Securing Digital Evidence Information in Bitcoin

A Case Study in Directorate General of Taxes

Dimaz Ankaa Wijaya
Faculty of Information Technology
Monash University
Melbourne, Australia
dawij5@student.monash.edu

Dony Ariadi Suwarsono
Direktorat Jenderal Pajak
Jakarta, Indonesia
dony.suwarsono@pajak.go.id

Abstract—***Bitcoin is the pioneer of cryptocurrency. Bitcoin replaces the need of a central authority with a decentralized system. The distributed ledger employed in the Bitcoin system is used to record every transaction in the network. As the distributed ledger is publicly available, hence it can also serve as a storage system. Moreover, the proof-of-work (PoW) mechanism makes it infeasible for anyone to change the information already stored in the distributed ledger. We propose a method of storing digital evidence information into Bitcoin's distributed ledger to prove the authenticity of the digital evidence obtained during digital forensic activities. The method utilizes Bitcoin's blockchain to act as an unforgeable timestamping system for the hash values of the digital evidence.***

Keywords—Bitcoin; blockchain; digital evidence; digital forensic; hash

I. INTRODUCTION

A. Bitcoin

Bitcoin is the first cryptocurrency ever created in the world. The system was first developed by someone under pseudonym of Satoshi Nakamoto in 2008 [1]. Unlike the fiat currency such as US Dollar or Indonesian Rupiah in which they are managed by central banks, similar system does not exist in Bitcoin. There is no central authority that controls how the system works. Instead, the Bitcoin system replaces the centralized system with a decentralized system. The system runs in a peer-to-peer network; each Bitcoin node acts as the server and the client of the system [1]. Bitcoin employs cryptographic functions to verify and validate transactions created in the network [2].

Each Bitcoin node maintains an exact copy of a distributed ledger called blockchain. The blockchain is a database used to store the transactions. Several transactions that are sent into the network in a period of time is grouped into a block. The newly created block is then chained to the preceded block as such that all of the created blocks create a chain-like logical structure.

The blockchain is also protected by a mechanism called proof-of-work (PoW). The PoW was adapted from a concept called Hashcash created by Adam Back [3]. Hashcash is a technique to reduce spamming by forcing the email sender to do a complex computation before sending the email, thus exhausts the resource of the spammer which in turn reduces the number of spam emails sent by the sender. PoW is used to protect any attacker in modifying any information stored in the blockchain. In order to create (or modify) a block, a user is forced to do a certain level of computation. Only if the computation satisfies a predetermined level, then the user is allowed to create or modify a block. The chained blocks make it even more infeasible for the user to modify a block already confirmed after a certain depth, because the user also needs to modify the superseded blocks as well [1].

B. Tax Fraud Preliminary Investigation

A tax fraud preliminary investigation is a mechanism of collecting evidence to build a tax fraud case. The duty of investigating tax fraud cases is carried by Directorate of Law Enforcement which was previously known as Directorate of Intelligence and Investigation. The directorate assembled a new division under Subdirectorate of Forensic and Evidence. As implied by its name, the sub directorate provides a technical support towards tax fraud investigations throughout its jurisdiction. One of the most important task of this sub directorate is providing digital forensic services in acquiring evidences in digital form.

The tasks and authorities of the Directorate of Law Enforcement in the preliminary investigation cases are regulated under The Ministry of Finance Regulation Number 239/PMK.3/2014 [4] and the technical details are explained in The Letter of Director General of Taxes Number SE-23/PJ/2015 [5]. The regulations describe the procedures of the digital forensic procedures in which the tax authorities are legally authorized to "borrow" any data owned by tax payers suspected to be related to tax fraud cases. The data can be in the form of physical data as well as digital data.

In the process of "borrowing" the data owned by the tax payers, the tax investigators are obliged to provide official letters as proofs of borrowing. If any of the borrowed data is in digital form, then the official letter must include appendices that contain the detail of the digital data acquisition processes [5]. The digital data itself will be represented as hash values. The appendices and the official letters are then signed by both parties: the tax investigators and the taxpayers. Both of the parties keep copies of the official letters. If the case proceeds to the court, the acquired data could be presented as an evidence to support the case under Information and Electronic Transaction Act [6]. These letters could then be used as a proof that the data was legally acquired from the tax payers and no data has been modified since its acquisition.

C. Problem

The mechanism of providing an official letter as a proof of digital forensic acquisition relies heavily on a trusted system. The official letter could be intentionally or unintentionally tampered by either the tax investigators or the taxpayers. The current regulations do not have a mechanism of resolving problems should there be any discrepancies between these 2 copies. The letter-tampering case could then be a problem that occupies the government's time and resources which could actually be allocated to focus on collecting tax and processing real tax fraud cases.

D. Our Contribution

We propose a trustless method of storing the hash values of digital evidence acquired during digital forensic activities. The method employs Bitcoin transaction as the method of storing the hash value. The Bitcoin system acts a timestamp which proves the existence of an information at the mentioned time. The information stored in the Bitcoin system is then protected from being tampered. This scheme mitigates the risk of the information within the official letters being modified. In this case, the information stored in the Bitcoin system can be referenced to determine which information should be considered as the original copy.

II. PRELIMINARIES

A. Hash Value

The hash value is a value generated by a hash function which takes the data as the input. The hash value represents an arbitrary length of data into a fixed length of value [7]. The hash value could be used to detect any change to the data; any small change in the data produces a significant change in the hash value. A good hash function must satisfy several requirements: preimage resistance, second-preimage resistance, and collision resistance [8]. Several hash functions usually used in digital forensic operations are MD5, SHA-1, and SHA-2.

B. Bitcoin Transaction

A Bitcoin transaction is constructed by series of mathematical puzzle which will be evaluated by the system. The mathematical puzzle is built by using Bitcoin Operation Codes (OpCodes) to determine what the system should do with the data. A bitcoin transaction consists of 2 parts: ScriptSig and ScriptPubKey [2]. The ScriptSig is a requirement that needs to be fulfilled by the redeem transaction, while ScriptPubKey is the answer to the requirement of the referenced transaction.

Several types of Bitcoin transactions are: Pay To Address (P2A), Pay To Script Hash (P2SH), and Null Data [9]. P2A is a transaction that pays to a Bitcoin Address, P2SH is a transaction that pays to a predetermined script, while Null Data does not actually pay but rather a mechanism to embed a metadata to the transaction.

C. Multisignature

Multisignature is a type of Bitcoin transaction which requires multiple signatures [10]. The scheme is denoted as m-of-n multisignature with $m \leq n$, m is the minimum number of signatures required for the transaction to be redeemed, and n is the total possible signatures which can sign the transaction. Multisignature is commonly applied by using P2SH scheme.

III. RELATED WORKS

Ever since the Bitcoin is introduced, the applications utilizing the technology grow in number and type. The Bitcoin is known not only used as a digital payment system but also a system to permanently record information into the blockchain which is available as public information. Several electronic notarization and asset management services was introduced [11], [12], [13], [14], [15]. These services put the hash value of the digital data into Bitcoin transactions and send the transactions to the network. The idea of the services is to prove the existence of a digital data (digital artworks, digital music, poems, etc) in a point of time, proven by their hash values already exist in the date and timestamped by the Bitcoin system. In case of anyone tries to claim the ownership of the digital data, then the original owner can prove that the digital data represented by its hash value has already exist in the past. The owner can also verify that he/she already owned the digital data in a certain time in the past, breaking the claim of the ownership of the digital data by others.

The aforementioned electronic notarization services have a limitation on the length of the data which can be embedded into Bitcoin transactions. This is caused by the limitation of the Null Data transaction allowing the maximum of 80 bytes information embedded into a single Bitcoin transaction [16]. The Null Data transaction uses a specific Bitcoin operation called OP_RETURN.

A more sophisticated protocol was proposed to insert data into Bitcoin transaction [17]. The protocol inserts long data into the public keys which can be used to avoid censorships. The protocol is similar to a technique called steganography. The protocol allows a user to insert an arbitrary length of data into Bitcoin transactions with the average transaction fee of 16 satoshis per byte.

IV. SECURITY MODEL

Let there exist 2 parties in the system: the tax investigator as the prover and the taxpayer as the verifier. The prover and the verifier agrees on a record of a hash value over a digital data acquired at a certain time. The prover tries to prove that there is no change made to the data acquired from the verifier by comparing the past record containing the hash value and the hash value computed at present time. The verifier verifies both data.

The attack against the system is defined as an effort made by the prover or the verifier by tampering the record of the hash value to validate or invalidate the computed hash value. The system is considered as secure against tampering if the effort of modifying the record is infeasible.

V. THE PROPOSED METHOD

A. Overview

The proposed method utilizes P2SH scheme to create a 2-of-2 multisignature transaction that pays to a predetermined destination address which pools all the similar data. The

transaction sends a small amount of bitcoin to the destination address with the hash value of the evidence embedded in the transaction. Each hash value is recorded in a separate transaction. Both parties sign the transaction and whenever needed, they refer to the Bitcoin transactions that contain the hash values.

B. Bitcoin Address Generation

Each of the parties creates a Bitcoin public key pair. As a government official, the prover must derive the key pair from a master key owned by the government by using a hierarchical deterministic wallet scheme [18]. The public key part of these public key pairs are used to identify the parties. Let the address of the prover be called `PRVR_ADDR` and the address owned by the verifier be called `VRFR_ADDR`. The private keys of these addresses must be kept secret and only known by their owners.

It is also assumed that the government already provides a Bitcoin address to collect the records of all hash values of the evidence. The address must be published as being used as a special purpose. Let this address be called `GOVT_ADDR`.

C. Verifying The Participants

A digital certificate scheme is utilized to verify the participants of the scenario. Such scheme has been developed by MIT Media Lab by using Bitcoin to store the hash value of the digital certificate issued by an authority [19]. The digital certificates verify the identity of the parties as well as verify the Bitcoin addresses involved in the protocol. The digital certificate issuance is done prior to data embedding process to the blockchain.

D. Data Insertion

The 2-of-2 multisignature transaction is created by an address by the prover and another address by the verifier. It means that the transaction can only be commenced only if both of the parties sign the transaction. The 2-of-2 multisignature pays `BTC` to the `GOVT_ADDR` in a Bitcoin transaction `TX` with the embedded data `Hash` in the scheme below.

```
(PRVR_ADDR, VRFR_ADDR) -> TX(GOVT_ADDR, BTC|Hash)
```

The Bitcoin transaction `TX` produces a unique transaction identifier TXID which could be used to refer the hash value. The Bitcoin system timestamps the `TX` once it is confirmed into the blockchain and propagated in the network.

VI. Evaluation

A. Limitation and Assumption

As it may also happen to other security model, the proposed method could not stand against both parties cooperate together to tamper the system by creating a new transaction with a new hash value embedded in the transaction. It is also assumed that the digital certificate authority always behaves honestly.

B. Security Evaluation

Both of the prover and the verifier provide their digital signature to the transactions. Assuming that the private keys are known only by their owners, then the non-repudiation characteristics and the security of the private keys of the scheme rely on the unforgeability of Elliptic Curve Cryptography.

If the prover tries to attack the system by forging the evidence and create a new record in the Bitcoin system, then the verifier can verify that the new record does not conform the receipt given to the prover at the time when the evidence is acquired. The verifier could then deny the new record. These 2 records can be distinguished by looking at the timestamp of the transactions. Assuming that the fake data is created after the original data, then the transaction containing the fake data tends to have a later date and time compared to the original transaction. Similar thing applies in the case where the verifier is the attacker.

If any of the party tries to attack the system by creating dummy addresses to create the transaction, they also need to forge a valid digital certificate. Assuming that the digital certificate authority always behaves honestly, then the possibility of attacking the system by using such scenario is negligible.

C. Performance and Transaction Fee

Based on P2SH specification, there are 2 transactions that need to be performed: commit transaction and redeem transaction. An experiement shows that these transactions can be confirmed in the same block, and therefore they require roughly 10 minutes to be confirmed and considered as permanent transactions in the system. The time required to confirm these transactions depend on several factors, including the time required by Bitcoin miners to produce a valid block and queued transactions waiting to be confirmed. These 2 factors could vary over time and not subjects to control by any Bitcoin users.

Each transaction requires at least 10,000 satoshis to be paid to Bitcoin miners, and therefore at least 20,000 satoshis are consumed for both transactions. If the price of 1 bitcoin is worth Rp. 7,610,500.00 [1] then the total transaction fee for both transactions will be at the rate of Rp. 1,522.00. The amount of bitcoin sent in these transactions should not less than 10,000 satoshis, but because this bitcoin is not actually spent, then it is not considered as an expense.

VII. Conclusion

We propose a method of storing hash values to act as a trustless failsafe which prevents the tax auditor or the taxpayer modifying the data acquisition records as an effort to deny the evidence or delay the tax fraud case. The method relies on the security of Bitcoin in which any of the confirmed transactions in the blockchain is infeasible to tamper after certain depth.

[1] Bitcoin price taken from https://vip.bitcoin.co.id as per 31 May 2016

Future work could be used to implement the scheme and use it in real case scenarios. Then, analysis could be done in the system to improve the efficiency of the method.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[2] P. Franco, *Understanding Bitcoin: Cryptography, engineering, and economics.*: John Wiley & Sons Ltd., 2015.

[3] A. Back, "Hashcash-a denial of service counter-measure," ed, 2002.

[4] *Tata Cara Pemeriksaan Bukti Permulaan Tindak Pidana di Bidang Perpajakan,* Menteri Keuangan Republik Indonesia 239/PMK.03/2014, 2014.

[5] *Petunjuk Teknis Pemeriksaan Bukti permulaan Tindak Pidana di Bidang Perpajakan,* Direktorat Jenderal Pajak SE - 23/PJ/2015, 2015.

[6] *Undang-Undang Tentang Transaksi dan Informasi Elektronik*, 2008.

[7] D. R. Stinson, *Cryptography Theory and Practice*: CRC Press, 1995.

[8] P. Rogaway and T. Shrimpton, "Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance," in *Fast Software Encryption*, 2004, pp. 371-388.

[9] D. A. Harding. (2015, 12 January 2016). *Bitcoin Developer Guide*. Available: https://bitcoin.org/en/developer-guide

[10] G. Andresen. (2011, September 28, 2015). *M-of-N Standard Transactions*. Available: https://github.com/bitcoin/bips/blob/master/bip-0011.mediawiki

[11] M. Araoz and E. Ordano. (2012, 28 January). *What is proof of existence?* Available: https://www.proofofexistence.com/about

[12] F. Charlon. (2013, 28 January). *Open Assets Protocol (OAP/1.0)*. Available: https://github.com/OpenAssets/open-assets-protocol/blob/master/specification.mediawiki

[13] T. M. D. Holtzman. (2015, 28 January). *Towards an Ownership Layer for the Internet, V1.03, 20150624*. Available: https://d1qjsxua1o9x03.cloudfront.net/live/trent@ascribe.io/ascribe%20whitepaper%2020150624/digitalwork/ascribe%20whitepaper%2020150624.pdf

[14] Omnilayer. (2016, 28 January). *Omni Layer*. Available: http://www.omnilayer.org/

[15] P. Snow, B. Deery, J. Lu, D. Johnston, and P. Kirby. (2014, 28 January). *Factom Business Processes Secured by Immutable Audit Trails on the Blockchain* Available: https://github.com/FactomProject/FactomDocs/blob/master/Factom_Whitepaper.pdf?raw=true

[16] W. J. v. d. Laan. (2015, 8 February). *Change the default maximum OP_RETURN size to 80 bytes*. Available: https://github.com/bitcoin/bitcoin/pull/5286

[17] K. Okupski, "(Ab)using Bitcoin for an Anti-Censorship Tool," Department of Mathematics and Computer Science Eindhoven University of Technology2014.

[18] P. Wuille. (2012, 29 February). *Hierarchical Deterministic Wallets*. Available: https://github.com/bitcoin/bips/blob/master/bip-0032.mediawiki

[19] J. Nazaré. (2016, 14 June). *What we learned from designing an academic certificates system on the blockchain*. Available: https://medium.com/mit-media-lab/what-we-learned-from-designing-an-academic-certificates-system-on-the-blockchain-34ba5874f196#.665ixy9f8