Blockchain and the Importance of Hashing

Kyusub Hwang
Hanyang University
kyusubhwang@hanyang.ac.kr

Emily Bauer
University of Florida
em.bauer@ufl.edu

***Abstract*—Blockchain is a decentralized system which records transactions and distributes the log to its network participants. It is transparent and public, relying on participants in the network to verify those transactions and add them to the blockchain. The process is not so straightforward, and requires a lot of computational power and time, contributing to its security. Many claims have been made about this seemingly infallible security. This paper explores some of the specifics of the blockchain, as well as the hashing algorithm that makes it possible, in order to better understand these claims. Potential weaknesses of blockchain are discussed, along with real instances of fraud that have occurred even with the security measures currently in place. Furthermore, real JSON data containing a block of Bitcoin is analyzed. The paper ultimately comes to the conclusion that the hashing algorithm is the core countermeasure against fraud, and that verifying these hash values is critical. Thus, code is used to verify the hash values, emphasizing just how complex and essential these hashing algorithms are to the mechanism of blockchain.**

Keywords—blockchain, hashing, SHA-256, cryptocurrency

I. Introduction

Bitcoin, and cryptocurrencies in general, has been on the rise over the past decade. Behind these cryptocurrencies and the lofty claims of their superiority to fiat currency lies the blockchain technology. Blockchain is often cited as one of the most revolutionary technologies, especially in terms of security. However, blockchain is both new and complicated, meaning that there is more to consider than simply its popularity and the excitement surrounding it. The weaknesses and instances of fraud related to both blockchain and cryptocurrency must also be explored. Furthermore, a more detailed look into the process of hashing and how it can be verified should also take place.

II. Blockchain and Bitcoin

A. What is Blockchain?

The most well-known application of blockchain technology is Bitcoin, the leading cryptocurrency. The major security behind Bitcoin is based around the use of blockchain technology, so it is an imperative feature. Before understanding Bitcoin, and other blockchain transactions, the basic mechanisms of blockchain must be explained. The base of a blockchain is a single block. "In its simplest form, a block is an encrypted aggregation of a set of transactions" [1]. So, in the case of Bitcoin transactions, the blockchain is able to represent a complete, decentralized ledger of transaction history in an encrypted format that makes fraudulently changing the ledger quite difficult [2]. Each computer in the network has a copy of the ledger. Thus, if a change is made, every copy will be updated.

A blockchain, as the name suggests, is similar to a linked list where each block references the previous block through the use of its hash. The block also contains the hash of its own contents [1]. Thus, changing the contents of any one block would alter the hash of that block, meaning that the reference to that block would be altered in the next block, and so on. This means that all subsequent blocks in the chain will also be altered, invalidating them.

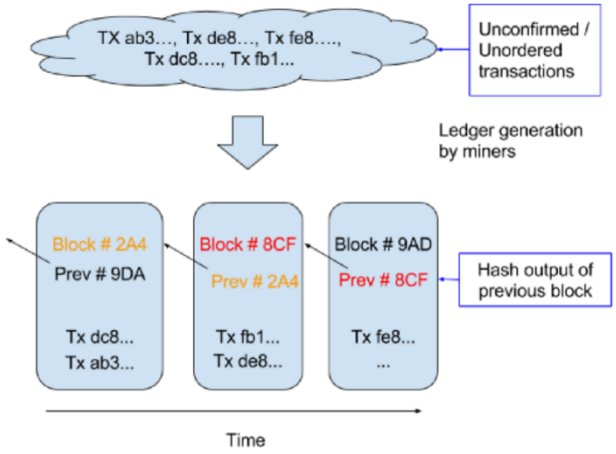

Fig 1. Graphic showing details of the blockchain [14]

B. Proof of Work

An authentic Bitcoin transaction requires a set of public and private keys that are used as a digital signature to verify the transaction [3]. The network can then confirm that this transaction is valid, using information about previous transactions, and the transaction can be written to a block [3]. However, writing to a block is not a straightforward process. The only way to effectively alter a block is to alter that block and all the subsequent blocks, which requires a lot of work [4]. This work comes from the process needed to validate a block within the blockchain, which is done by a network of competing computers called miners. "…Bitcoin makes it computationally difficult to hash a block, by requiring that the resulting hash have specific numeric properties" [5]. This is known as proof of work. More specifically, in order to validate the block, the miners must find a special number called a nonce via random guessing. The correct nonce inserted into the hash function will create a certain desired hash value, usually ones starting with a certain number

of zeros. Finding the correct nonce takes quite a bit of computational work, as well as time. Thus, having to do this multiple times, once for every block affected by changing the blockchain, becomes a nearly infeasible task. This is the basis of the security of blockchain technology.

III. Blockchain Hashing

Clearly, hash functions are a core part of how the blockchain is able to function securely. Thus, part of truly understanding the security behind blockchain means also being aware of the specifics of the hash function.

A. What is a Hash Function

A hash function is a mathematical algorithm that produces output which cannot be traced back to its input [6]. There are several key features:

- It is deterministic; the same output is guaranteed from the same input.
- It is irreversible; cannot reverse output back to its input.
- It is regular; fixed size of output is returned from arbitrary size of input.
- It is obscure; small change of input dramatically affects its output.
- It is unique; computationally cannot find another input which produces the same output.

B. SHA-256

Secure Hash Algorithms (SHA) are a family of cryptographic hash functions published by the National Institute of Standards and Technology (NIST) [7]. There are various versions of SHA, and SHA-256 is one of the most broadly used in blockchain protocols. Along with other multiple coins, Bitcoin uses SHA-256 hashing algorithm.

Hashing Algorithm	Input Bits	Output Bits	Operations	Collision
SHA-0	< 2^64	160	+,and,or,xor,rotl	detected
SHA-1	< 2^64	160	+,and,or,xor,rotl	detected
SHA-224	< 2^64	224	+,and,or,xor,shr,rotr	-
SHA-256	**< 2^64**	**256**	**+,and,or,xor,shr,rotr**	**-**
SHA-384	< 2^128	384	+,and,or,xor,shr,rotr	-
SHA-512	< 2^128	512	+,and,or,xor,shr,rotr	-

TABLE I. Comparison of SHA hash functions [8]

At its most basic level, SHA-256 takes an input which is a sequence of bytes and pads it so that it is a multiple of 512 bits [8]. This input is parsed into blocks and the hash is computed using the different operations shown in the table above [8].

IV. Blockchain Weaknesses and Fraud

Despite the promising nature of blockchain technology, it is not without weaknesses.

A. Scalability

As aforementioned, the process of inserting a block into the blockchain is quite arduous. This has serious implications for the scalability of the blockchain system. In fact, when compared to traditional systems, the number of transactions that can be handled per second is quite low [9]. As the popularity of Bitcoin increases, so do the amount of transactions that occur, and thus need to be added into the blockchain. Due to this increase, the time that it takes for a transaction to be verified has increased as well. In 2017, the time was up to almost 30 minutes per transaction [10]. Space has also been shown to be an issue, with each network node requiring more than 170 GB of storage [9]. Of course, if the popularity of Bitcoin dies down this problem will be minimized. However, the underlying problem remains the same in other applications of blockchain technology. Due to the very process that makes it so attractive from a security standpoint, the scalability of blockchain technology leaves much to be desired. This also raises the question of how this scalability issue could potentially be taken advantage of by malicious individuals or groups to commit fraud.

B. Blockchain Attacks

There have also been several very real incidents of cryptocurrency theft, despite the lofty claims of impenetrable security that often come up in relation to blockchain. In 2018 alone, $30 million of coins were stolen after the Korean cryptocurrency platform, Bithumb, was hacked [11]. Before 2017, an estimated $15 billion worth of coins had been stolen [11]. That being said, cryptocurrency theft is most often a result of traditional methods such as “such as spear phishing, social engineering, distribution of malware, and website defacement” [12]. These methods would have little to do with the security of blockchain itself and more to do with user error.

Regardless, there is a notable method in which blockchain can be exploited known as the double-spending attack, or the related 51% attack. Essentially, the attack occurs in conjunction with an attempt to spend coins twice, once with a normal transaction and once with a fraudulent transaction [4]. Due to the nature of blockchain, if the fraudulent transaction is verified with proof of work and inserted into the blockchain, and this fraudulent chain becomes the longer chain, it will be accepted into the ledger [4]. The 51% attack is so powerful because “if attackers are able to get 51% of the hashing power or more, they will be in a position to drive the longest chain by persuading the

network nodes to follow their chain" [13]. While this type of attack is supposedly rather unlikely, in one month of 2018 alone, at least five companies were subject to the 51% attack [11]. This attack is directly related to hashing power, and who controls that hashing power. The implications of such an attack highlight the importance of hashing as the critical component of blockchain security.

V. Real Data – Block Analysis

A. Basics of the Data

As aforementioned, each block consists of two parts, the header and the body. The former contains a summary of the block and the latter enumerates transactions that belong to the block. The same applies to the given data. Due to the heavy size of the given JSON data, it was effectively impossible and quite inefficient to load the whole data with limited computing resources. Thus, the ijson python library was used to read and analyze the JSON line by line.

Fig 2. Breakdown of the given JSON file

As mentioned above, we can check that a block consists of header and body information. Briefly, 'height' indicates that this is the 556459th block. Using the same logic, the first block of bitcoin must have a 1 instead of 556459. This block contains 946 transactions according to the nTx (number of Transactions). But how do we check that this header information is intact?

B. Block Hashing Algorithm

By hashing the header information of a block, we can get the main hash value of the block. It guarantees the integrity of the block. Among the header information, what we need to verify in terms of the block's integrity are the version, previous block hash, Merkle root, time, bits (also known as difficulty), and nonce.

- Version: The block version in hexadecimal to determine which block validation rule to follow
- Previous block hash: SHA-256 hash value of its previous block
- Merkle root: SHA-256 hash value of its root node of all transactions in the block
- Time: Block timestamp (Unix)
- Difficulty/bits: Target threshold of the block hash
- Nonce: Hash counter

After parsing and applying the double SHA-256 hashing algorithm, one can verify that the block's header information is correct because any small alterations of the block will result in unpredictable hash values, due to the effect explained earlier.

C. Code Analysis

Fig 3. Code: https://github.com/Kyusub-Hwang/BlockChain_Fraud_Detection

A simplified explanation of the steps involved is as follows:

- Step 1: Input block header information
- Step 2: Conversion to little endian hex
- Step 3: Concatenation of header info
- Step 4: Application of double-SHA256
- Step 5: Conversion to big endian hex

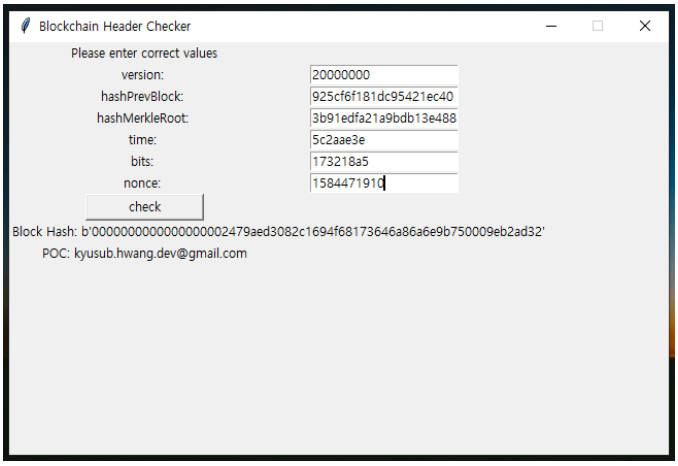

Fig 4. Blockchain Hash Calculator using double SHA-256 (tkinter Python module)

Additionally, to strengthen the security of the block along with checking the block's hash, we can also verify each piece of information in the block header. For example, the Merkle root is a hash value of the block's tree of transactions called a Merkle tree. A Block Transaction Fraud Detector can be

generated by checking each node's hash value from the Merkle root, possibly using Inorder(LRV) tree traversing algorithm. This way, we can get the root node's hash value by calculating each node in the tree. Then, it is guaranteed that each node (transaction) is intact.

VI. Conclusion

Blockchain is a technology based around hash algorithms. It is often seemingly unfeasible to forge the data inside of the blockchain due to the characteristics of hash algorithms mentioned above. Previous attacks were mostly caused due to users' misconfiguration rather than the technology itself. However, blockchain is not infallible and attacks which exploit its features have occurred in the past. Understanding the details of the blockchain and how hashing contributes to its security, or potential lack thereof, is the first step to making improvements to the technology. To further reduce the chance of attack, verification of hash values of each block or transaction would help to guarantee the integrity of blocks and strengthen the security against any fraud.

References

[1] Gupta S., Sadoghi M. (2018) *Blockchain Transaction Processing*. In: Sakr S., Zomaya A. (eds) Encyclopedia of Big Data Technologies. Springer, Cham. https://doi.org/10.1007/978-3-319-63962-8_333-1

[2] Nofer, M., Gomber, P., Hinz, O. et al. *Blockchain*. Bus Inf Syst Eng 59, 183–187 (2017). https://doi.org/10.1007/s12599-017-0467-3

[3] Phan, Linh; Li, Suhong; and Mentzer, Kevin, "Blockchain Technology and The Current Discussion on Fraud" (2019). Computer Information Systems Journal Articles. Paper 28. https://digitalcommons.bryant.edu/cisjou/28

[4] Pierre-Olivier Goffard. Fraud risk assessment within blockchain transactions. 2019. ffhal-01716687v2f

[5] Danny Bradbury, The problem with Bitcoin, Computer Fraud & Security, vol. 2013, no. 11, 2013, pp. 5-8, ISSN 1361-3723, https://doi.org/10.1016/S1361-3723(13)70101-5.

[6] Cryptographic Hash Function. (2021, June 8). In Wikipedia. https://en.wikipedia.org/wiki/Cryptographic_hash_function#cite_note-tIZZx-2

[7] R. Martino and A. Cilardo, "SHA-2 Acceleration Meeting the Needs of Emerging Applications: A Comparative Survey," in IEEE Access, vol. 8, pp. 28415-28436, 2020, doi: 10.1109/ACCESS.2020.2972265.

[8] Y. National Institute of Standards and Technology, *Secure Hash Standard (SHS)*,FIPS 180-4, August 2015, http://dx.doi.org/10.6028/NIST.FIPS.180-4

[9] V. Gatteschi, F. Lamberti, C. Demartini, C. Pranteda, and V. Santamaría, "Blockchain and Smart Contracts for Insurance: Is the Technology Mature Enough?," Future Internet, vol. 10, no. 2, p. 20, Feb. 2018, doi: 10.3390/fi10020020.

[10] A. Chauhan, O. P. Malviya, M. Verma and T. S. Mor, "Blockchain and Scalability," 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2018, pp. 122-128, doi: 10.1109/QRS-C.2018.00034.

[11] Rebecca M. Bratspies, *Cryptocurrency and the Myth of the Trustless Transaction*, 25 Mich. Telecomm. & Tech. L. Rev. 1 (2018).
Available at: https://repository.law.umich.edu/mttlr/vol25/iss1/2

[12] "14 cyber attacks on crypto exchanges resulted in a loss of $882 million." Group-IB, October 17, 2018. [Online]. Available: https://www.group-ib.com/media/gib-crypto-summary/

[13] S. Sayeed and H. Marco-Gisbert, "Assessing Blockchain Consensus and Security Mechanisms against the 51% Attack," Applied Sciences, vol. 9, no. 9, p. 1788, Apr. 2019, doi: 10.3390/app9091788.

[14] M. Crosby, Nachiappan, Pradhan Pattanayak, Sanjeev Verma, and Vignesh Kalyanaraman. Blockchain Technology: Beyond Bitcoin, *Applied Innocation Review*, vol. 2, no. 2, June 2016.