**Creditworthiness – A Pillar of Trust**

**in Cryptocurrency Transactions**

Sudip Bhattacharyya, Dan Freeman, Timothy McWilliams,

Craig Hall, Pablo Peillard

Southern Methodist University

6425 Boaz Lane

Dallas, TX 75205

{sbhattacharyya, freemand, tmcwilliams}@smu.edu,  
craig@caudicum.com, pablo@caudicum.com

**Abstract.** In this paper, we present a novel authentication and reputation scoring algorithm for transactors in the cryptocurrency marketplace using historical block- and transaction-level Bitcoin data. While the decentralized aspect of blockchain technology on which cryptocurrencies operate has many benefits, the lack of regulation in this space has provided illegitimate users an alternative to cash for harboring their criminal activities. In order to distinguish law abiding cryptocurrency users from criminal users, we characterize and identify historical transaction patterns associated with the ownership of particular coins. PROBABLY NEED ANOTHER SENTENCE ON WHAT TECHNIQUES ARE USED. We find that [our main numerical result]. [our main conclusion].

1 Introduction

Money has its strong presence in the complex structure of today’s fiat-based financial world. The global financial market today relies on a centralized system in which an authority, usually a bank, is responsible for validating transactions between parties. This model inherently requires involvement of a neutral third party in which the parties taking part in a transaction place trust. Moreover, the role of the third party is important in case of any disputes. HOW SO? WHAT DOES TRUSTED THIRD PARTY DO IN CASE OF DISPUTES? THIS SHOULD RELATE TO WHY NOT NEEDED WITH BITCOIN AND ITS USE OF THE BLOCKCHAIN LEDGER.

With the advent of the first cryptocurrency, Bitcoin, in 2009, the traditional centralized system of financial attestation has been challenged. Cryptocurrencies operate within a decentralized distributed ledger platform, known as blockchain, in which transactions are validated by other nodes in the network via public key cryptography. This decentralized architecture significantly reduces the probability of unauthorized access which can be obtained much more easily with a centralized repository. Despite the tremendous potential of establishing a secure environment for global finance, blockchain is often argued to widen the channels of money laundering because of lack of rules and regulations associated with cryptocurrency activity**.** In recent years, criminal organizations have been known to take advantage of the unregulated nature of cryptocurrencies, usually as a means of money laundering, in order to facilitate their illicit activities NEED CITATIONS TO SUPPORT THIS CLAIM. WHAT’S YOUR PROBLEM? YOU HAVEN’T DEFINED A PROBLEM TO BE SOLVED WITH YOUR WORK. BE SPECIFIC AND APPROPRIATE.

As reputation systems for transactors in cryptocurrencies would assist in identifying potential criminal actors, we present an algorithm to authenticate the identity and reputation of transactions and transactors. This makes the cryptocurrency marketplace more secure and trustworthy HOW?.

NEED DESCRIPTION OF WHAT YOU DID, PARAGRAPH ON MAIN RESULTS AND PARAGRAPH ON MAIN CONCLUSIONS.

The remainder of this paper is structured as described here. Section 2 describes how continuous research on digital currency laid the foundation for the rise of cryptocurrencies. This section also presents a few facts and figures on the latest available data on the cryptocurrency market. The basic architecture of blockchain as a decentralized and distributed ledger technology along with the security protocols used in blockchain are explained in Section 3. Section 4 describes the fields of the dataset we use as well as the cryptographic details behind how some of the fields are derived. Section 5 contains a biography of our advisor, Craig Hall.

2 Evolution of Cryptocurrency

The barter system is the oldest form of economic exchange, involving a direct trade of goods and services between parties. Even though this system has existed since ancient times, it has always faced the challenge to match the needs of different parties involved in a transaction. Often an arrangement of additional parties is inevitable to align the requirements at the same place at the same time which slowly motivated the innovation of currencies through successive stages of evolution as a neutral medium of exchange. A cash-based system addresses this problem, but no trade can take place without an initial allocation of cash, and the credit-based system is not without issue either as there is a risk of the borrower defaulting on the repayment of their debt.

Online payments started gaining popularity in the early 1990s, but consumers were heavily hesitant to disclose information required for the transaction to materialize. To address this concern, an architecture named SET has been developed by Mastercard and VISA in collaboration with tech-giants such as IBM, Microsoft, Verisign and RSA [1]. This architecture used an encryption for the purchase details along with the consumer’s credit card information that no one except the intermediary could decrypt. The intermediary had the authority to approve transactions only if the consumer’s purchase details matched with merchant’s view. Unfortunately, this system failed over time due to lack of user friendliness as the system required the buyers to get a digital certificate which was a tedious task. NO. THIS DID NOT PAVE THE WAY FOR CRYPTOCURRENCY.

The first idea of a digital currency was conceptualized in 1983 by an American cryptographer, David Chaum, who developed ecash and later founded an electronic finance company named Digicash [1]. Chaum introduced a ‘blind signature’ that concealed the content using the encryption algorithm RSA. This content then could be unblinded and authenticated by the recipient via a digital signature. This electronic cash still needed financial institutions operating as a centralized authority to ensure a trustworthy system for its users. After the downfall of Digicash, the world economy has observed numerous initiatives of an electronic cash, including e-Gold, Bitcash, Bitgold, B-money, PayPal to name a few [1].

The emergence of the world’s first cryptocurrency known as Bitcoin was presented by Satoshi Nakamoto (a pseudonym of a person or a group of people) in YEAR [CITE PAPER]. Nakamoto also developed the first ever blockchain database where the genesis block (the first block) has a timestamp of 18:15:05 GMT on 3 January 2009 [2]. The concept of blockchain, a decentralized, peer-to-peer trust protocol presents a challenge to an established structure of a centralized authority for banking and finance. Transactions here take place between willing parties based on a cryptographic proof, not on trust in a third party. Nakamoto’s white paper suggests that this concept of blockchain provides a robust solution to the possible issue of “double spending” with a peer-to-peer network. The timestamped transactions are appended to an existing chain of transaction blocks with a hash-based proof of work where it is unalterable, as it requires a redoing of this proof of work which is an impossible task [3]. Moreover, since appended transactions are irreversible, blockchain provides a protection from fraud. YOU HAVE AN ENTIRE SECTION ON THE DETAILS OF BLOCKCHAIN. MENTION THAT.

THIS SECTION DOESN’T QUITE LIVE UP TO ITS TITLE. WHERE’S THE EVOLUTION? TWO POINTS 16 YEARS APART IS NOT AN EVOLUTION. ALSO, THEY ARE BOTH COMPLETELY DIFFERENT TECHNOLOGIES, SYSTEMS, AND DESIGNS.

SHOULD HAVE A TABLE WITH THE CURRENT CRYPTOCURRENCIES, YEAR INTRODUCED, BASIS, ETC.

MAKE THIS A SEPARATE SECTION – AND CREATE YOUR OWN FIGURES AND TABLES. YOU CANNOT INCLUDE COPYRIGHTED MATERIAL WITHIN YOUR PAPER

Since the innovation of Bitcoin, the world economy has observed the introduction of numerous cryptocurrencies to market such as Litecoin, Ethereum, Ripple, Dash and Monero. Even though most of the cryptocurrencies operating in the market today are predominantly cloned from Bitcoin, there are several others built on a different functionality such as a new consensus mechanism and smart contracts. To date, there are more than 1.000 types of cryptocurrencies operating in the market with a total market capitalization of more than $800 billion (Fig.2.1) [4]. The current active users of cryptocurrencies across the globe are estimated to be between 2.9 and 5.8 million years old?. The industry has observed a steady growth (exponential in a few cases) for prices of most cryptocurrencies over recent years (Fig.2.2) [5].

**Fig. 2.1.** Total market capitalization for cryptocurrencies

**Fig. 2.2.** Pricing for major cryptocurrencies

By the end of 2017, total Bitcoin supply crossed 16 million coins with an average 24-hour trade volume of $4.9 billion. This has enabled Bitcoin to capture more than 70% of the current market WHAT MARKET? across 96 countries where the usages are unrestricted. Since its inception in 2013, Ethereum, featuring smart contracts, has emerged as the biggest competitor IN WHAT SENSE? of Bitcoin (Fig.2.3). Some numbers regarding Bitcoin are displayed in following figure (Fig. 2.4) [5].

**Fig. 2.3.** Percentage share of market capitalization for major cryptocurrencies

**Fig. 2.4.** Facts & figures on Bitcoin

Customers can trade cryptocurrencies in exchange for assets like traditional money through exchanges spread across the globe. Over half (52%) of small exchanges WHAT EXCHANGES? WHAT IS AN EXCHANGE? WHAT IS IT EXCHANGING? SENSE YOU ARE TALKING ABOUT EXCHANGES FOR A COUPLE OF PARAGRAPHS, YOU NEED TO DEFINE THEM FIRST. and 35% of large exchanges have a formal government license for operation (Fig. 2.5). Exchanges in Asia-Pacific are mostly unlicensed (85%) whereas licensed exchanges (78%) are dominant in the North America region [5].

**Fig. 2.5.** Distribution of licensed & unlicensed exchanges

The operational risk factors for exchanges are presented in Table XX [5]. Small exchanges are often more heavily impacted by fraudulent activity than large exchanges.

**Table XX**

The global scenario of the cryptocurrency industry is summarized in the following map (Fig. 2.7). YOU NEED TO TALK ABOUT THIS NOT JUST POINT TO A FIGURE. Based on a decentralized architecture, the cryptocurrencies have tremendous promise to revolutionize the global economy beyond any border and barrier [5]. REVOLUTIONIZE HOW?

**Fig. 2.7.** The crypto-world

3 Blockchain Technology

“The blockchain is an incorruptible digital ledger of economic transactions that can be programmed to record not just financial transactions but virtually everything of value [6].” – Don & Alex Tapscott

START WITH A DEFINITION OF BLOCKCHAIN NOT A TONGUE IN CHEEK STATEMENT Blockchain is a continuously growing chain of blocks. New transactions are grouped to form a block and then the block is appended to the existing chain. The blockchain is a digital ledger distributed across multiple computers or servers, known as ‘nodes’, connected over a peer- to-peer network (Fig. 3.1). Any transaction over this network is made secure via cryptography. In addition, blockchain ensures authenticity of any new transaction with validation by the miners in the network. Once validated and added, transactions cannot be erased, only another transaction can be added to reverse an erroneous transaction which makes the length of the chain ever-growing.

**Fig. 3.1.** Basic structures of centralized vs distributed system [6]

Before getting into a detailed operating model, we define a few important terms, as follows:

* **Parent Block**: Any block preceding the current block is called a parent block. The parent block’s structure is hashed and included in the current one as an input.
* **Genesis Block**: The first ever block created in a chain is known as the genesis block. For Bitcoin, Satoshi Nakamoto created the genesis block in 2009 just after the cryptocurrency was conceptualized.
* **Blockchain Head**: The latest block added to a chain is called the blockchain head and the future blocks are appended to the current head.
* **Block Height**: The height of a block refers to its chronological order.
* **Miner**: Mining is an act of searching a difficult ‘proof-of-work’ to append a new block to an existing chain where nodes perform as miners in the context. Multiple nodes engage themselves into a contest to solve partial hash inversion of the existing chain and add new blocks to the chain.
* **Fork and Orphan Block**: There may arise a situation when different miners arrive at different solutions and be ready to append the new blocks at the same time, but the longest chain wins in this situation. This scenario is known as ‘fork’ and the discarded blocks are called ‘orphan’ blocks (Fig. 3.2).

**Fig. 3.2.** Forks in a blockchain [7]

A proof-of-work mechanism is followed in Blockchain operations. Each block in a blockchain contains a set of validated transactions. These transactions, clubbed with a nonce and hash value of the previous block using SHA-256, form the current block. The nonce solves the hash inversion of the parent block and is the key element that helps a node/miner to win the contest of appending the new block. The nonce determines the number of zeros in the hash and thus needs to be adjusted to produce the required ‘0’ bits (Fig. 3.3).

**Fig. 3.3.** Basic Blockchain architecture [7]

Sequentially chained transactions are secured with a digital signature based on public-private key encryption (Fig. 3.4) that enables a payee to verify the chain of ownership. The blocks are also time-stamped to avoid the issue of duplicate transactions.

**Fig. 3.4.** Encrypted transactions in a block [7]

To grow, the network nodes identify the longest chain on which to work. Once a new transaction is broadcast to the chain, all available nodes collect the transaction to their own block and initiate a difficult proof-of-work for the block. The quickest node with a valid proof-of-work is then selected to grow the chain and the others will be left as orphans for future use. A Merkle tree structure with a pruning mechanism (Fig. 3.5) is considered in the blockchain network to save disk space with an ever-growing network.

**Fig. 3.5.** Merkle tree and pruning [7]

The key advantages of a blockchain network are as follows [8]:

* **Decentralization**: In contrast to a centralized system, a blockchain database is distributed across nodes connected in a distributed network. The network operates peer-to-peer, with the nodes together managing the database.
* **Durability and robustness:** Since it is built on the Internet, blockchain automatically inherits the durability of the Internet. Moreover, since it cannot be controlled by any single entity or node and there is no single point of failure, blockchain is expected to produce a more robust result.
* **Transparency and incorruptibility:** Blockchain works in a consensus. A self-auditing system reconciles transactions in regular intervals. Any block of transactions is visible to all the participants and data cannot be altered, once validated and entered.
* **Enhanced security:** First, with a distributed database architecture, a threat of attack on any centralized point is eliminated. Moreover, the proof-of-work mechanism and the use of hash functions and public-private keys make this blockchain network very secure from attack.

4 Data

We obtained complete historical Bitcoin blockchain data from Kaggle [9]. THE DATA SET…GIVE DESCRIPTION OF IT HERE. NEED TO INCLUDE HOW MANY ATTRIBUTES, ROWS, ETC

**Table 1.** Block level. NEED BETTER CAPTION – BELONGS IN BLOCK LEVEL SUBSECTION – GIVE EXAMPLES OF VALUES

|  |  |
| --- | --- |
| Field | Description |
| Timestamp | Current time as seconds since 1970-01-01T00:00 UTC. |
| Transactions | A transfer of a coin value, Bitcoin, that is broadcasted to the network and collected into blocks. |
| Block ID | 256-bit hash that represents a given block. |
| Previous Block | 256-bit hash of the previous block header. |
| Merkle Root | 256-bit hash based on all of the transactions in the block. |
| Nonce | 32-bit number used once. Used to generate the block and to allow variations of the header to compute different hashes. |
| Version | Block version number. |
| Work Tera Hash | The hash rate, a measure of mining performance. |
| Work Error |  |
| Difficulty Target | A measure of how challenging it may be to find a hash below a given target. The target is a 256-bit number that every client shares. |
| Row Number |  |

**Table 2.** Transaction level. NEED BETTER CAPTION – BELONGS IN TRANSACTION SUBSECTION – GIVE EXAMPLES OF VALUES

|  |  |
| --- | --- |
| Field | Description |
| Transaction ID | A hashed value that represents a given transaction. |
| Inputs | A reference to an output from a previous transaction. |
| Outputs | Contains the instructions for sending coins, Bitcoins. |

**4.1 Block level**

The block is where the transaction data is permanently recorded and stored. Each block is organized in a linear order over time. When a new block is created, it is added to the end of the chain. Once added to the network, it can never be changed or removed. Inside the block level, there is a timestamp, block ID, previous block, Merkle root, the transactions, a nonce, version, Tera hash, work error, difficulty target and a row number.

The block ID is a SHA256 hash that is calculated from the version, previous block, Merkle root, timestamp, the difficulty target and the nonce [10]. The calculation of the 256-bit hash value also requires the standard SHA256 padding, making two 64-byte portions. The SHA256 algorithm needs these two portions in order to calculate the block ID hash.

The block timestamp is attached to each and every block. It serves as a source of variation for the block hash as well as making it more challenging for the block chain to be manipulated. When the timestamp is greater than the median timestamp of the previous eleven blocks and less than the network-adjusted time plus two hours, only then a timestamp is valid [11].

Each block contains a Merkle root which is a part of the Merkle tree. The Merkle tree is a binary tree of hashes. To obtain the Merkle root of the Merkle tree, a two-step procedure is required. First, the bottom row of the tree is formed with the double-SHA256 hashes of the transactions in the block. Next, the row above the bottom row consists of half that number of hashes. “Each entry is the double-SHA256 of the 64-byte concatenation of the corresponding two hashes below it in the tree” [10]. If a row that is not the root of the tree contains an odd number of elements, then the double hash is duplicated. This ensures that the row will have an even number of hashes. Now, this procedure will repeat recursively until there is a row that contains just a single double hash; this row is the Merkle root [10]. For example, we have a block with five transactions a, b, c, d and e. The Merkle tree is then: DON’T INLINE CODE

# Double hash each transaction

# Since there is an odd number, the final double hash is duplicated

dh1 = SHA256(SHS256(a))

dh2 = SHA256(SHS256(b))

dh3 = SHA256(SHS256(c))

dh4 = SHA256(SHS256(d))

dh5 = SHA256(SHS256(e))

dh6 = SHA256(SHS256(e))

# Concatenate

dh7 = SHA256(dh1 + dh2)

dh8 = SHA256(dh3 + dh4)

dh9 = SHA256(dh5 + dh6)

# Get the Merkle root

MR = SHA256(dh7 + dh8 + dh9)

The difficulty is a measure of how challenging it may be to find a hash below a given target. A target is a 256-bit number that every client shares. In order for a block to be accepted to the network, the SHA256 hash of the block must be lower than or equal to the current target [12]. Each individual block stores a representation for the hexadecimal target which can be derived. For example, say we have a packed target on the block that is 0x1b0404cb [12]. Then we can derive this into a hexadecimal current target by:

|  |  |
| --- | --- |
| current\_target = 0x0404cb \* 2(8 (0x1b - 3)) . | (**1**) |

The highest possible difficulty target is defined as 0x1d00fff [12]. We can derive this into a hexadecimal target by:

|  |  |
| --- | --- |
| difficulty\_1\_target = 0x00fff \* 2(8 (0x1d - 3)) .  Where *difficulty\_1\_target* is the highest possible target, difficulty 1. | (**2**) |

Then, we can calculate the difficulty by:

|  |  |
| --- | --- |
| D = difficulty\_1\_target / current\_target . | (**3**) |

4.2 Transaction level

Inside each block, there are multiple transactions. “A transaction is a transfer of Bitcoin value that is broadcast to the network and collected into blocks” [13]. The transaction will reference previous transaction outputs as the new transaction inputs. Then dedicates all input coin values to new outputs. Since transactions are not encrypted, every single transaction can be seen inside a block.

The input is a reference to the output from the previous transaction. There can be many inputs per transaction. All of the inputs for a single transaction are added up and then used in the outputs of the new transaction [13]. The output contains the instructions for sending the cryptocurrency over the network. Just like the inputs, there can be multiple outputs in a transaction that will share the added values of the inputs [13].

**Fig. 4.** A sends 100 BTC to C and C generates 50 BTC. C sends 101 BTC to D. D sends the 101 BTC to someone else, but they haven't redeemed it yet. Only D's output and C's change are capable of being spent in the current state [13].

In order to verify that the inputs are authorized to collect the values that are referenced in the outputs, a Forth-like scripting system is used. There are two parts to the scripting system for all transactions made, scriptSig and scriptPubKey [14]. The inputs are a part of the scriptSig and the referenced outputs are a part of the scriptPubKey.

The script is composed of a list of instructions that are recorded with each transaction. These instructions describe how well the next client can gain access to the cryptocurrency being transferred. The client that is spending the cryptocurrency must provide two things: a public key that has the destination address rooted in the script and a signature to prove the ownership of the private key that corresponds to the public key [15].

WHAT DOES THE REST OF YOUR PAPER STRUCTURE LOOK LIKE?

References

1. A. Narayanan, J. Bonneau, E. Felten, A. Miller, S. Goldfeder “Bitcoin and Cryptocurrency Technologies: A comprehensive Introduction,” Princeton University Press. [Online]. Available: https://press.princeton.edu/titles/10908.html
2. Wikipedia on Satoshi Nakamoto. [Online]. Available: https://en.wikipedia.org/wiki/Satoshi\_Nakamoto
3. Satoshi Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”. [Online]. Available: https://bitcoin.org/bitcoin.pdf
4. A. Marshall “Combined Crypto Market Capitalization Races Past $800 Bln”Jan 07, 2018. [Online]. Available: https://cointelegraph.com/news/combined-crypto-market-capitalization-races-past-800-bln
5. Dr. G. Hileman, M. Rauchs University of Cambridge – Judge Business School “Global Cryptocurrency Benchmarking Study”. [Online]. Available: https://www.jbs.cam.ac.uk/fileadmin/user\_upload/research/centres/alternative-finance/downloads/2017-global-cryptocurrency-benchmarking-study.pdf
6. Blockgeeks, “What is Blockchain technology? A step by step guide for beginners,” [Online]. Available: https://blockgeeks.com/guides/what-is-blockchain-technology/
7. R. Franco, Chapter 7, “Understanding Bitcoin Cryptography Engineering and Economics*”*, 2014.
8. S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” [Online]. Available: http://static.bitcoingroup.com.au/wp-content/uploads/2015/09/Satoshi-White-Paper.pdf.
9. Bitcoin Blockchain; Complete Live Historical Blockchain Data. Available: https://www.kaggle.com/bigquery/bitcoin-blockchain
10. bitcoinwiki, "Protocol Documentation," [Online]. Available: https://en.bitcoin.it/wiki/Protocol\_documentation#block.
11. bitcoinwiki, "Block Timestamp," [Online]. Available: https://en.bitcoin.it/wiki/Block\_timestamp.
12. bitcoinwiki, "Target," [Online]. Available: https://en.bitcoin.it/wiki/Target.
13. bitcoinwiki, "Difficulty," [Online]. Available: https://en.bitcoin.it/wiki/Difficulty.
14. bitcoinwiki, "Transaction," [Online]. Available: https://en.bitcoin.it/wiki/Transaction.
15. bitcoinwiki, "Transaction," [Online]. Available: https://en.bitcoin.it/wiki/Transaction.