Secure University Decentralized Data Storage Solutions

Krisztián Bálint

Óbudai University, Keleti Károly Faculty of Business and Management,
balint.krisztian1@uni-obuda.hu

Abstract The secure storage of university data is crucial these days, as it involves the storage of sensitive data. Everything must be done to protect the personal data of students and the data of university employees. The university's reputation may depend on outsiders having access to this data. In the past, centralized data storage solutions did not provide sufficient protection, so it is advisable to investigate decentralized blockchain-based data storage options. In the blockchain, data is stored as blocks, so each block contains the previous block's hash value. For a hacker to modify the blockchain, he must change all the blocks going back to the Genesis block, which is a problematic solution depending on the size of the blockchain. DDoS attacks can also be easily avoided by using a university blockchain since the data is not stored centrally on a server but in different nodes.

Keywords: Blockchain, Database security,

1 Introduction

In the future, database security will become even more prominent as data volumes continue to grow. With the advent of IoT, Internet communication is no longer only between people, but also smart devices can communicate with each other. The servers can thus slowly become overloaded, and serious computer capacities will be needed to process the data.

The 5G network is currently being built, which can open new horizons in communication. It is easy to see that without adequate database security, 5G solutions can be at risk. Educational institutions must prepare for these threats. It is not only important to store university data, but also the personal data of students and employees that universities have. Thanks to modern eLearning solutions, teaching materials can no longer be prepared only in the form of presentations, but can also be made more colorful by means of video recordings. Establishing a secure and fast Internet connection between the university and the students should be an important aspect, even in the case of increased data volumes. In the future, educational institutions may be threatened by the following new types of threats:

- Malware can become more difficult to identify as it learns to mimic user behavior. This will make them more difficult to recognize,
- Traditional cloud-based centralized solutions can increase the risk of possible successful attacks, as attackers have been able to successfully identify their weak points until now,

Automated attack solutions using artificial intelligence may appear. These intelligent attacks will presumably be more difficult to defend against [1], that's why it's important to check which are currently the most secure data storage solutions.

2 Types of Data Storage Solutions

Storing data on physical devices is well known to everyone. Recently, data stored in the cloud has opened new horizons. Cloud-based solutions have already proven their advantage through easy access and synchronization, which facilitate data capture. In order to secure university data storage, it is necessary to examine both centralized and decentralized data storage solutions. In practice, data storage looks like this:

- Saving on physical devices. This can be the hard drive of the NVR, as well as pen drives and various discs (CD, DVD, Blue-Ray),
- Data storage in a centralized cloud. The data is stored in the cloud, which has a separate owner and operator. The server operator is responsible for data security.
- In the case of decentralized cloud-based data storage, the data is stored on a decentralized network. The data is not saved on a company's server, but on computers operated by independent individuals in many parts of the world. You can connect to such networks through smart contracts.

2.1 Centralized Data Storage Solutions

In the case of centralized cloud data storage, there have been a number of undesirable events in the past that may cause concern. Data leakage has occurred in many cases. In the case of university databases, the goal is to achieve the highest possible security, so it is necessary to examine the available data storage options and to search for new solutions that can provide the most effective security possible. Centralized data storage solutions are characterized by:

- Increasing storage costs. With the advent of IoT and Industry 4.0, the amount of data has increased significantly, requiring more bandwidth and storage capacity.
- Censorship and surveillance. Many people are concerned about the fact that others can observe, read, and in some cases even modify their data.

- As a result of DDoS attacks, the servers of many large companies were attacked. With decentralized storage, this cannot happen, since there is no central server that can fall victim to a directed attack. [2]

The local storage of data is more and more in the background, because as the amount of information increases, so do the amounts of data. The capacity of hard drives is limited, and access to local computers from a distance is difficult. Thanks to the centralized cloud-based solution, personal data can be accessed from almost anywhere, as long as an Internet connection is required. As a result of centralized storage, data loss due to hardware errors can be reduced, as backup copies are saved not only on the hard drive, but also in the cloud. The figure below illustrates the centralized data storage solution:

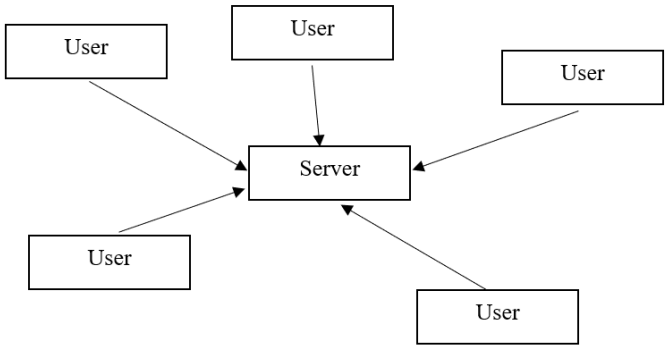

Figure 1
Centralized data storage solution

In the case of cloud-based data storage, the basic assumption is that the third party is a reliable service provider whose goal is to ensure that the data is always secure and available 24 hours a day. It may happen that the third party damages the data for their own personal benefit. You can modify them, release them to different bodies, or delete them. In order to prevent this, several cloud-based storage platforms are used at the same time. The disadvantage is that this method generates a lot of network traffic and bandwidth. [3]
In order to protect the data from third parties, it is recommended to encrypt it and upload it to the cloud. Centralized storage systems have the following weaknesses:

- Safety. If an unauthorized person has access to the server data, it can be compromised.
- Reliability. The server may become overloaded if too many requests are received at once. This is how DDoS attacks work.
- Data transfer speed. A fast connection to the server is essential. If users' computers are in different countries (which is usually the case), the data transfer speed may decrease, and some countries may impose restrictions.

- Scalability. Due to the centralized design, the capacity of the server is limited, and the data traffic is also regulated [4].

2.2 Decentralized Data Storage Solutions

Thanks to the decentralized data storage solution, the data is more secure than in the case of cloud-based storage, since it is distributed across many nodes. Furthermore, storage systems use public key encryption. Data is distributed flexibly between nodes and smart contracts are also used for automatic execution [5]. A Advantages of decentralized data storage:

- Performance is balanced as nodes share data volumes proportionally,
- High availability. Most of the hubs are available 24 hours a day. If some nodes become unavailable, the other nodes will continue to serve the user.
- A high degree of independence. Each node is independently responsible for following the rules, thus forming the blockchain ecosystem. An outside person or authority does not limit or regulate its operation.
- It divides the users' data into several pieces and then encrypts it to the nodes. In the event of a DDoS attack, the system remains operational.
- If some nodes do not work or become unreachable in the event of an attack, the other nodes can continue to function without interruption. In the centralized system, if the central server stops, the whole system will most likely be inoperable, therefore the data cannot be accessed [6].

Disadvantages:

- Due to the lack of central supervision, there is no chain of command that could give orders to perform various tasks,
- The so-called "usual" regulatory oversight is missing. The creator of the private blockchain defines the rules that are enforced by the smart contract. This is difficult to apply in some cases
- Determining which node has failed is critical, as each node must be checked,
- It is difficult to determine which node responded to the request since the same data is available on several nodes through a decentralized system [7].

In conclusion, it can be stated that in the case of decentralized data storage, the data is stored completely independently of each other on many nodes, which is a safety-enhancing factor, and therefore this type of blockchain technology is suitable to be part of the university data recording in the long term. Therefore, it is advisable to further investigate blockchain technology.

2.3 Blockchain technology

With the appearance of large amounts of data (Big Data), networks often become overloaded. Due to their division, blockchains enable more efficient data processing

and cost reduction. And by distributing the processing, they positively influence the Internet of Things (IoT) [8].

In their view, it will become technology such as the steam engine, energy supply, information, and Internet technology [9].

The first block of the blockchain is the genesis block, on which the rest are built. After that, each block is connected to the previous so-called parent block. The block consists of a header and a body. Their structure is as follows:

- Block version: contains the rules necessary to validate the block,
- Parent block hash: this is a 256-bit value that always points to the previous block. Without this, the chain could not be created,
- Merkle tree root extract: forms the extract of all transactions of all blocks,
- Timestamp: current timestamp in seconds. This is necessary for authentication,
- nBits: the current hash value expressed in a compact format,
- Nonce: 4-byte field that starts with 0 and continuously increases during hash calculations. [9] The figure below shows the structure of the blockchain.

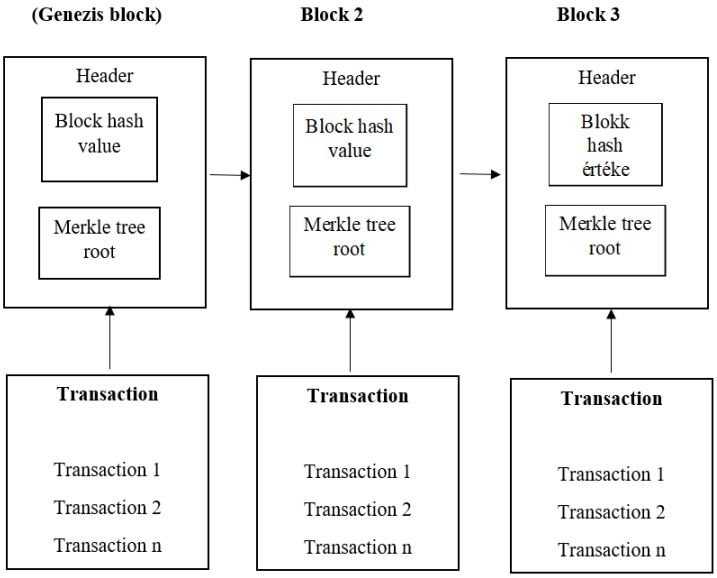

Figure 2
Blokchain strucutre [10]

3 Creation of a Universtiy Blockchain

In the case of the constitution of a self-sufficient, faculty-based blockchain, the educational institution may determine the advantageous and convenient conditions of data storage. These may be the following:

- Broader access to the blockchain in question,
- The definition of the size of blocks,
- The definition of terms of use,
- The original block (genesis block) to which will all the other blocks will connect, stays in the domain of the faculty,
- The limiting of access to the blockchain (only the authorized persons may use it),
- The definition of data protection policy,
- The blockchain may be started on multiple servers to uphold the security,
- The nodes are more easily monitored,the system will become more transparent,The eventual data compromising will be more easily identified [11], [12].

When creating the university blockchain called UDSC (Universities Data Storage Chain), the first step is to create the genesis block. The creation of the genesis block is shown in the third figure.

```
{
  "config": { // the config block defines the settings for our custom chain and has certain attributes to create a private blockchain
    "chainId": 987, // identifies UDSC blockchain
}
    "homesteadBlock": 0, // Homestead version was released with a few backward-incompatible protocol changes, and therefore requires a hard fork. UDSC chain however won't be hard-forking for these changes, so leave as 0
    "eip155Block": 0, // Homestead version was released with a few backward-incompatible protocol changes, and therefore requires a hard fork. UDSC chain however won't be hard-forking for these changes, so leave as 0
    "eip158Block": 0
  },
  "difficulty": "0x400", // This value is used to control the Block generation time of a Blockchain. The higher the difficulty, the statistically more calculations a Miner must perform to discover a valid block
  "gasLimit": "0x8000000",
  "alloc": {}}
}
```

Figure 3
Creating a genesis block [13]

The faculty-based blockchain may be constituted in the following manner, presented in the third figure. This blockchain called UDSC needs to be created for the purpose of storing the university teaching materials

.

University chain-util generate UDSC

the default settings would be used:

/default ~ university chain/UDSC/chainsettings.dat

chainsettings.dat include:

Database addresses [receiver (cloud storage) IP address, sender (university) IP address],

Database system addresses [receiver (university database) IP address, sender IP address],

Terms of GDPR database.

Next, the UDSC blockchain would be initialized, and the genesis block would be created

universitychain UDSC

The server will be started in those few seconds after the genesis block has been found, then the node address needs to be connected:

UDSC@192.168.0.1:8008

After these steps, the connection can be attempted from a second server:

universitychain UDSC@192.168.0.1:8008

After the message confirming the chain has been initialized, permission is not given for connection to the database. The address would be copied and pasted: 192.168.0.2

finally, permission for connection would be granted:

universitychain UDSC grant 192.168.0.2 connect.

Figure. 4
The creation of an UDSC blockchain [11,12,13]

Conclusions

In order to securely store university data, it is recommended to search for new solutions instead of already proven and convenient solutions. The advantage of centralized data storage is that it is fast and easily accessible. The data stored in the cloud can be kept up-to-date by the continuous synchronization of the devices, however, thanks to the central server, the system can easily become vulnerable.

The data stored in the blockchain can also withstand traditional DDoS attacks, since the data is not stored in one place, but in different nodes, which means that a hacker attack focused on one place cannot be launched against it.

However, for the sake of long-term and secure data storage, it is advisable for universities to think about creating their own blockchain, which allows them to regulate access to their own blockchain. By applying this method, they could have blockchain-based administrator rights, which could even form the basis of a secure data storage at the university level in the future.

References

[1] Deloitte: Protecting against the changing cybersecurity risk landscape. https://bit.ly/3k1gWEo (accessed: 2023.04.23), 2022.

[2] Medium: What Is Decentralized Storage? https://bit.ly/2UK9jaG https://bit.ly/3k1gWEo (accessed: 2023.04.23), 2022.

[3] Huige Li-Fangguo, Zhang Peiran, Luo1-Haibo, Tian1-Jiejie He: How to retrieve the encrypted data on theblockchain, KSII Transactions on Internet and Information Systems vol. 13(11), Nov. 2019, pp. 5560-5579.

[4] Andrew Tar: Decentralized and Distributed Databases, Explained. https://bit.ly/2ZEBqtl (accessed: 2023.04.23), 2022.

[5] Peng Jiang-Fuchun Guo-Kaitai Liang-Jianchang Laib-Qiaoyan Wen: Searchain: Blockchain-based Private Keyword Search in Decentralized Storage, Elseiver, 2017, pp. 781-792.

[6] Amer Rosic: Centralized vs Decentralized Storage, Redefining Storage Solutions with Blockchain. https://bit.ly/2NxIrX9 (accessed: 2023.04.23), 2022.

[7] Geeks for Geeks: Comparison – Centralized, Decentralized and Distributed Systems. https://bit.ly/3dYwLYw

[8] AMBRUS Éva: Blockchains, Hadmérnök, XII issue, 2. 2017, jun. 2017, pp. 1-11.

[9] Azaria Asaph: Medrec: Using blockchain for medical data access and permission management; 2016 2nd International Conference on Open and Big Data (OBD). IEEE, 2016, 99.pp. 25-30.

[10] Tanzeela Sultana Ahmad, Almogren Mariam, Akbar Mansour Zuair Ibrar Ullah, Nadeem Javaid: Data Sharing System Integrating Access Control Mechanism using Blockchain-Based Smart Contracts for IoT Devices Applied Sciences 10(2) 2020 pp. 1-21.

[11] Bálint Krisztián: The connection of a Blockchain with Students' Attendance Register based on Security Cameras, IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY 2021), Subotica, Serbia, 2021, 191 p. pp. 67-70.

[12] Bálint Krisztián: Data Security Structure of a Students' Attendance Register Based on Security Cameras and Blockchain Technology, IEEE Joint 22nd International Symposium on Computational INTELLIGENCE and Informatics and 8th International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo 2022) : Proceedings, Budapest, Hungaray, 2022, 41pp. 185-189.

[13] Bálint Krisztián: Modern, Decentralized Blockchain-Based Solutions for Saving Video Footage, IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY 2020) Danvers (MA), Amerikai Egyesült Államok: IEEE, 2020, pp. 11-14.