**Reputation Scoring for Blockchain Transactions**

Sudip Bhattacharyya1, Dan Freeman1, Timothy McWilliams1,

Craig Hall1, Pablo Peillard1

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 transaction-level Bitcoin data. While the decentralized aspect of blockchain technology on which cryptocurrencies operate has many benefits, the anonymity of users in this space has provided criminal users an alternative to cash for harboring their illicit 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. Historical incidents are analyzed to profile transactions to develop a reputation score. We find that [our main numerical result]. In conclusion, [our main conclusion].

1 Introduction

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. Since ledgers are updated centrally by financial institutions beyond customers’ visibility, disputes can occur during transactions. Registered third parties come into play to resolve a dispute between parties based on the circumstances of the transaction along with the applicable laws and regulations. With blockchain, discussed in detail later in this paper, a public ledger with complete, irreversible history of transactions are visible to all the members of the network. A consensus on the ledger is the integral feature of a blockchain network that makes third parties unnecessary.

With the advent of the first cryptocurrency, Bitcoin, in 2009, the traditional centralized system of financial attestation has been challenged. Cryptocurrencies operate within a distributed ledger platform in which transactions are validated by other nodes in the network via public key cryptography [1]. Despite the tremendous potential of establishing a secure environment for global finance, blockchain has afforded terrorist and organized crime entities a less regulated means of laundering money from illegitimate sources. Developing a user reputation score for cryptocurrency transactions would enable a marketplace in which buyers and sellers can trust each other.

As reputation systems for transactors of cryptocurrencies would assist in identifying potential criminal actors, we present an algorithm to authenticate the identity of transactors. This makes the cryptocurrency marketplace more secure and trustworthy. Based on historical Bitcoin data at the transaction level, we investigate the history of ownership of coins in order to identify transaction patterns as a differentiator between lawful and unlawful transactors. For this purpose, we have identified a number of criminal activities in the past. We profile these incidents and replicate the model in our dataset to identify the patterns of transactional behaviors. These patterns will serve as the basis of a reputation score for each participatory entity.

The remainder of this paper is structured as described here. Section 2 provides an overview of the concept and history of cryptocurrency as well as data on the digital currency marketplace. The architecture of blockchain as a distributed ledger technology are explained in Section 3. Section 4 describes the fields of the dataset we use and how we prepare the data for our analysis. In Section 5, we detail our methodology for building a reputation score for cryptocurrency users. In Section 6, we apply a [insert advanced statistical model here] model to develop reputation scores and analyze the results. Section 7 examines the ethical implications of profiling and scoring cryptocurrency transactors. In Section 8, we conclude by mentioning future areas of research that can add further value to the work we have done for this paper.

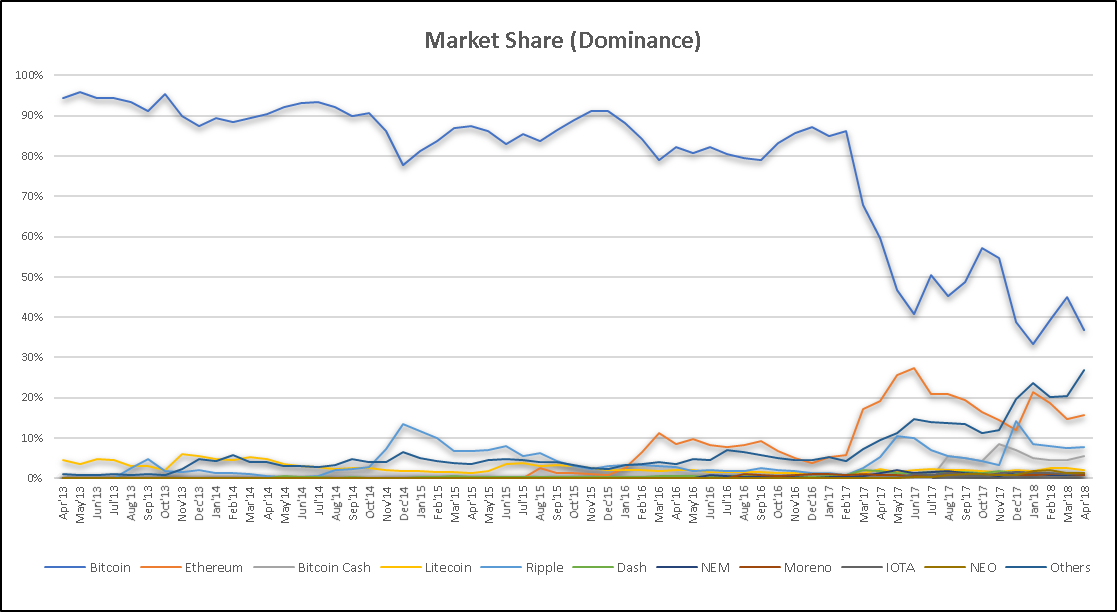
2 Cryptocurrency

The world’s first cryptocurrency, known as Bitcoin, was conceptualized by Satoshi Nakamoto (a pseudonym of a person or a group of people) in 2008 [1]. 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 peer-to-peer distributed ledger, challenges the 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. The key features of blockchain are described in Table 2.1 below.

**Table 2.1.** Features of blockchain. [3]

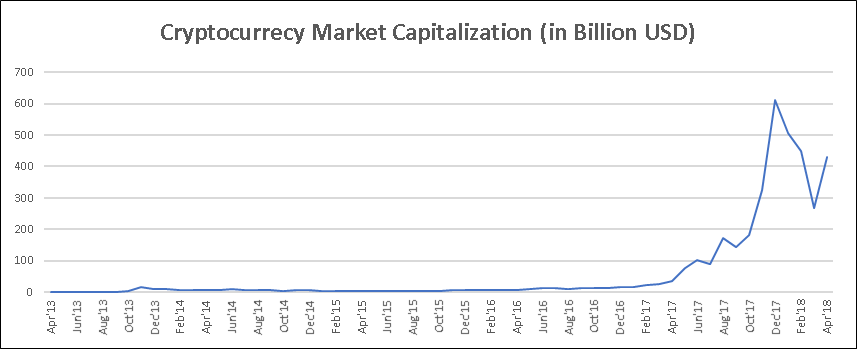
|  |  |
| --- | --- |
| Feature | Example |
| Replicated Ledger | All participants in the network have a replication of the ledger where all the historical transaction details are recorded. |
| Cryptography | Authentication and privacy of transactions are ensured by cryptographic encryption mechanism. |
| Consensus Logic | The consensus protocol, Markle Tree, is used to maintain blockchain’s data integrity. |

Since the innovation of Bitcoin in 2009, the world economy has observed the introduction of numerous cryptocurrencies to market such as Litecoin, Ethereum, Ripple, Dash and Monero, to name a few. 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. Being the debutant in the industry, Bitcoin had enjoyed almost a monopoly with over 70% market share of global cryptocurrency industry until early 2017. Slowly other players entered the market, and, to date, there are more than 1,500 types of cryptocurrencies operating in the market. Since its inception in 2013, Ethereum, featuring smart contracts, has emerged as the biggest competitor to Bitcoin based on market share. As of April 2018, Bitcoin captures almost 37% of the cryptocurrency market, followed by Ethereum (16%), Ripple (8%) and Bitcoin Cash (6%) (Figure 2.1).



**Figure 2.1.** Global market share of major cryptocurrencies [4]

The industry has experienced an exponential growth in the number of active users in the past few years. The current number of users is reported to be more than 23 million and is expected to cross 200 million by 2024 at the present growth rate [5]. At the beginning of Q2 2018, the market capitalization of cryptocurrencies was over $430 billion (Figure 2.2). By the end of 2017, total Bitcoin supply crossed 16 million coins with an average 24-hour trade volume of $4.9 billion across 96 countries where the usages are unrestricted.

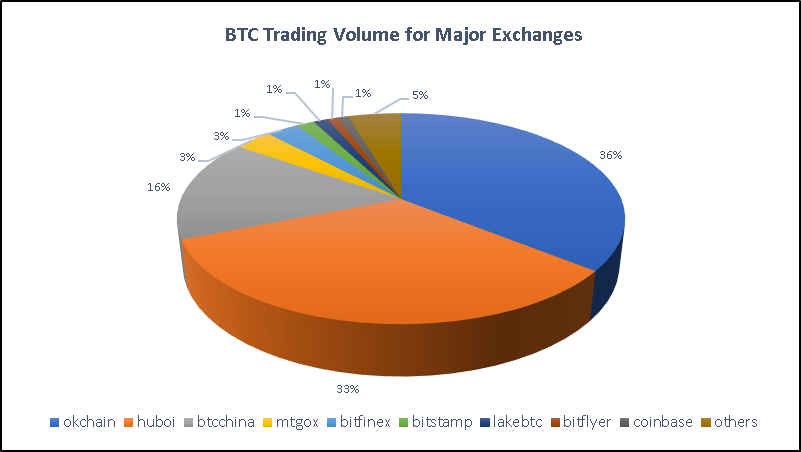


**Figure 2.2.** Total market capitalization for cryptocurrencies

Exchanges are an essential component in cryptocurrency transactions. These exchanges, spread across the globe, offer an online marketplace where customers can buy and sell cryptocurrencies against assets like traditional money or other cryptocurrencies. The cryptocurrencies bought through an exchange are stored in customers’ cryptocurrency wallets. A wallet generally stores the public and private keys required to make a successful transaction of crypto assets and monitors customers’ funds. Mt.Gox was the first exchange to be established in 2010. Due to comparatively relaxed regulations, exchanges located in the Asia-Pacific region became more popular over time for cryptocurrency trading. Exchanges such as OKChain, Huobi and BTCChina have been widely used in recent years (Table 2.2 & Figure 2.3).

**Table 2.2.** Trade volumes and market share for major exchanges. [6]

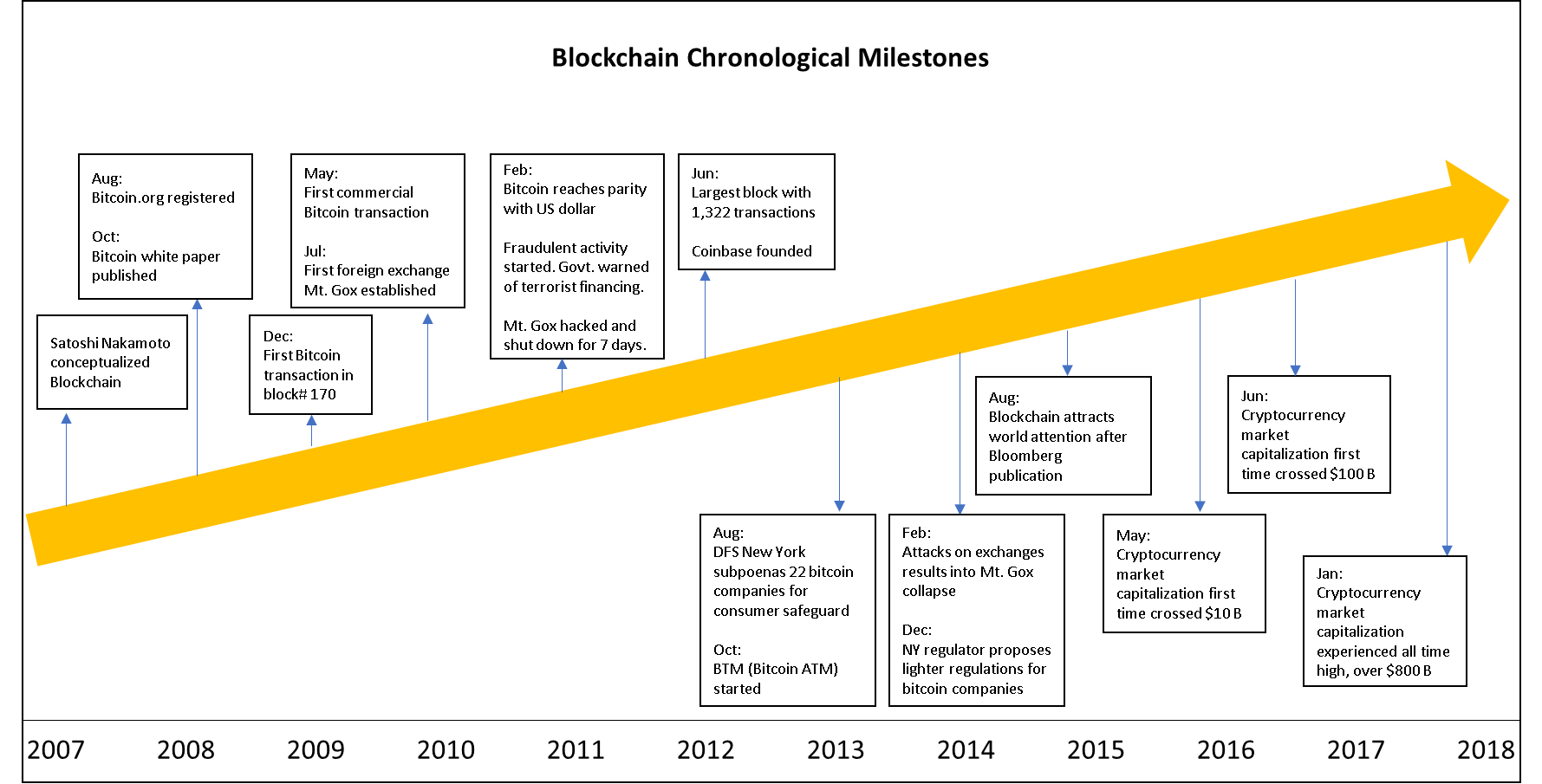
|  |  |  |
| --- | --- | --- |
| Exchange | BTC Volume | Market Share |
| okchain | 586,565,939.90 | 35.76% |
| huboi | 544,852,241.80 | 33.21% |
| btcchina | 258,007,950.80 | 15.73% |
| mtgox | 53,861,912.21 | 3.28% |
| bitfinex | 45,461,276.87 | 2.77% |
| bitstamp | 25,910,946.75 | 1.58% |
| lakebtc | 20,767,853.48 | 1.27% |
| bitflyer | 15,022,795.60 | 0.92% |
| coinbase | 14,915,059.72 | 0.91% |
| others | 75,147,603.32 | 4.58% |



**Figure 2.3.** Popular exchanges for trading

With ever-growing popularity of cryptocurrency trading, there comes a risk of security breaches and attacks. The crypto industry has matured over time but on the other side it has opened doors for channelizing illegal assets. Criminals often consider this market place as a potential target for their illicit activities due to lack of regulations and compliance for cryptocurrency operations. Money sourced from different criminal activities are stored in form of cryptocurrencies to avoid being tracked. The crypto world has experienced several attacks and thefts which caused considerable loss of funds. The largest bitcoin exchange, Mt. Gox, was hacked in early 2014 where bitcoins valued at approximately $460 million were stolen by hackers. This incident forced them to file for bankruptcy. Recently, in January 2018, it was reported that NEM tokens of worth $533 million was stolen from Tokyo based exchange Coinchek [7]. This industry is almost unregulated which makes it extremely difficult to hold any authorities accountable in case of any such scam. Due to the anonymous nature of transactions, senders and recipients are unidentifiable which is the main road-block for implementing KYC and AML compliance. Moreover, since exchanges and wallet services operate across the globe, regulating them under any geographical jurisdiction of law enforcement is quite difficult.

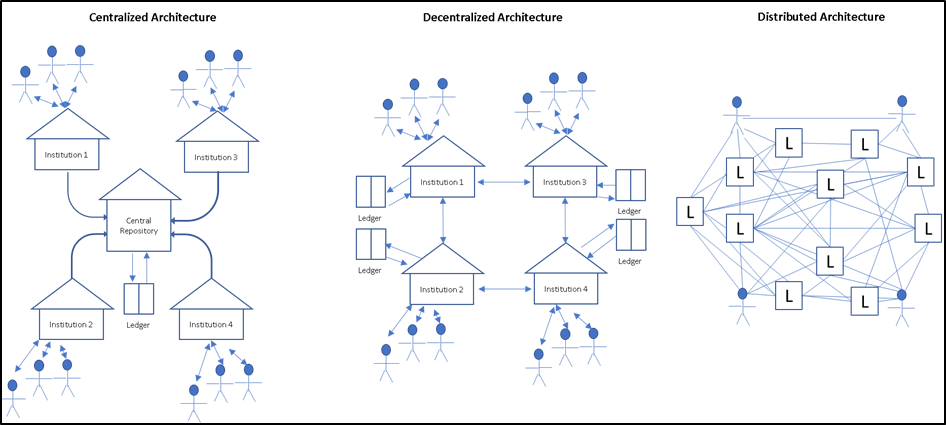
The diagram below (Figure 2.4) illustrates the major milestones that have occurred in the roughly decade-long existence of the cryptocurrency world.



**Figure 2.4.** Cryptocurrency milestones [8]

3 Blockchain Technology

Blockchain is a distributed ledger with no single point of control which was first introduced in 2008 by a person or a group of people with the pseudonym Satoshi Nakamoto. Though the concept of blockchain was associated with the innovation of the first cryptocurrency named Bitcoin, it has been widely used for other coins as well. Unlike a centralized system, Blockchain never requires any central repository of financial assets. This digital ledger is distributed across multiple computers or servers, known as “nodes”, connected over a peer-to-peer network (Figure 3.1). Any transaction made on blockchain network are secured with a cryptographic encryption mechanism. 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. This makes the length of the chain ever-growing.

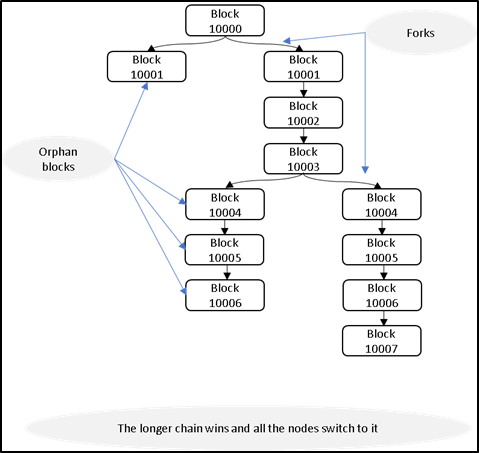


**Figure 3.1.** Ledger Architecture

Before exploring a detailed operating model, we define a few important terms in Table 3.1:

**Table 3.1.** Definitions of frequently used terminologies.

|  |  |
| --- | --- |
| Terminology | Description |
| 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. |
| 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. |
| 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 (Figure 3.2). |



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

The steps to run the network are as follows [1]:

1) New transactions are broadcast to all nodes.

2) Each node collects new transactions into a block.

3) Each node works on finding a difficult proof-of-work for its block.

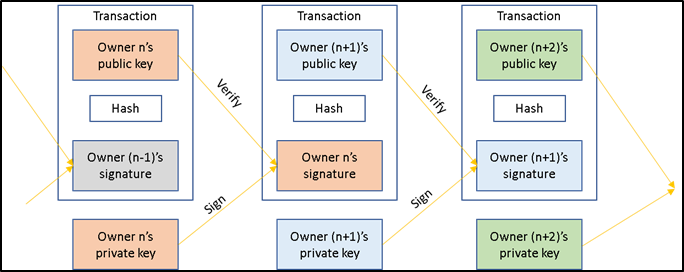
4) When a node finds a proof-of-work, it broadcasts the block to all nodes.

5) Nodes accept the block only if all transactions in it are valid and not already spent.

6) Nodes express their acceptance of the block by working on creating the next block in the chain, using the hash of the accepted block as the previous hash.

Each block in a blockchain is timestamped and contains a set of validated transactions which are appended in order. These transactions, the timestamp, a nonce and a SHA-256 hash value of the previous block form the current block. The block then gets appended to end of each chain in the network. Sequentially chained transactions are secured with a digital signature based on public-private key encryption (Figure 3.3) that enables a payee to verify the chain of ownership. Timestamps on the block avoid the issue of duplicate transactions.

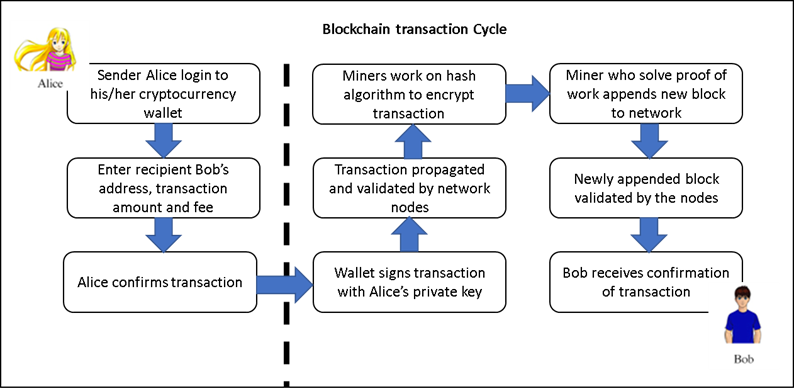
The difficulty of the proof-of-work is a measure of how challenging it may be to find a hash below a given target. A target is a 256-bit number that every client shares. In order for a block to be accepted to the network, the SHA-256 hash of the block must be lower than or equal to the current target [14]. Each individual block stores a representation for the hexadecimal target (e.g., 0x1b0404cb) which can be derived.



**Figure 3.3.** 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 is considered in the blockchain network to save disk space with an ever-growing network.

A typical cryptocurrency transaction on a blockchain network is illustrated in Figure 3.4.



**Figure 3.4.** Blockchain transaction cycle [9]

Table 3.2 describes the key advantages of a blockchain network.

**Table 3.2.** Blockchain advantages.

|  |  |
| --- | --- |
| Advantages | Description |
| 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 trans[[1]](#footnote-2)action data from Kaggle [9]. The data has 21 attributes and 313,423,709 rows. After taking a random sample and cleaning the data there are now 5 attributes and 5,419,268 rows. The 5 attributes are: timestamp, transaction ID, input address, output address and the satoshis. These are key attributes needed for mapping the transaction history and computing a reputation score. There are 2 additional attributes that will be generated based on the 5 listed above: a quantifiable metric to score the individual transactions per address and the reputation score for the address. The computation of the 2 additional attributes will be described later in the section.

Every transaction is a simple transfer of a certain amount of Bitcoin from a sender to a receiver. The transaction will reference previous transaction outputs as the new transaction inputs and then dedicates all input coin values to new outputs. Since transactions are not encrypted, every single transaction can be seen inside a block.

**Table 1.** Outline of the transaction level data.

|  |  |  |
| --- | --- | --- |
| Field | Description | Example |
| Timestamp  Transaction ID | Current time as seconds since 1970-01-01T00:00 UTC.  A hashed value that represents a given transaction. | 1489132192 is equivalent to 03-10-2017 T07:49 UTC  943108050c5ff069e7c3177ab105791faeb7a7634d13f05a9915cd1d89326160 |
| Input Address | An identifier of 26-35 alphanumeric characters that represents the sender. The format of the address here is P2PKH. | 138Q5rFKXdfjVEgz82LZ2tLS6JeLXUaioU |
| Outputs Address  Satoshis | An identifier of 26-35 alphanumeric characters that represents the receiver. The format of the address here is P2PKH.  The amount of Bitcoin sent from the Input Address to the Output Address. | 1Miuy5qySDSo95TL1uizgEYStSKwTaXiH9  1357920 BTC |

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 [12]. 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 [12].

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 [13]. 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 [14].

5 Our Approach for a Reputation Score

A reputation score is a gauge of a transactor’s trustworthiness. Users are unlikely to engage in transactions with an entity with a low reputation score. The steps we take to build a reputation score are profiling transaction histories, flagging transactions, and creating a blacklist of transactors. The score serves as a differentiator between transactors and helps them to improve their reputation score by avoiding bad actors in the network.

**5.1 Profiling Transaction History**

When building the reputation score, there are two major attributes that act on each other for generating the actual score value. A user’s score will be based on their addresses and transactions. Each transaction will be flagged and scored which then, in turn, will change the overall score of the address. Creating this profile of transaction history will allow certain users to be categorized as a good or bad actor.

There are four parameters that are taken into consideration when distinguishing a bad actor from a good actor: frequency of transactions, amount of the transaction, the ratio of inputs to outputs and whether it is a newly minted coin that goes to the same input address. The behavior of these parameters will determine the type of flag the transaction will be given. The flags allow us to compute the user’s reputation score.

**5.2 Flagging Criteria**

The flagging of transactions will be the backbone of the reputation score. This step is where each transaction will be categorized into its respective category. If a transaction is of a newly minted coin that goes to the same input address, then the transaction will be flagged as a pooled transaction. A pooled transaction occurs when a group of miners combine their hashing power together and then split the coins equally based off the amount of shares a user contributes to solving the block [18].

The amount of the transaction is also a flagging criterion. The average transaction can vary from less than 1 to about 50 Bitcoin [10]. Therefore, if there are transactions with a high amount of BTC’s, the transaction will be flagged. Transaction behavior of past criminal incidents will allow us to know how the bad actors behave. Profiling such behavior allows us to flag other transactions with similar behavior.

5.3 The Blacklist

For Bitcoin transactions, there needs to be a baseline to quantify which addresses are associated with illegal activities. Certain transactions are associated with criminal actions and thus the address that made the transaction should not be trusted. Events such as the Mt. Gox hack, May 2012 Bitcoinica Hack, BTC-E Hack, and the Allinvain Theft are all examples of times in which certain addresses were used to commit illegal activities. Therefore, if a user transacts with one of the addresses on the list of bad actors, then the score of that user would decrease.

If transactor A is associated with the Mt. Gox hack and transactor B is associated with transactor A, then the score of transactor B would decrease. This score will be made available to everyone who has an address. Therefore, if transactor C wants to send Bitcoin to transactor B, they could view the score and decide not to send Bitcoin to transactor B because doing so will lower transactor C's reputation score.

This list of bad actors will contain all the major Bitcoin heists, thefts, scams and losses [17]. The list will act as a baseline for identifying actual bad actors. The transactions on the list will receive a low reputation score and thus will lower the reputation score of the address associated with the transaction.

6 Reputation Scoring

**7 Ethical Discussion**

8 Future Work

References

1. Satoshi Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”. Published in 2008 on Bitcoin.org [Online]. Available: <https://bitcoin.org/bitcoin.pdf>
2. Wikipedia on Satoshi Nakamoto. [Online]. Available: <https://en.wikipedia.org/wiki/Satoshi_Nakamoto>
3. D. Godfrey-Welch, R. Lagrois, J. Law, R. S. Anderwald, "Blockchain in Payment Card Systems". Published in 2018 on SMU Data Scoence Review [Online]. Available: https://scholar.smu.edu/datasciencereview/vol1/iss1/3/
4. Coin Market Cap, "Cryptocurrency Market Capitalizations," [Online]. Available: https://coinmarketcap.com/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. Statista, "Number of Blockchain Wallets 2018" [Online]. Available: <https://www.statista.com/statistics/647374/worldwide-blockchain-wallet-users/>
6. Data.bitcoinity, "Bitcoin trading volume," [Online]. Available: <https://data.bitcoinity.org/markets/volume/30d?c=e&t=b>
7. Coindesk, "Coincheck Confirms Crypto Hack Loss Larger Than Mt.Gox," [Online]. Available: <https://www.coindesk.com/coincheck-confirms-crypto-hack-loss-larger-than-mt-gox/>
8. Next.autonomous, "Blockchain Blockbuster," [Online]. Available: <https://next.autonomous.com/blockchain-blockbuster/>
9. Researchgate, "Life Cycle of Bitcoin Transaction," [Online]. Available: <https://www.researchgate.net/figure/LIFE-CYCLE-OF-BITCOIN-TRANSACTION_fig2_318850089>
10. Bitinfocharts, "Bitcoin Avg. Transaction Value Historical Chart," [Online]. Available: https://bitinfocharts.com/comparison/bitcoin-transactionfees.html
11. 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
12. Blockgeeks, “What is Blockchain technology? A step by step guide for beginners,” [Online]. Available: https://blockgeeks.com/guides/what-is-blockchain-technology/
13. R. Franco, Chapter 7, “Understanding Bitcoin Cryptography Engineering and Economics”, 2014.
14. bitcoinwiki, "Difficulty," [Online]. Available: https://en.bitcoin.it/wiki/Difficulty.
15. bitcoinwiki, "Transaction," [Online]. Available: https://en.bitcoin.it/wiki/Transaction.
16. M. Kiran, M. Stannett, "Bitcoin Risk Analysis," NEMODE Policy Paper, 2015, [Online]. Available: http://www.nemode.ac.uk/wp-content/uploads/2015/02/2015-Bit-Coin-risk-analysis.pdf
17. Bitcoin Forum. "List of Major Bitcoin Heists, Thefts, Hacks, Scams, and Losses," [Online]. Available: https://bitcointalk.org/index.php?topic=83794
18. Bitcoinmining. "Bitcoin Mining Pools," [Online]. Available: https://www.bitcoinmining.com/bitcoin-mining-pools/

1. https://www.kaggle.com/bigquery/bitcoin-blockchain [↑](#footnote-ref-2)