SECURE BIG DATA AND IOT WITH IMPLEMENTATION OF BLOCKCHAIN

Daniela Mechkaroska, Aleksandra Popovska-Mitrovikj and Vesna Dimitrova

Faculty of Computer Science and Engineering,
University "Ss Cyril and Methodius" - Skopje, P.O. Box 393, R. of Macedonia
emails:{daniela.mechkaroska, vesna.dimitrova, aleksandra.popovska.mitrovikj}@finki.ukim.mk.

Abstract: *BlockChain is a distributed database of records or public ledger of all time stamped transactions saved in all computers in one peer-to-peer network. It allows a secure and transparent transfer of digital goods including money and intellectual property. Bitcoin - a digital decentralized cryptocurrency, is the first application of BlockChain. The second application is an agreement called Smart contract that enables exchanging a value or assets between two owners based on a set of conditions included in the contract.*

In this paper, we analyze the possibilities for application of BlockChain in Big Data and IoT. Implementation of BlockChain in Big Data confirms that data is accurate and secure and sharing of data will become more simple. In industries like financial services, government and healthcare there is a need to combine BlockChain and Big Data because these industries have repositories full of important data. They must store and share these large amounts of data. Implementation of BlockChain technology provides security of data and ensures its integrity when shared. BlockChain technology is also seen as a way to secure the Internet of Things (IoT). Application of BlockChain in IoT enables IoT devices to participate in BlockChain transactions and invents new styles of digital interactions. This technology will provide a simple infrastructure for devices to directly and securely transfer data or money using Smart contract.

KEYWORDS*: BLOCKCHAIN, BITCOIN, SMART CONTRACT, PEER-TO-PEER NETWORK, IOT, BIG DATA.*

1. Introduction

BlockChain is a distributed ledger that records transactions in blocks that form a chain. This chain is secure, immutable, and transparent. Therefore, solutions that use BlockChain technology can be a pillar of data handling and processing systems within organizations. Integrating BlockChain with Big Data has many advantages because it enables better management of enormous volumes and variety of information.

Big Data and BlockChain are correlative: Big Data has processing capacities that can handle BlockChain complexities and its big expansion and vice versa. Implementation of BlockChain in Big Data confirms that data is accurate and secure and sharing of data will become more simple.

First, in this paper, we explain what the term Big data means, where they are used, and which tools are used to process them. Then, we will briefly explain the basic principle of work of BlockChain technology.

The rest of the paper is organized on the following way. In section 2 and 3 we explain what is Big Data and BlockChain technology. Next, in section 4, we consider the benefits of application of BlockChain in Big Data and Big Data analytics. Also, in this section, we analyze several implementations of BlockChain technology in Big Data: in the financial service industry, in industries outside of banking and in Supply Chain Monitoring and we give some examples for improving Big Data using BlockChain. In section 5 we investigate benefits of application of BlockChain in IoT. At the end, we give some conclusion.

2. What is big data?

Big Data is a term used for a collection of large and complex data sets. For example, Facebook has collected 300 petabytes of personal data since its inception, as they are the records kept, messages sent, videos published, transaction data from online transactions and shopping among many other sources of big data.

It is hard to process and analyze Big Data with traditional techniques, because it refers to a big volume of structured and unstructured data. An important role in collecting, storing and processing these big amounts of data has the growth of cloud storage in helping companies. Tools like Hadoop, Plotly, Bokeh, etc. can provide massive storage for any kind of data and enormous processing power, internet of things and unstructured sources. Studies show that 2.5 quintillion bytes of data are generated daily.

Despite such a contribution to the effective management of large amounts of data, there is a growing concern about user privacy and big data security. There are many examples that show it, like the large-scale scientific experiment conducted by Facebook without informing its users explicitly or the government has often come under attack for its observation on its citizens without their explicit permission. Today the great businesses are careful in accepting big data, in order to ensure their security and privacy [1].

3. BlockChain technology

BlockChain is a decentralized ledger of transactions, distributed on all computers in one peer-to-peer network where all details of transactions are visible to everyone connected to the network. Namely, this is a growing list of linked blocks. The blocks consist of valid transactions, a timestamp and a hash pointer as a link to the previous block in the chain.

The sequence of linked blocks creates a secure, interdependent chain. Blocks must first be validated to be added to the BlockChain. When a block is verified, it is distributed through the network (added to the existing chain) and each node adds the block to the majority BlockChain. Then the transaction is completed [2].

Three data attributes that are in BlockChain, and which are rarely provided in conventional centralized data management are [3]:

- **Security -** It is almost impossible for data to be changed or corrupted if BlockChain is used.
- **Integrety -** BlockChain data offers audit trails, certainty of origin, and digital signature of the message also provides integrity through the transfer.
- **Value -** Data generated with BlockChain is complete and there is no doubt for its value.

4. Benefit of application of BlockChain in Big Data

Regardless the variety, velocity and volume of data, Big Data tools can process it and BlockChain enables transparency and simplicity to processes in every area. The importance of Big Data and the development of BlockChain technology in the last years enables abandoning of the old information processing structures and business transaction processing. Big Data and BlockChain are correlative: Big Date has processing capacities that can handle BlockChain complexities and its big expansion and vice versa [4].

Some of the benefits of using BlockChain technology in Big Data analytics are [5]:

- reduces costs (significantly reduces storage costs);
- increasing traceability (each product or document has "digital password" which ensures tracking of its origin and journey);
- enhanced data quality (data is complete and structured, weak points in a big data analytics which increase accuracy and makes analysis easier);
- facilitates data access (users from different departments can access the data for the analysis process and this shortens the time cycle of data access and analysis);
- enhancing security (the system is decentralized and transparent, so the risk of fraudulent activities is reduced).

4.1. Analyses of several implementations of BlockChain technology in Big Data

a) Speeding up the financial service industry

The connection of BlockChain and Big Data in financial institutions will allow to estimate risk and identify suspicious patterns in real time. Using BlockChain as a means of conducting transactions will help to protect banks and their customers from fraud, to speed up the process of transactions and reduce the cost of money transfers [6].

For example, in the last few years, in order to simplify money transfers between bank accounts using BlockChain technology, an organization of 47 Japanese banks joined a BlockChain startup called Ripple. The aim of this is to perform real-time transfers at a lower cost. The traditional real-time transfers are expensive, because of the potential risk factors as double spending that can be avoided with BlockChain technology. Big Data analyses together with BlockChain can identify risky transactions faster than the current ones. This reduces the cost with real-time transactions [7].

b) Security in industries outside of banking

To handle data and prevent hacking and data leaks, healthcare, public administration, enterprises have started using BlockChain.

For example:

- **Healthcare**: a project at the MIT Media Lab known as Medrec is starting to create BlockChain system that gives priority to patient agency, which authorizes patients to share their records. In order to make the sharing of permissions to access data easier, Medrec uses a private Ethereum chain [8].

- **Insurance:** A controlled master policy from the UK, and three local policies from the U.S., Singapore and Kenya, have been written into a "smart contract" that enables a shared insight into policy data and documentation in real-time. Clarity into coverage and premium payment at the local and master level is provided by Blockchain. It also enables automated alerting to participants in the network after payment events [9].

c) Supply Chain Monitoring

The goods included in Supply Chain are added to the BlockChain and a Mobile App is used to monitor the status of the goods while being transported. The benefit of using BlockChain in Supply Chain are [10]:

- reduce or eliminate fraud and errors;
- improve inventory management;
- minimize courier costs;
- reduce delays from paperwork;
- identify issues faster;
- increase consumer and partner trust.

For example, in order to improve food safety by increasing the monitoring of products from the place of origin to the time it is sold to the consumer, Walmart uses BlockChain technology. In this way, users get credible insights of the origin of food. Receiving unchangeable, credible and traceable data is of great importance to success of Walmart's operations, because it produces 40 petabytes of data every day [1].

4.2. Examples for improving Big Data using BlockChain

In this subsection we present several examples of implementation of BlockChain in Big Data.

- Example 1: GOLEM

Golem network is a decentralized network of computers that utilizes computer power of the network users to make the supercomputer functional. Laptops, desktops, and data centers are just a few examples of devices that can contribute computing power to the network. A decentralized sharing economy is the form of the network and the computing power is shared. In this way, people can make money from renting their computer power and software. They are also able to lower the costs and not to depend on a centralized company for computing services for running some programs [11].

- Example 2: PATH

Path aspires to unite both BlockChain and Big Data allowing users to rent their extra bandwidth. This comes as a result of the fast acquiring of BlockChain technologies among people and companies. "Path Mining Nodes" are installed onto companies' computers and they work passively in the background - earning tokens for providing insights into performing of their websites or the loading time of their applications etc., back to Path and the clients they service. The real value of BlockChain data is the quality and how it is stored on the public ledger. Thus, this platform, together with the others BlockChain platforms that help companies to improve data, can become middlemen between companies and the users [12].

- Example 3: ESTONIA

Keyless Signature Infrastructure (KSI) is a BlockChain technology designed in Estonia to ensure that network and data are free of compromise and retain data privacy. KSI is used to store public data in BlockChain by Estonia, which is striving to be the world's most advanced digital society. Using this, the government can observe any changes in the database thus ensuring that the data is transparent. This has two benefits: cuts down external falsified/tampered records, and makes harder for unauthorized government officials to interfere with information and data [1].

- Example 4: SKRY

Financial institutions, law enforcement, and Bitcoin companie*s can detect suspicious activities using the first product of Skry. It can also identify legal and illegal entities, which enables following regulations and investigation of cyber-crimes like ransomware extortions, terror financing or drug trafficking on the dark web.

5. Internet of Things (IoT) and BlockChain

Application of BlockChain in IoT enables IoT devices to participate in BlockChain transactions and invents new styles of digital interactions. This technology will provide a simple infrastructure for devices to directly and securely transfer data or money using Smart contract.

5.1. Challenges of a secured IoT model

The current IoT ecosystem is based on a centralized model – client/server model, where all devices are identified, authenticated and connected through cloud servers. This structure is the biggest challenge of IoT security. Even, if the devices are at a short distance the connection between them will have to go through the cloud. These models continue to be used in today IoT networks, although they might not be able to handle the growing needs of giant IoT ecosystems in the future. The costs for existing IoT solutions, that use centralized model, are another big obstacle for the future growth of the IT network. The clouds, large server farms, and networking devices have higher costs for infrastructure and maintenance. Also, in order to protect IoT devices and platforms form physical tampering, new security technologies are needed. In addition, many devices in IoT system use simple operating systems and processors that not support advanced security systems [13].

BlockChain in IoT can be used in tracking connected devices, processing of transactions and coordination between devices. The decentralized nature of BlockChain technology will produce a more flexible system for running the devices and with the cryptographic algorithms data will be more private. The BlockChain ledger cannot be manipulated by malicious users since it does not exist in any single location, which means that it is a tamper-proof. Also, BlockChain enables secure peer-to-peer messaging between IoT devices. Capabilities of BlockChain, such as decentralization, autonomy and confidentiality, make this technology ideal for a fundamental component of IoT solutions [13].

From this, we can conclude that the main benefits of using BlockChain technology in IoT system are:

- build trust
- reduce costs
- accelerate transactions.

5.2. Example of IoT and insurance

Usage-based insurance (UBI) models are some of the new products that will develop accurate actuarial models which is valuable to insurers. For example, in the auto insurance, we can use encrypted data for driving times, distances, acceleration, braking and other behaviors for identification of high-risk drivers and validation of the information included in the applications. On this way, we can provide consumers bigger control over their premiums. The question is how to manage the enormous volume of data and logic because of the communication between millions devices. Large complex networks can be managed by BlockChain with devices communicating between each other on a peer to peer basis, securely and accurately, instead of building an expensive data center. This type of managing devices is cheaper than the data central model [14].

Conclusion

In this paper, we analyze benefits of using BlockChain technology in Big Data and IoT and possibilities for improving security. We present several examples of implementation of BlockChain technology in Big Data and IoT (in the financial service industry, in industries outside of banking, in supply chain monitoring, auto insurance, …).

From the presented analyses we can conclude that the main benefits of using BlockChain in Big Data are: reduced costs, increased traceability, enhanced data quality, facilitated data access and enhanced security.

For IoT the main gains from implementation of BlockChain are: build trust, reduced costs and accelerated transactions.

References

[1] https://medium.com/ethereum-dapp-builder/how-blockchain-will-improve-big-data-da1547dd268

[2] Mechkaroska D., Dimitrova V., Popovska-Mitrovikj A.: A Survey on Applications of BlockChain Technology, Proc. of the 15th International Conference on Informatics and Information Technologies CIIT, 2018 (in print)

[3] https://jaxenter.com/blockcahin-big-data-146218.html

[4] https://blog.dataone.io/big-data-and-blockchain-c84c6c02544c

[5] https://www.forbes.com/sites/bernardmarr/2018/02/07/5-blockchain-opportunities-no-company-can-afford-to-miss/#9893de51a83d

[6] https://jaxenter.com/blockcahin-big-data-146218.html

[7] https://www.kdnuggets.com/2017/09/introduction-blockchain-big-data.html

[8] Casey M., Crane J., Gensler G., Johnson S., Narula N.: The Impact of BlockChain Technology on Finance: A Catalyst for Change, Geneva, International Center for Monetary and Banking Studies 2, 2018.

[9] https://www.datanami.com/2018/06/28/blockchain-solutions-usher-in-era-of-trusted-big-data/

[10] https://www.ibm.com/blockchain/industries/supply-chain

[11] https://blockchain.wtf/cryptocurrency-blockchain/erc20-tokens-projects/golem/

[12] https://www.inc.com/nicolas-cole/web-analytics-will-get-a-massive-improvement-with-blockchain-technology.html

[13] https://www.bbvaopenmind.com/en/a-secure-model-of-iot-with-blockchain/?utm_source=views&utm_medium=article06&utm_campaign=MITcompany&utm_content=banafa-jan07

[14] https://www.ey.com/Publication/vwLUAssets/EY-blockhain-in-insurance/$FILE/EY-blockhain-in-insurance.pdf