Blockchain Databases

Isabelle Schmit
University of Rochester
MS Data Science
Rochester, NY
Email: ischmit@u.rochester.edu

Thomas Varner
University of Rochester
MS Data Science
Rochester, NY
Email: tvarner@u.rochester.edu

ABSTRACT
Blockchain Databases were introduced in 2008 by Satochi Nakamoto as a precursor to Bitcoin, an online cryptocurrency. These databases allow for secure, peer-to-peer online transactions that are reconciled with a distributed public ledger of transaction history. The theoretical concept of Blockchain is based on those of hashchains and public key cryptography, and blockchains are implemented using mutual distributed ledgers with a network of functional nodes to perform tasks like routing, storing, and mining. Blockchain benefits from being decentralized and pseudonymous, removing the need for users to trust a third party, and can be applied in situations outside of Bitcoin and finance, such as for online document authentication and smart contracts.

1. INTRODUCTION
Nearly all transactions carried out today, from financial transactions to notarized documents to government contracts, are completed using third party. While trusted third parties are standard, many individuals worry about the security of those transactions, and whether the trusted third party can truly be trusted. These parties, such as banks, government agencies, or notaries, are known to take actions for their own gain, rather than that of their customers, so many individuals have sought out ways to eliminate third parties in various types of transactions. Blockchain databases were introduced as a method to remove intermediary third parties, while also solving other issues that previous peer-to-peer transactions failed to do, like the double-spending problem for financial transactions. By using a distributed and openly available ledger system, blockchain databases provide a promising alternative to trust-based third party transaction systems in a wide variety of applications.

2. BITCOIN
Bitcoin was the first example of blockchain technology to gain significant public traction and widespread use, hence it serves as a convenient case study into the implementation of a blockchain database as well as the implications for society. The cryptocurrency was theorized in 2008 by Satoshi Nakamoto, and released for market-use in early 2009. Tokens known as bitcoins with real market value could be exchanged through a highly-secure online protocol preserving anonymity. A bitcoin itself is a solution to a mathematical formula, with a theoretical limit guaranteeing no more than 21 million bitcoins will ever be found. In November 2013, the price of one bitcoin peaked at $1,124.76, with a market capitalization of $13.9 billion [6]. Bitcoin operates using a public ledger of all transactions, this ledger can be used to determine whether or not any proposed transaction is valid or not. At launch, all machines that participated in the network performed two important functions: verifying proposed transactions and mining new solutions to the equation, though the roles of machines in the network soon bent to match the evolving demands of the market [10]. The blockchain technology allows bitcoin to automatically validate transactions before committing them, safeguard transactions against fraudulent activity, and maintain a historical record of all transactions that have ever been performed [6]. For non-cryptocurrencies, these operations are performed by a trusted third-party, introducing inefficiency and room for human error and corruption into transaction-based systems. Free of these timeless constraints, bitcoin introduced the world to the paradigm shifting technology of blockchain networks.

3. SUPPORTING TECHNOLOGY
The conceptual structure of blockchain is built on the concepts of hashchains and public key cryptography.

3.1 Hashchains
Blockchain technology is built on the foundation of hashchains, sequences of data blocks linked by a hash function. Hashchains are formed by feeding data into a hash function then representing a block as a hash that points to a payload. Given hash function h(x) of length n for any string x, the hashchain would follow the structure $h_1(h_2(...(h_n(x))...))$ which can be denoted as $h^n(x)$. In the first block, simply the payload is hashed. Once the first block is formed, the next block in the hashchain is constructed by feeding the entire previous block, including its hash, into the hash function, yielding the hash key for the next block. For all successive blocks, both the hash and payload of the previous block are fed into the hash function, so the last block in the chain's hash depends on all previous blocks.

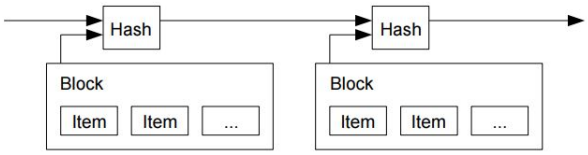

Figure 1: Hashchain Structure

This structure enforces one of the most important aspects of blockchain, that the data of previously added blocks cannot be changed without changing the entire hashchain [7].

3.2 Public Key Cryptography
While hashchain structure is helpful, it does not ensure that blocks within the hashchain cannot be changed, it simply enforces that attempts to change a block would change the entire hashchain. To add true immutability, blockchain uses public key cryptography to ensure that only one user can add a new block to the hashchain.

To do this, the current user's public key and the previous user's digital signature are stored within each block, so the current user can create the next block with another user's public key, then use their own private key to sign the block and pass control onto the next user [7].

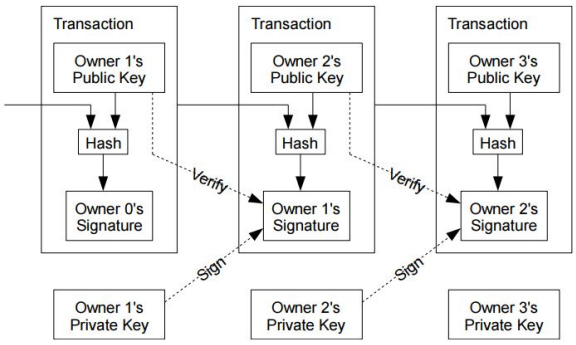

Figure 2: Public Key Implementation in Hashchains

This adds more security to the hashing function, because not only are the hash and payload put into the hash function, the owner's public key is as well. By using the owner of the current block's public and private key to create, verify, and sign the next block, this implementation of public key cryptography keeps any user from changing previous transactions, allowing the hashchain to be truly unchangeable [8].

4. BLOCKCHAIN STRUCTURE

4.1 Conceptual Structure

The structure of Blockchain combines hashchains with public key cryptography by creating a global hashchain where blocks contain internal hashchains. The global hashchain is constantly creating new blocks and old internal blocks cannot be added to new external blocks [7]. In Satochi Nakamoto's original paper introducing the blockchain, in each block of the global hashchain, the internal hashchain is stored in a Merkle Tree that can be pruned to store only the most relevant information [8].

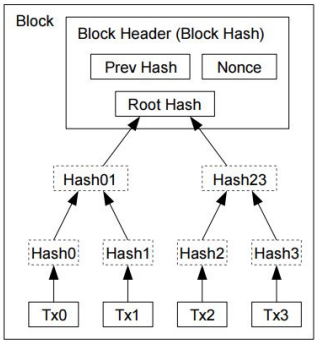

Figure 3: Structure of an External Block in the Global Hashchain

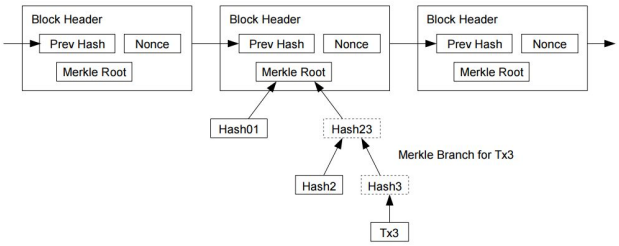

Figure 4: Structure of the Global Hashchain with Pruned Trees

4.2 Network Structure

Blockchain databases are structured as mutual distributed ledgers (MDLs). MDLs are mutual by allowing all users full access to the contents of the database, distributed in that the database can be stored either in part or in full in many decentralized locations, and ledgers in that records of all previous transactions are maintained [6]. The MDL is shared among a "peer-to-peer" network of computers, where each computer is a node and edges connect each computer to many other computers. The network typically takes the form of a single giant connected component, devoid of small cliques or isolated nodes. This allows the entire network to remain up-to-date and maintains the notion of consistency [5].

Nodes in the network are all considered equivalent to each other, and may take on various functionalities depending on the nature of the blockchain network; these may include routing functionality, mining functionality, blockchain storage functionality, and endpoint functionality. All nodes in a blockchain network must have routing technology; this allows nodes to propagate transactions throughout the network through the edges that connect the nodes. Routing nodes also bear the responsibility of validating new transactions before committing them to the blockchain. This requires communicating with other nodes in the network to receive a copy of historical transactions necessary to validate a new transaction. Once a transaction is validated, proof of validation is shared between routing nodes to that the new transaction can be added to the database. Nodes may also have blockchain storage functionality, where they maintain historical transactions. "Full" nodes carry a complete copy of the ledger, and are are capable of validating a new transaction without querying any other nodes for transactions. The network as a whole is not dependant on full nodes, as valid transactions can still be verified by routing nodes that use composite copies of the ledger it obtained from several sources; this gives blockchain networks consistency as well as resilience to attack [1]. Nodes with mining functionality are responsible for creating new blocks by solving the proof-of-work algorithm. Full nodes are capable of mining in isolation whereas other nodes participate in pooled mining, an automated process in which a central node autonomously manages the distribution of mining tasks among nodes in the pool [10]. Nodes with endpoint functionality allow users to interact with the database by querying information or requesting a transaction.

The role a node plays in the network is based on its composition of these four functionalities. Using bitcoin as an example, a popular open-source reference client for functionalizing nodes in the network is known as Bitcoin Core, which comes with routing, storing, mining, and endpoint technology. Storage nodes only contain routing and storage functionality; these nodes are responsible for holding the distributed copies of the database

throughout the network. Mining nodes contain routing, mining, and often storage technology; mining farms range in size from a single hobbyist machine to thousands of machines in data warehouse [10]. Simplified payment verification nodes are lightweight nodes consisting of routing and endpoint technology; these are typically run on a desktop or smartphone whenever a user would like to request a blockchain transaction [1]. Beyond bitcoin, a blockchain database could be built atop network of nodes with a more diverse set of functionalities.

5. BENEFITS OF BLOCKCHAIN

One of the main benefits of using blockchain databases for various applications is their decentralized nature. Unlike other systems, which are owned and maintained by a single entity, blockchains are freely owned and available as part of a network using distributed ledgers. Users feel more comfortable using blockchain because they do not have to trust a corporation or bank to keep their transactions, they trust the structure of the blockchain and the individuals that maintain it, who are not an organized entity and do not benefit from the transaction, but have collective self interests to maintain the authenticity of the blockchain. In financial situations, blockchain is seen as ideal by some because it removes the need for banks, which are often motivated by profit. Before blockchain was introduced, some central authority was necessary to solve the double spending problem, where the same money or coins cannot be spent in multiple transactions, so transactions are traditionally checked by a mint. Due to public key cryptography, the blockchain eliminates the need for a trusted third party like a bank, which is why it has been so successful in the financial sector [8].

The other main benefit of blockchain is that it ensures anonymity to those involved in transactions. Because blockchain require a public-private key pair to create and authenticate transactions, each user has a pseudonym, so the history of their transactions can be tracked, but their public key cannot be traced to the users actual identity. In traditional banking systems, accounts are linked to user's social security numbers and other identifying information, so blockchain, and more specifically Bitcoin, are ideal for those who want to keep their online transactions and spending habits private and secure [9].

6. APPLICATIONS OF BLOCKCHAIN

By providing a framework that is decentralized, resilient to cyberattack, and ensures anonymity, blockchain databases have been identified as potentially "radically disruptive" to existing technologies and markets [4]. Blockchain databases could make document authenticity easy, removing the need for a centralized authority to verify the existence, ownership, and integrity of a document [3]. This effectively removes the need for a third party such as a notary public to examine a document and make a decision about its validity, speeding up the process while ensuring its integrity. As many of these processes are conventionally carried out by a trusted third party such as a government, many functionalities of a government could be replaced by a technologies built on blockchain databases; a special report by the UK Government Office For Science concluded that "[Blockchain] has the potential to redefine the relationship between government and the citizen in terms of data sharing, transparency and trust," [4]. The music industry is still working out how to keep track of royalties for artists, labels, and songwriters in the era of streamable services; one possible solution could be to manage music royalties with smart contracts stored in a blockchain database. Smart contracts are automatically enforced electronically, a process that is enabled with blockchain technology [3].

Although blockchain technology is still in its infancy, several applications beyond bitcoin are already on the market for consumption and development. Etherium is a blockchain-based distributed computing platform that uses virtual machines to run scripts enabling development, and compensates nodes in the network for their computational efforts with a value token called "ether" [2]. Applications are currently being developed on the ethereum platform that introduce novel approaches of blockchain technology to areas including governance, autonomous banks, smart contracts, and keyless access. One such company, Everledger, maintains diamond certification with a blockchain database, storing physical characteristics of each diamond as well as supplying an API for creating, reading and updating claims [3]. Even large companies including IBM, Samsung, Overstock, Amazon, UBS, Citi, Ebay, and Verizon Wireless are currently researching how blockchain technology could be used to improve their respective services [3]. In the age of online processing, blockchain databases could allow our transaction to be faster, safer, more secure, and autonomous.

7. CONCLUSION

Many transactions that we make on a day to day basis require the approval of a trusted third-party. Transferring money, notarizing documents, and finalizing contracts require time and money while introducing room for corruption and human error. Blockchain databases make direct peer-to-peer transactions possible, bypassing the need for a trusted third party while maintaining the anonymity which we require for much of our online information. Blockchain's consensus-based voting system makes it easy to spot and remove fraudulent activity while building a historical transaction record. These characteristics make blockchain databases powerful tools with the potential to change the way we share information at the height of the digital era.

8. ACKNOWLEDGMENTS

Our thanks to Professor Biswas and the TAs.

9. REFERENCES

[1] Antonopoulos, A. M. (2014) Chapter 6: The Bitcoin Network, Mastering Bitcoin. O'Reilly Media, 1st Edition.

[2] Buterin, V. "White Paper · ethereum/wiki Wiki · GitHub" 2014.

[3] Crosby, Nachiappan, Pattanayak, Verma, Kalyanaraman. "Blockchain Technology: Beyond Bitcoin" Applied Innovation Overview. Issue No. 2 June 2016.

[4] Hancock, H. M., Vaizey, E. "Distributed Ledger Technology: beyond block chain" Government Office for Science, UK (2015).

[5] Harrigan M. "A Network Analyst's View of the Block Chain". Coin Desk. May 2014.

[6] Mainelli, M., Smith M. "Sharing ledgers for sharing economies: an exploration of mutual distributed ledgers" The Journal of Financial Perspectives. Winter 2015 | Volume 3 – Issue 3.

[7] Mazonka, Oleg. "Blockchain: Simple Explanation" Journal of Reference, 2016

[8] Nakamoto, Satochi. “Bitcoin: A Peer-to-Peer Electronic Cash System” bitcoin.org, 2009

[9] Swan, Melanie. “Blockchain: Blueprint for a new economy” O'Reilly Media, Inc., 2015.

[10] Swanson, T. “Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems.” Report, available online, Apr 2015.