recommender systems, the Internet of Things, artificial intelligence, machine learning, and big data analytics. He is on the reviewer board of several international journals and has been a program committee member for several international/national conferences and workshops. He also serves as the guest editor for various special issues of reputed international journals. He is serving as a research supervisor and successfully guided five research scholars and a visiting expert to various universities in India and Abroad.