2022 International Balkan Conference on Communications and Networking (BalkanCom) | 978-1-6654-8764-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BalkanCom55633.2022.9900867

OVERVIEW OF BLOCKCHAIN DATA STORAGE AND PRIVACY PROTECTION

Avni Rustemi
Faculty of Electrical Engineering
and Information Technology (FEIT)
Ss. Cyril and Methodius University
in Skopje
Skopje, North Macedonia

Vladimir Atanasovski
Faculty of Electrical Engineering
and Information Technology (FEIT)
Ss. Cyril and Methodius University
in Skopje
Skopje, North Macedonia

Aleksandar Risteski
Faculty of Electrical Engineering
and Information Technology (FEIT)
Ss. Cyril and Methodius University
in Skopje
Skopje, North Macedonia

Abstract - Blockchain nowadays is finding great application in many areas, based on its security, decentralized mode of operation, however many things during its use remain challenging include the issue of data privacy, such as safe and effective data management, maintaining data integrity and data quality, etc. Through this paper we will try to make an overview of existing blockchain technologies, their use today and sublimating existing and future trends. We will briefly describe application of blockchain technology in various fields suggesting solutions, advantages and disadvantages of blockchain, cloud platforms for blockchain data storage. We will focus on data privacy by making a description of the strategies used in this regard. We will conclude the paper by giving the latest research regarding data storage, the importance of data storage and our views regarding the future of data management and storage.

Keywords: blockchain, cloud platforms, data storage, data privacy, data management, future trends.

I.INTRODUCTION

Blockchain is one of the most used technologies nowadays thanks to the various cryptographic techniques used as well as the specific characteristics it has shown for the protection, processing and transmission of data [1]. Data security undoubtedly plays a very important role in creating and managing different systems, so the main focus of researchers is how to enable more secure access and storage of data in different sources. One of the many reasons that blockchain has found great application is decentralized form of data processing which undoubtedly has its advantages. It is very important the impact of the blockchain on the trade network, respectively on the supply chain [2]. Blockchain today is finding application in almost all spheres of life, including medicine, government, IoT devices, thanks to the way it works, it is currently recognized as one of the safest for application. Through this paper we want to make a brief summary for important issues in terms of data security and management, strategy used to maintain their privacy, characteristics of blockchain, the advantages and disadvantages of blockchain, application of blockchain in different fields, also giving our views on some use cases of this technology. Most important are the views that researchers present in terms of the future of data storage and data privacy. An analysis will be made of cloud platforms based on blockchain technology that are mostly used today, giving a diagram where we graphically present the market cap for each of them. Before we conclude our paper, we describe related work, future trends, detailing the most recent research on privacy and data storage, and our aim towards other researches in this direction.

II. APPLICATION AND CHARACTERISTICS OF BLOCKCHAIN

The application of blockchain technology, big data and cloud computing is being seen in various areas of life, as everyone is interested their data to be of a higher security, because decryption of data can lead to catastrophes and great damage, escpecially data privacy violations. Blockchain offered a higher security of data protection, higher security during transmission of data and has reduced the possibility of data being deciphered by other people, which has made it applicable in many spheres of life. Blockchain technology, since its existence, has gone through three phase of development. The main characteristic of the first phase is the earning of money and the creation of a large number of cryptocurrencies, the second phase is characterized by the integration of the smart contract as well as the third phase of development is about the application of blockchain not only for material benefits, but its extension in medicine, education as well as many spheres of life which can facilitate the work with its application [3]. Blockchain in IoT devices can find great use, such as during the creation of various applications for education, whether for schools or even for home use. Various home appliances today are being replaced by IoT devices [4]. However, with the application of blockchain for the creation of these smart devices, the safety and confidence of customers has undoubtedly increased [5].

978-1-6654-8764-1/22/$31.00 ©2022 IEEE

Table 1 present advantages and disadvantages that this technology has.

Table 1. Advantages and disadvantages of blockchain

Advantages of blockchain	Disadvantages of blockchain
A shared data store is used and data can be accessed by different people from different places at the same time.	Although the mining process is useful and very important for transactions, it requires expensive hardware equipment.
It provides trust and security to the parties involved in conducting transactions	The process of making transactions consumes a lot of electricity.
Every process is transparent and customers can see every change.	The process of inserting data is slow, transparency can harm the privacy of customers.
It is characterized by data stability, and a decentralized work system	Problems with data storage space due to data duplication, redundancy, due to data security

The application of blockchain in banking systems is inevitable, for safe and fast transactions. Also in medicine, during the performance of various operations, where patients with specific diseases are automatically intervened, electronic data management through sophisticated systems working on platforms based on blockchain, the interconnection between different systems working for different purposes during operations and etc [6]. Also the cloud platforms that are used nowadays, most of them work based on blockchain technology. This means that blockchain has found wide application in data storage and management [7]. Figure 1 shows some use cases where blockchain technology has been applied.

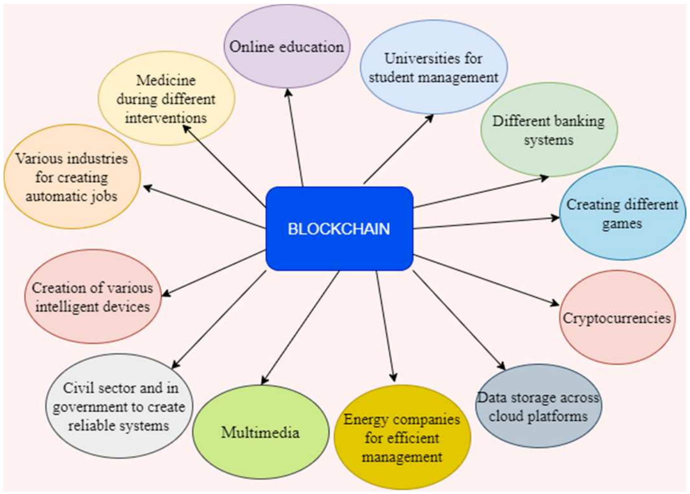

Online education
Universities for student management
Medicine during different interventions
Different banking systems
Various industries for creating automatic jobs
BLOCKCHAIN
Creating different games
Creation of various intelligent devices
Cryptocurrencies
Civil sector and in government to create reliable systems
Multimedia
Energy companies for efficient management
Data storage across cloud platforms

Figure1. Blockchain technology use cases

III. CLOUD PLATFORMS FOR BLOCKCHAIN DATA STORAGE

Cloud platforms are of particular importance for managing, transporting and processing data. They implement blockchain technology due to more reasons, among which we can mention faster data transfer, security during data transfer, privacy protection, etc. We will clarify some of cloud platforms by describing the advantages and disadvantages, as well as their usability. IPFS it is a file system for transferring data based on a decentralized way, where users can exchange their data through certain channels by having the hash codes in advance, but it is necessary that both sides to be online during the transmission. For the distribution of data and files, IPFS uses the BitSwap protocol, which is composed of blocks [8]. Unlike IPFS, which is a web-oriented, open and public method for data transfer and distribution, MTFS it is also a file system which aims to carry and transfer data, but offers a more private way of data transfer using cryptographic algorithms to ensure secure transfer between two parties exchanging data [9]. Sia is another cloud platform, which uses the cryptocurrency Siacoin, through which it enables users to buy data storage space, which space is decentralized and can be used by more users. The reasons why Sia is used, considering that there are a large number of cloud platforms, is the speed of data download from data storage. In the analyses made in [10], it turns out that Sia has a high download speed, while the data upload speed is slower, where one of the most used cloud platforms, that of Google Cloud, was taken as a comparison. Filecoin is also a decentralized system that works based on the blockchain, similar to IPFS, because they work in distributed networks for data transfer. Like IPFS, Filecoin works on a peer-to-peer network basis, and is mostly used to transfer data from a desktop PC or laptop to a cloud server. Their disadvatange is that both work only with online internet when transferring data from the desktop to the server, although various researches are being done in this direction on how to transfer data even while the devices are not connected to the internet, considering that both desktops, laptops and other intelligent devices have their own static memory which can be used temporarily to transfer data to another device, and then when there is Internet access to be transferred to the cloud servers assigned [11]. SWARM is also another cloud data distributed technique based on blockchain that uses intelligent algorithms to manage data on a server. Swarm is a technique which is usually used more to balance the data in a cloud server, considering that in a cloud server there is a lot of data, and at the same time many virtual machines are installed, which means that at the same time they are transferred data from more countries, it is very important those data to be provided to the customers as quickly as possible. Therefore, to avoid long waits, since at the same time there may be more requests for data access, someone must deal with the distribution of data and the management of data coming from different VMs. SWARM is suitable for data balancing [12]. Azure is also another cloud platform, created by the Microsoft company, which we can freely say is currently in the group of most used platforms by customers, based on the company's benefits from this platform. The organization of data on this platform is done through the data centers. Access to the Azure server can be done in several ways, including through Power Shell, the Azure portal, or Json templates. As a disadvantage of Azure is the non-equal performance of all servers, because Azure is known for the multiple distribution of servers in more countries of the world. [13]. Google cloud platform is a platform that offers its customers to use some of the technologies that the company Google is using itself, where customers have the opportunity to store data on Google's servers. It is worth noting that large companies such as Google, Microsoft, Amazon all offer secure services, secure data transfer, but depending on the needs of customers, the circumstances that suit their needs, the software that customers use, they can choose any of the cloud platforms that these companies offer. All these cloud platforms have included blockchain technology in their services [14].

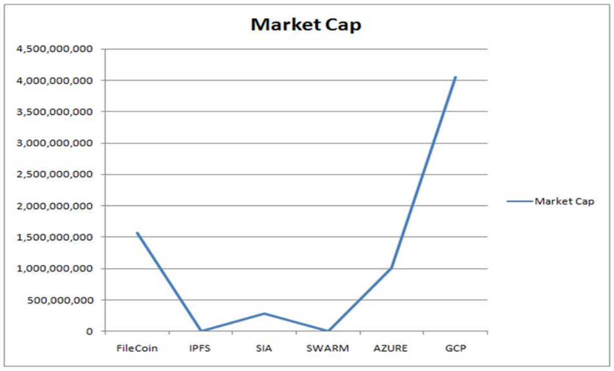
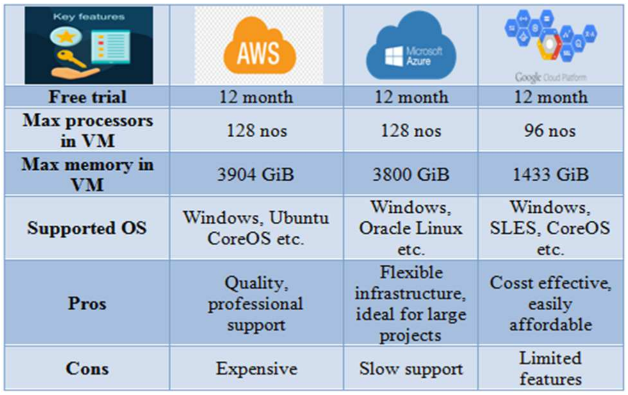


Figure 2. Market cap of cloud blockchain platforms

In the figure 2 we have presented a graph based on the market cap of all the cloud platforms that we have explained in this section, and it is clear that Microsoft Azure and Google Cloud are among the platforms that have the greatest benefits, resulting from the use of more large number of customers. The market cap values to create the diagram is taken from [15]. Figure 3 present the three most used cloud platforms presenting in short points the main characteristics for each of them based on the analyzes made in [16].

IV. DATA PRIVACY PROTECTION MECHANISM

Data is undoubtedly among the most important values in every sphere of life. Any data corruption by third party people would cause catastrophic consequences and data privacy would be called into question. Although there is always an attempt to incorporate the newest techniques that guarantee the preservation of privacy into platforms that enable data storage and use more secure databases, research and innovations in the field of data privacy have not stopped.



Key features	AWS	Microsoft Azure	Google Cloud Platform
Free trial	12 month	12 month	12 month
Max processors in VM	128 nos	128 nos	96 nos
Max memory in VM	3904 GiB	3800 GiB	1433 GiB
Supported OS	Windows, Ubuntu CoreOS etc.	Windows, Oracle Linux etc.	Windows, SLES, CoreOS etc.
Pros	Quality, professional support	Flexible infrastructure, ideal for large projects	Cosst effective, easily affordable
Cons	Expensive	Slow support	Limited features

Figure 3. The most used cloud platforms

There are always opportunities for some research related to privacy and data storage. The inclusion of blockchain technology in cloud platforms had a positive impact on the detection of data access opportunities by unauthorized persons. This is because any access to the system by other people, the blockchain technology immediately detects interference and immediately performs checks of the initial hash value, and if there is a change in the hash value, it means there was an attempt to access that data and immediately informs the client and the service provider [17]. Transferring data to cloud servers and maintaining data privacy is a very difficult and challenging process for the providers that offer these services. This is due to the many changes and upgrades that occur in cloud server services, due to the fast dynamics of data transfer, the complexity of the internal organization of servers, and on the other hand, preserving data privacy is a very difficult issue that must be analyzed in detail so that the copyright and private rights of customers not to be violated at any time. Figure 4 shows the architecture for data protection in the cloud, where all implementation details you can find in [18]. We will briefly describe some mechanisms used for privacy protection that work based on blockchain technology, such as bitcoin mixing services, signature scheme, multi-party computation, zero knowledge proofs and homomorphic encryption. Zerocoin is a distributed electronic system that enables online Bitcoin transactions, but by using sophisticated cryptographic techniques, and such offers higher security and data privacy protection. This mechanism is actually based on prior knowledge that the client must have, and certain proofs that prove that during the transaction there really was no interference with the coin [19].

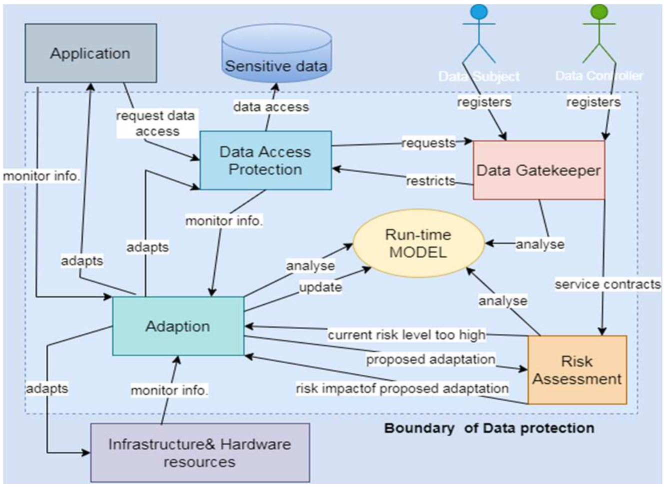


Figure 4. Architecture for data protection in the cloud

Coin Mixing technique is another technique for maintaining the privacy of customers while making transactions. This technique is based on the fact that there is always an intermediary between the buyer and the seller, who will perform all the services between them. By using the intermediary, the origin of the cryptocurrency is hidden, but also the identity and privacy of both parties. Also, this mechanism has the ability to detect even if this intermediary tries to cooperate with any of the parties, either the customer or the coin seller [20]. Ring signature is technique that is used to preserve the privacy of the client, where before a message is sent, in this case to the blockchain before the transaction is carried out, a ring is formed, where, among other things, it must be part of the ring as well the person who wants to perform the transaction. To carry out transactions, the private key and the public key of the other members of the ring are needed. This method is very efficient in protecting privacy because no one who is not part of the ring can get information about the members of the ring, nor what transactions they perform. Although there are different schemes of this technique developed to date, the essence of the operation of this protective mechanism can be found in [21]. Homomorphic encryption is mechanism that enables mathematical operations to be performed on the data that are encrypted, respectively to manage the data that is encrypted in the cloud, and to perform various operations on the same that are requested by customers [22]. To be even more clear, if we consider the data to be encrypted as plaintext, and the encrypted data as ciphertext, then, homomorphic encryption, allows us to perform calculations with plaintext in this case, using ciphertext concerning them, but without affecting the privacy of the data. Different schemes have been developed for this mechanism too, where more details are given in [23].

Table 2 shows the main characteristics for each of the protective mechanisms, which are part of our paper.

Table 2. Privacy protection technology strategies

Protection strategy	Characteristics
Zero-knowledge proof	works based on proof and prior knowledge
Coin mixing technology	use intermediaries to carry out transactions
Ring signature	transactions are made only after a ring is formed and only those members can exchange coins among themselves
Homomorphic encryption	enables the computing of encrypted data

Undoubtedly, even these strategies are not perfect, they have their own limitations, however, they find application in cloud platforms with different modifications of their schemes.

V. RELATED WORK

With the development of blockchain technology, the integration of various fields and blockchain has become more and more close, exposing more and more privacy security issues, so the original privacy protection technology needs to be further improved. Although it finds great application in many fields, for blockchain technology there are still many areas for research, and we have tried to make a summary in terms of the blockchain privacy issue, respectively we have described some strategies that are used to protect and preserve data privacy in the blockchain. Blockchain is a technology that is increasingly being applied in more fields, but our focus will be mainly on data management, their security and how much this technology has found application in cloud platforms. BigchainDB will also be part of our research, to research security and data management in this database the application of the same in cloud platforms and finally to create our prototype where we will elaborate on all these challenges.

VI. FUTURE TRENDS

The main purpose of this paper is to analyze and summarize existing knowledge regarding privacy and data storage in blockchain technology. Although many techniques are being developed for data management and storage, the problem with data capacity is becoming increasingly complex. This is because we are facing large data capacities, which require large storage, which is interrelated with many economic, social factors. New data storage techniques are being considered to be implemented as 5D Optical Storage, Cold Storage, DNA Storage [23], with the purpose of higher data security and the reduction of data storage space. Data privacy plays an important role in data management. This means that different data storage must necessarily guarantee security and privacy. Despite the fact that there are many different techniques and strategies used in data management, the attempt of people to hack them will not stop. Therefore, even research in this regard will not stop because there is always something new that should be part of various researches in terms of data privacy and data storage. Blockchain technology currently remains one of the most lucrative technologies, and that has radically changed the issue of data management.

VII. CONCLUSIONS

The paper discusses blockchain technology, sublimate existing and future trends regarding data management and storage. After describing the application of blockchain technology in different fields, particular importance is given to the cloud platforms used for data storage, giving brief details about each of them. Several mechanisms for data storage and privacy are explained, describing the characteristics for each of them. Our future challenge is to research in detail the limitations of the mechanisms for data storage and privacy and to propose a protection scheme through which we would overcome these limitations and improve efficiency, perfomance and security.

VIII. REFERENCES

[1] Sri Santhoshi Devi Arigela, Persis Voola. "Detecting and Identifying Storage issues using Blockchain Technology", 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022.

[2] Jabbar, Sohail; Lloyd, Huw; Hammoudeh, Mohammad; Adebisi, Bamidele; Raza, Umar (2020). Blockchain-enabled supply chain: analysis, challenges, and future directions. Multimedia Systems, (), –. doi:10.1007/s00530-020-00687-0, Springer.

[3] Gatteschi, V., Lamberti, F., Demartini, C., Pranteda, C., & Santamaria, V. (2018). To Blockchain or Not to Blockchain: That Is the Question. IT Professional, 20(2), 62–74. doi:10.1109/mitp.2018.021921652.

[4] Alam, Tanweer. (2019). Blockchain and its Role in the Internet of Things (IoT)., IJCRSEIT, DOI : https://doi.org/10.32628/CSEIT195137.

[5] Fei Chen, Zhe Xiao, Laizhong Cui, Qiuzhen Lin, Jianqiang Li, Shui Yu, "Blockchain for Internet of things applications: A review and open issues", Journal of Network and Computer Applications, Volume 172, 2020, Science Direct, Elsevier.

[6] Abid Haleem, Mohd Javaid, Ravi Pratap Singh, Rajiv Suman, Shanay Rab, "Blockchain technology applications in healthcare: An overview", International Journal of Intelligent Networks, Volume 2, 2021, Pages 130-139, Science Direct.

[7] M. R. Dorsala, V.N. Sastry, S. Chapram, "Blockchain-based solutions for cloud computing: A survey", Journal of Network and Computer Applications, Volume 196, 2021, ScienceDirect.

[8] B. Guidi, A. Michienzi, L. Ricci, “Data Persistence in Decentralized Social Applications: the IPFS approach”, 2021 IEEE 18th Annual Consumer Communications &amp; Networking Conference (CCNC) |©2021 IEEE |.

[9] J. Kan and K. S. Kim, "MTFS: Merkle-Tree-Based File System," 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 2019, pp. 43-47, doi: 10.1109/BLOC.2019.8751389.

[10] P. Austria, C. H. Park, A. Hoffman and Y. Kim, "Performance and Cost Analysis of Sia, a Blockchain-Based Storage Platform," 2021 IEEE/ACIS 6th International Conference on Big Data, Cloud Computing, and Data Science (BCD), 2021, pp. 98-103, doi: 10.1109/BCD51206.2021.9581866.

[11] R. Kothari, B. Jakheliya and V. Sawant, "Implementation of A Distributed P2P Storage Network," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298375.

[12] A. Dave, B. Patel, G. Bhatt and Y. Vora, "Load balancing in cloud computing using particle swarm optimization on Xen Server," 2017 Nirma University International Conference on Engineering (NUiCONE), 2017, pp. 1-6, doi: 10.1109/NUICONE.2017.8325618.

[13] John Savill, "The Cloud and Microsoft Azure Fundamentals," in Microsoft Azure Infrastructure Services for Architects: Designing Cloud Solutions , Wiley, 2020, pp.1-46, doi: 10.1002/9781119596608.ch1.

[14] Dan Sullivan, "Overview of Google Cloud Platform," in Official Google Cloud Certified Associate Cloud Engineer Study Guide , Wiley, 2019, pp.1-14, doi: 10.1002/9781119564409.ch1.

[15] Online (https://coinmarketcap.com/), Accessed: 06.06.2022

[16] Shabana R, "Top Cloud Service Providers in 2021: AWS, Microsoft Azure and Google Cloud Platform", Online ("https://blog.cloudthat.com/top-cloud-service-providers-in-2021-aws-microsoft-azure-and-google-cloud-platform/"), Accessed:30.07.2022.

[17] E. A. Kanimozhi, M. Suguna and S. Mercy Shalini, "Immediate Detection of Data Corruption by Integrating Blockchain in Cloud Computing," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), 2019, pp. 1-4, doi: 10.1109/ViTECoN.2019.8899394.

[18] N. Gol Mohammadi, Z. Á. Mann, A. Metzger, M. Heisel and J. Greig, "Towards an End-to-End Architecture for Run-Time Data Protection in the Cloud," 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2018, pp. 514-518, doi: 10.1109/SEAA.2018.00088.

[19] I. Miers, C. Garman, M. Green and A. D. Rubin, "Zerocoin: Anonymous Distributed E-Cash from Bitcoin," 2013 IEEE Symposium on Security and Privacy, 2013, pp. 397-411, doi: 10.1109/SP.2013.34.

[20] N. Lu, Y. Chang, W. Shi and K. -K. R. Choo, "CoinLayering: An Efficient Coin Mixing Scheme for Large Scale Bitcoin Transactions," in IEEE Transactions on Dependable and Secure Computing, vol. 19, no. 3, pp. 1974-1987, 1 May-June 2022, doi: 10.1109/TDSC.2020.3043366.

[21] Xuanwu Zhou, "Study on ring signature and its application," 2009 Chinese Control and Decision Conference, 2009, pp. 1388-1393, doi: 10.1109/CCDC.2009.5191586.

[22] X. Song and Y. Wang, "Homomorphic cloud computing scheme based on hybrid homomorphic encryption," 2017 3rd IEEE International Conference on Computer and Communications (ICCC), 2017, pp. 2450-2453, doi: 10.1109/CompComm.2017.8322975.

[23] J. Kim and A. Yun, "Secure Fully Homomorphic Authenticated Encryption," in IEEE Access, vol. 9, pp. 107279-107297, 2021, doi: 10.1109/ACCESS.2021.3100852.

[24] D. Dawson, "The Future of Data Storage", Url ("https://circleid.com/posts/20220107-the-future-of-data-storage"), Accessed: June, 2022.