# EXCHANGING PERSONAL DATA THROUGH BLOCKCHAIN TECHNOLOGY TO FULFIL INCREASING PRIVACY NEEDS

Petrus J. Roos and Henri E. van Rensburg
*North-West University, Potchefstroom, South Africa*

**ABSTRACT**

The risk of privacy intrusion is a developing concern due to the advancement in data mining techniques and the increased value of data. This study focuses on developing a system that implements blockchain technology to allow the exchange of personal data while ensuring the authenticity and privacy of the exchanged data are preserved. The privacy of data is ensured using encryption while the authenticity is intuitively built into blockchain technology with the consensus algorithm and cryptographic signatures. The blockchain interaction was enhanced by developing a companion application that simplifies exchanging data over the blockchain using QR codes. The expectation is to create awareness of the potential for using blockchain technology for secure-, private-, but verifiable data exchange, the possible implementations for future systems, and to reduce negative sentiments in regards to data sharing.

**KEYWORDS**

Blockchain, Privacy, Secure, Cryptography, Personal, Data

## 1. INTRODUCTION AND BACKGROUND

The risk of privacy intrusion is a developing concern due to the advancement in data mining techniques and the increased value of data. Blockchain technology can be used to prevent this issue. Blockchain technology is a distributed ledger where each record is encapsulated in a block, each block is connected to a previous block forming a chain (Heister & Yuthas, 2020:2).

The focus of this paper is to develop and implement a system where the capability of blockchain technology is tested by securely exchanging personal data. The main focal point is ensuring the authenticity of data stored in the blockchain, along with preserving the privacy of data exchanged to third parties. The proposed blockchain system enables a fast and effective exchange of personal data through the use of QR codes to authenticate trustworthy third-parties and consent to share personal data with.

Facebook disclosed the data of 87 million users to Cambridge Analytica for user profiling, which they used on various social media platforms to persuade voters in the United States elections (Isaak & Hanna, 2018:56-58). With more than 87 million people affected through the action of two companies, it indicates that personal data is not protected. Identity theft is when personal information is used to impersonate an individual such as to open a credit account leaving the victim many times unknowingly with debt (Holvast, 2007:23). Moreover, identity theft affects up to 7% of the United States population with the amount of money embezzled reaching figures up to 16 billion US dollars (Graves & Sexton, 2017:217). The Centre for International Governance Innovation (CIGI-Ipsos, 2019) conducted a survey which stated that 75% of participants claimed that social media are the main cause for skepticism in internet privacy with 45% of participant claiming that this skepticism prevented them from disclosing personal information online. In retrospect, the intrusion of personal data poses a great threat to the trustworthiness of respectable businesses in the Information Technology (IT) sector. Legislations are put in place for a reason however, as demonstrated, powerful companies does not follow these legislations, thus alternative measures must be taken. Blockchain technology can perhaps prevent the intrusion of privacy, encouraging better trust in IT businesses and professionals.

The key features of blockchain technology comprise out of immutable blocks and a decentralized network (Haughwout, 2017). This means that, once a block is added to the blockchain and distributed across the network

it cannot be removed. Haughwout (2017) also stated that with each addition to the chain, a timestamp is added to the block as well as a cryptographic signature known as a hash which is then used to verify the validity of the chain. Figure 1 illustrates three blocks in a blockchain. The blocks are connected by storing the previous hash in the block similar to a linked list. Suppose that a block is manipulated, the hash is then recalculated and will not correspond to the next block anymore declaring the whole blockchain null and void. Figure 1 also demonstrate that each block has a timestamp represented in epoch Unix time and is calculated once a block is added to the chain. Moreover, in Figure 1 the cryptographic signature is calculated using the participant's public key to encrypt the hash forming a digital signature to enhance the authenticity of each block added to the chain. Blocks are added to the blockchain in a process called mining (Bashir, 2018:167). The last observation from Figure 1 is a simple example of personal data to be stored in the blockchain.

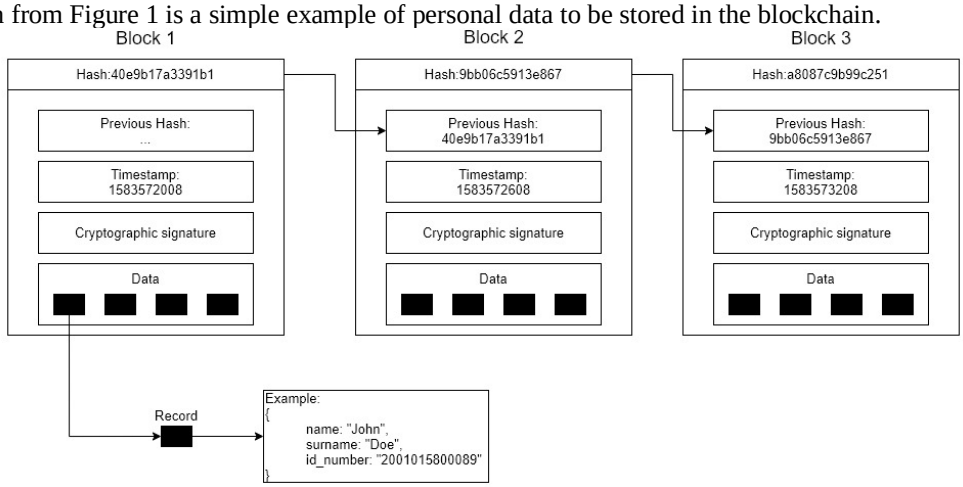


Figure 1. A graphical representation of blocks in a blockchain containing personal data

By utilizing blockchain technology, data stored in the blocks are immutable and authentic. However, there is an issue storing data in a general blockchain since data is publicly available. Hence a lightweight algorithm should be implemented to encrypt and decrypt data asymmetrically allowing only the participating users to provide access to their data. Therefore the proposed system implements blockchain technology for data exchange to allow participants to exchange personal data while ensuring that the authenticity and privacy of exchanged data are preserved.

## 1.1 Data Privacy and Protection

The data-driven world which emerged in the last few decades created platforms which saw a rise in privacy issues. These issues range from the illegal selling of personal data to big corporate companies and direct crimes such as identity theft and fraud. In the last decade, Facebook was one of the biggest culprits where they sold the data of millions of users (Isaak & Hanna, 2018:56-58).

The most prominent implementation of privacy protection is to implement legislation to penalize companies after infringing these laws. The European Union enforced legislation called the General Data Protection Regulation (GDPR) which protects the privacy of citizen in the European Union (Herrle & Hirsh, 2019). According to Wolford (2020), there are seven principles when processing data which include the essential principle of ensuring the confidentiality and integrity of the data being processed. The penalty of not being GDPR compliant are fines that reach a value of 20 million euros and these fines could also apply to offenders outside of the European Union (Wolford, 2020). Although this is the case for the European Union, not all regions of the world have concrete legislation to protect the privacy of citizens.

According to van den Hoven and Warnier (2020), privacy intrusion might be caused by IT systems, but IT can also help to improve privacy protection. Moreover, development methodologies should cater to privacy protection needs along with implementing tools to enhance privacy such as cryptography to encrypt personal data (van den Hoven & Warnier, 2020).

The legislation enforced to enhance privacy protection will only prevent privacy violations from happening to some extent since companies could still infringe these laws and pay the fines. In contrary, the implementation of a technical solution to enhance privacy protection will provide the required preventative measures to stop it from happening at all.

## 1.2 Blockchain Technology for Privacy

The research proposes the utilization of blockchain technology to fulfil increasing privacy needs to ensure the privacy and authenticity when exchanging personal data. The key features of blockchain technology, as discussed previously, comprise out of immutable blocks that are distributed across the network to ensure the authenticity of the data (Haughwout, 2017); however, it does not ensure the privacy of the data exchanged.

Since the blockchain is distributed, data stored in the blocks should be encrypted to ensure it is kept private. The RSA encryption algorithm is an appropriate light weight encryption algorithm to implement for the blockchain based private data transmission. According to Singh and Chauhan (2017), there are three steps to follow when implementing RSA encryption which includes generating a public and private key, encrypting data using the public key and decrypting using the private key. When a public key is used to encrypt data and a private key is used to decrypt data, it is classified as asymmetric cryptography (Bashir, 2018:80).

RSA encryption is commonly used in blockchain solutions due to it being relatively computationally inexpensive; however, encryption will increase the size of the data (Singh & Chauhan, 2017). According to Bashir (2018:50), the largest downside to blockchain technology is that it can only store a small amount of data; therefore to reduce the amount of data to be stored on the blockchain, it is necessary to apply a compression algorithm to the encrypted data.

# 2. ARTEFACT IMPLEMENTATION

A blockchain based solution is developed to exchange personal data while retaining privacy and authenticity. The key aspects that enable the proposed blockchain implementation to be a viable solution is shortly discussed below.

### *A. Blockchain implementation*

The fundamental components of a blockchain was developed as a starting reference. The object-oriented class structure includes the transaction, the block, and the blockchain. The transaction object stores the sender and receiver's public keys along with a digital signature and in this case, the personal data to be exchanged. The block object stores an index, the previous hash, a timestamp and a list of transactions. This previous hash is a reference to the hash of the previous block, and the timestamp is the date and time of when the block was added to the chain. The blockchain object stores a list of blocks, the open transactions which should still be added to the chain and the public key of the participant.

Additional classes include; the repository, the hash utility, and the verification class. The repository class has two fields; the public- and private keys of a participant. The repository class provides functions to generate, save, and load public- and private keys as well as to sign- and verify transactions. The hash utility class has two functions; the hash string and hash block. The hash string function applies a SHA-256 hash function to a string. The hash block function uses the hash string method to hash a block by serializing the block as a string. Finally, the verification class receives a list of transactions and the previous hash to validate the chain of transactions and to verify the new transaction received from the repository to be added to the blockchain.

### *B. Personal data dictionary*

In the transaction class, instead of representing a single value, a data field is included to use a dictionary as a data type. The dictionary data type stores values as a collection of key-value pairs. Using a dictionary to store personal data provides flexibility when it comes to data storage and exchange. Depending on the purpose of each data exchange, various personal data fields may, or may not be included in the data dictionary of the transaction. The data dictionary provides flexibility of data that can be exchanged over the blockchain such as biographical, address, and medical information.

### *C. Privacy and Encryption*

To ensure privacy of the exchanged data, an encryption algorithm to encrypt the data dictionary was added to each transaction using the public key of the recipient. The recipient will then be able to decrypt the data

using their private key. This was achieved by using the asymmetric keys generated by the *Repository* class to encrypt and decrypt the data dictionary.

The RSA asymmetric encryption displayed a limitation that can only encrypt a fixed size of characters based on the byte size of each character. This limitation was overcome by splitting the data dictionary into a list of strings that can be encrypted and stored on the blockchain. This enabled the recipient to decrypt the list of strings, concatenating these strings-, and converting it back to a data dictionary for the personal data to be viewed.

### *D. Compression*

The literature review asserted that once data is encrypted the size of data increases. Thus a compression algorithm is necessary to reduce the size of the data exchanged on the blockchain. The Python compression module was used to seamlessly implement compression on the serialized data dictionaries. Compression was implemented on the personal data where the serialized dictionaries were compressed before it was passed on to the encryption functions and then decompressed after it was decrypted.

### *E. Network – (consensus algorithm)*

The term consensus refers to the agreement nodes have with each other on what the current state of the blockchain is (Bashir, 2018:35). The implementation of the consensus algorithm therefore ensures that all blocks on the decentralised network are identical. A publish-subscribe pattern was implemented which allowed nodes to communicate seamlessly across the network. Figure 2 displays the working of a publish-subscribe pattern where there are three entities the publisher, the subscriber, and the publish-subscribe server.


Figure 2. Publish-Subscribe architecture

The publish-subscribe server acts as a mediator between the publisher and the subscriber. The subscriber is a user who subscribes to a certain channel to get specific messages from publishers. The publisher is a user who publishes messages to a channel the publish-subscribe server forwards the message to the subscriber. The blockchain will require all nodes in the network to be publishers and subscribers simultaneously. It is important to add the required test cases to ensure that publishers and subscribers do not echo to each other. In the peer discovery, two test cases is added to verify if the sender is itself or if the node is added to the peer node list yet to prevent an echo loop to occur. The peer node list can now be used to broadcast new transactions, new blocks, and resolve nodes that are not up to date with the main chain.

### *F. User interaction interface / Companion application development*

A web user interface and a companion app is created to facilitate interaction with the blockchain by generating a QR code to represent the public key or scanning the QR code when a transaction is performed. The user is then able to add key-value pairs to a list view (data dictionary) and then send the data to the receiver. Figure 3 demonstrates the simplicity of the transaction process of sending private data over the blockchain using the companion application.

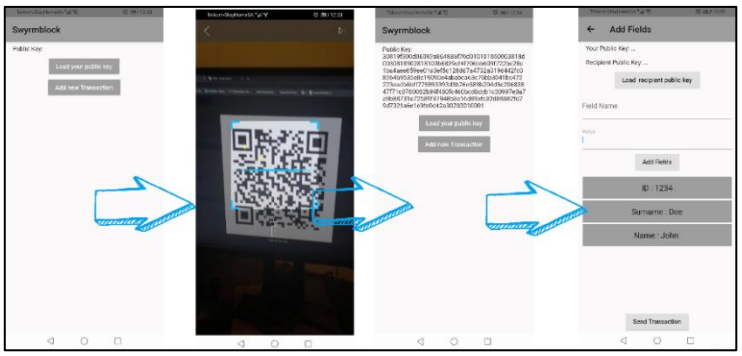


Figure 3. Companion application transaction process

## 3. SUMMARY AND RESULTS

The resulting system provided the ability to send private data using blockchain technology. The implementation of the blockchain makes it possible to exchange data while ensuring authenticity and privacy. Privacy is ensured by using an encryption algorithm to encrypt the personal data added to the blockchain. Authenticity is intuitively built into blockchain technology by making use of the consensus algorithm to ensure that the blockchain is stored identically across the decentralized network. The use of cryptographic signature is added to each transaction that authenticates the sender and receiver. The implementation of blockchain including the exchanging of personal data was programmed using python.

The user interface allows participants to send and receive transactions using the QR codes. After the transaction is sent to a recipient, the encrypted transaction can be viewed in an open transaction tab before it is added to the chain. The decrypted data can be viewed by the recipient using their private key to decrypt. The web application's front end was developed using HTML5, the backend developed using Flask and an API to communicate with the companion app that was developed using Flutter to enable native features such as accessing the camera to scan the public keys represented as QR codes.

## 4. CONCLUSION

The study aimed to implement a blockchain system that allows participants to exchange personal data while ensuring that the authenticity and privacy of exchanged data are preserved. The information gathered from the literature review indicated that encryption is required to ensure privacy of the exchanged data. The encryption size expansion issue was resolved by performing a compression algorithm before it is added to the blockchain in order to reduce the overall size of the blockchain. The companion application simplified the usability by providing an easy to use interface, allowing the exchange of customizable data through the use of QR codes.

In conclusion, this study encapsulates the importance of data privacy followed by a technological implementation of the proposed solution in the form of a blockchain based system that enables the exchange of personal data while ensuring the privacy and authenticity of the data is preserved.

## REFERENCES


Bashir, I. (2018). *Mastering Blockchain: Distributed ledger technology, decentralization, and smart contracts explained.* Second edition. Packt Publishing Ltd.

CIGI-Ipsos. (2019). CIGI-Ipsos Global Survey on Internet Security and Trust. www.cigionline.org/internet-survey-2019 Date of access: 29 February 2020.

Graves, P.E. & Sexton, R.L. (2017). Optimal Public Policy against Identity Theft. *American Economist*, Vol. 62, No. 2, pp. 217-221.

Haughwout, J. (2017). Blockchain: A Single, Immutable, Serialized Source of Truth: Blockchain technology could help introduce higher levels of security to and confidence in supply chain transactions. *Material Handling & Logistics*, Vol. 72, No. 8, pp. 27-29.

Heister, S. & Yuthas, K. (2020). The blockchain and how it can influence conceptions of the self. *Technology in Society*. Elsevier. Vol. 60, pp.101218.

Herrle, J. & Hirsh, J. (2019). The Peril and Potential of the GDPR. https://www.cigionline.org/articles/peril-and-potential-gdpr/ Date of access: 21 May 2020.

Holvast, J. (2007). History of privacy. *The History of Information Security*. Elsevier Science BV. pp. 737-769.

Isaak, J. & Hanna, M.J. 2018. User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, Vol. 51, No. 8, pp. 56-59.

Singh, P. & Chauhan, R.K. (2017). A Survey on Comparisons of Cryptographic Algorithms Using Certain Parameters in WSN. *International Journal of Electrical & Computer Engineering (2088-8708)*, Vol. 7, No. 4, pp. 2232-2240.

van den Hoven, J.B.M.P.W. & Warnier, M. (2020). Privacy and Information Technology. (In Zalta, E.N., ed. The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Wolford, B. (2020). What is GDPR, the EU's new data protection law? https://gdpr.eu/what-is-gdpr/ Date of access: 21 May 2020.