Blockchains, a Survey

John Edgar

CMPT454

Justin Liu - jzl1 - 301070053

What is a blockchain?

A blockchain, in its essence, is an immutable public ledger of transactions that have been expedited among all participating parties. Five key characteristics found in any type of blockchain is decentralization, persistency, auditability, user anonymity, and transparency. All transactions are appended to the ledger through a majority consensus among the participants in the chain. Once a transaction is added to the ledger, it cannot be removed. This decentralized environment is only possible through the support of technologies including cryptography and consensus algorithms. Asymmetric cryptography is the foundation that blockchain technologies build on to preserve anonymity and lend support to auditability in an enigmatic environment. The infrastructure which allows for the distributed and persistent nature of a blockchain ledger is largely rooted in a consensus algorithm, which ensures synchronicity in an asynchronous environment. The final, and often most reveled characteristic, transparency, is the result of a completely public ledger that any party involved in the chain are privy to view or make copy of.

As Bitcoin has brought a global spotlight on blockchain databases and may be heavily attributed to the continued enthusiastic development of the technology today, it is only fitting that this report will make reference to Bitcoin specifically to illustrate the anatomy and functionality of blockchains in general.

Anatomy of a Blockchain

Conceptually, a blockchain may be represented exactly as its name implies; consecutive blocks linked together by a chain. Each block holds multiple transactions, of which are not necessarily related, and are considered to have occurred simultaneously. Given a block, all ancestor blocks in the chain are considered to have occurred before it and any progenitor blocks are considered to have occurred after.

Block Contents (non-exhaustive)

Block Header
Merkel Tree Root
Parent Block Hash
Timestamp
Difficulty
Nonce
Block Transaction History

Merkel Tree Root: the hash value of all the transactions in the block.

Parent Block Hash: the unique hash value of the block immediately preceding the block being considered.

Timestamp: date and time of the block's creation

Difficulty: the condition set by the block chain which dictates the range of an acceptable block hash value (ie. hash must have five leading zeros).

Nonce: a variable set by Miners such that they can produce a hash value for a block satisfying the condition given by Difficulty (discussed in detail in the Mining subsection).

Database Management

Network

A blockchain is a decentralized public ledger. The network is represented by nodes which can be differentiated into three types: system nodes, client nodes, and mining nodes. System nodes contain full or partial copies of the entire chain and provide security to the network by enforcing the blockchain's protocols and verifying "candidate blocks". Client nodes are endpoint devices that utilize the blockchain's services but are not involved in verifying and validating transactions as they do not necessarily hold copies of the chain. In Bitcoin, client nodes are associated with wallets which are uniquely identified as hash values of their public keys. Blockchain client nodes have some form of account, as in Bitcoin, to hold and index all transactions involving the account to their respective blocks they are stored in the chain. Mining nodes are endpoints that are involved in creating "candidate blocks" that are considered by system nodes and appended to the chain after a final validation process.

Transaction Security

To execute a transaction, the two parties utilize asymmetric cryptography to encrypt sensitive transaction data. Asymmetric cryptography requires a sender to first digitally sign the data by encrypting it with the sender's private key, then again encrypt the data using the receiver's public key. When the data is transferred to the receiver, the receiver first uses his private key to decode the data, and then uses the sender's public key to decrypt the data which had been digitally signed. This process not only ensures confidentiality, but confirms the identities of both the sender and receiver, as for the receiver to decrypt the data they must know who the sender is (they must know who's public key to use), and the sender must know who the receiver is since they are using their public key to encrypt the data in the first place. Once a transaction is completed, it is broadcasted to all nodes in the blockchain for verification. The verification process depends on the blockchain, but in the case of bitcoin it requires two steps:

- verification that the sender owns the cryptocurrency
- the sender owns enough currency in his online wallet

Verification that a sender owns the cryptocurrency is done by system nodes, which would simply verify a sender's wallet address (the hashed public key of the sender) decrypts the data signed by the

sender's private key. The verification that the sender owns enough currency is done by iterating over the entire chain to determine the wallet balance. Once verification is completed, the transaction is broadcasted to all system nodes in the chain. Each node will then have a collection of transactions that miners will draw from to package into candidate blocks, later to be validated as per the chain's consensus algorithm.

Mining

Utilization of asymmetric cryptography allows for individual transactions to be verified for validity. Blocks containing transactions however still need to be created and verified, which is done respectively by miners and consensus among the participants in the system. An example of a consensus algorithm would be bitcoin's Proof of Work (PoW) consensus strategy. PoW is an algorithm which begins with forming a "candidate block" out of a collection of transactions through calculating a valid block hash value and ends with block validation in a system node. The hash value is determined through a hash algorithm which considers the hash of all transactions contained in the block (the Merkel tree root) and a variable input called a "nonce". A Miner changes the value of the nonce to recalculate a hash value for the block until it satisfies the condition set by the difficulty value in the block header. Once a valid hash is "mined", the block is broadcasted to all nodes in the blockchain for block validation.

Calculating the hash for the block is designed to be purposefully computationally intensive, aiming to limit the validation rate of blocks to a target average. Complexity in the hash computation stems not only from the execution of the algorithm that outputs a value, but by a difficulty value set by the blockchain's protocols, set in the block header. Given the computational overhead of the PoW algorithm, Miners are generally compensated in some way to incentivize continual support.

Validity

As discussed above, transactions are first verified by system nodes and then passed over to mining nodes to package them into a block and determine a block hash. Once a block has been filled and the PoW algorithm applied, the block is broadcasted back to all system nodes in the chain for block validation.

Block validation is a necessity as it is possible for a malicious Miner to return an invalid block; such a block may have altered transactions or an invalid block hash value. All system nodes would validate the candidate node by verifying all transactions again, and then test that the nonce value provided in the block header may be applied to the block hashing algorithm to arrive at the mined block hash. Only after a block is mutually validated by all system nodes is it appended to the chain.

Given blockchains are a decentralized system, there is the possibility that multiple valid blocks are appended to a copy of the chain before all system nodes have synchronized to the addition of their new block. Firstly, it is generally considered rare that multiple valid blocks contend for the same position in the chain (Zheng and Dai, 2018). A blockchain protocol generally aims to minimize such conflicts by altering the Difficulty target on new block headers, tightening the definition of a valid block. Such an example would be Bitcoin's target average time for a new block to be generated every 10 minutes (Zheng and Dai, 2018). Should two blocks be appended simultaneously on different system nodes, they may simply fork and continually build on their own unique chain. As determined by the specific chain's protocol, the first fork to append a predetermined number of new blocks would be the accepted branch

and broadcasted to all other system nodes to re-synchronize. Any other forks are orphaned and discarded.

System Security

It was discussed earlier that the PoW strategy, in addition to regulating the flow of new blocks into the system, also provide a layer of security against malicious actors during the block validation phase. The conjunction of the PoW algorithm and forking procedure also introduces another layer of security to the overall system. Should a malicious actor attempt to introduce a fictitious transaction into the chain, they would first need to pass transaction verification (twice), then race to compute a valid hash value for a block containing the transaction, and then also generate enough additional blocks to ensure the transaction even makes it into the accepted fork (Crosby et al., 2016)!

Challenges

Scalability

As the number of transactions continue to be appended to a blockchain and the chain becomes bulky, storage of the database may become an issue. Since the entirety of the distributed database is stored in each system node, each node must also have the appropriate hardware support to host an entire database. Multiple proposed solutions to combat this issue have been proposed, of which include decoupling transaction data from the infrastructural data of the block chain (Zheng and Dai, 2018). Though this complication is undoubtedly a major concern, a massive chain such as Bitcoin has yet to breach 300 gigabytes.

Anonymity

One of the major advantages of a blockchain is its capacity to uphold anonymity for its users. It is for this quality that the widespread adoption of blockchain technologies may be threatened by the possibility that users may not be as anonymous as they may believe. Recent issues in blockchain security have been noted ranging from fundamental design issues to user networking practices that threaten user anonymity. Blockchains by design are public, allowing participants to view a full ledger of transactions from the inception of the chain to its end. This design facilitates the necessary infrastructure for countless nodes to interact, validate, and build the database. This, however, may result in a loss of transactional privacy, resulting in the ability for transactions to be linked to reveal user information (Zheng and Dai, 2018). Additionally, since users typically transact from the same general location, a client may be uniquely identified by a set of system nodes it connects to (Zheng and Dai, 2018).

Selfish mining

Blockchains rely on Miners to help create blocks to be appended to the chain after validation. It is not uncommon for Miners, to improve their chances of solving the PoW algorithm before any others, to band together to enhance their overall computing power. A fatal weakness of a blockchain stems from this dependency on Miners and the forking procedure of the blockchain (when multiple system nodes

accept different valid blocks to be appended to the chain, they fork the chain and later accept the longest fork), as "selfish Miners" may fork the chain, hide successfully mined blocks, and privately build the chain until the fork procedure requirements are fulfilled. The result is that non-colluding miners would work on a separate fork, which is later replaced by the selfish Miner's chain. As a result, the honest miner would have wasted their efforts without the possibility of any remuneration (Zheng and Dai, 2018).

Applications

Blockchain technologies today are almost always spoken in reference to cryptocurrencies and their potential to overhaul the financial sector. It was why Bitcoin was chosen as the reference in describing blockchains in this review. The potential for blockchain databases to modernize industries is not limited to the financial sector, but rather to whether the five key characteristics of blockchains (decentralization, persistency, auditability, user anonymity, and transparency) are advantages for the application. Possible fields where a blockchain database may be beneficial include academic publications and management of property.

Academic research requires a formal request for peer review and approval before being published into its respective journal. Traditionally, the reviewers are experts in the field relevant to the author's research but remain anonymous while passing judgment. Though there are advantages to this system, there are also disadvantages such as possible bias introduced by the reviewer(s) or miscommunication between the reviewer, author, and publisher. A blockchain may instead be employed to allow for an open peer review process, where members in the community may openly review and critique each other's work. Members may also be incentivized to participate in peer reviews by disbursing renumeration (ie. cryptocurrencies) or through reputation system. Additionally, methodology and reproducibility are major themes in research and may be accurately represented in an open immutable file system.

Property transference and management can be reduced to a simple blockchain. Management of Vancouver real estate, for example can be greatly streamlined by having property owners as client nodes and transactions be the transfer of property title. As property title is legally in the public domain, it benefits greatly by a transparent transaction records while preserving property owners' anonymity. The immutable file system of blockchains would immortalize the property history, ensuring against fraud as ownership is linked to an account holder, rather than a (non-unique) name and claim.

The Future

Cryptocurrencies have made waves in the financial market, with major banks currently employing blockchain databases in international payments and future designs for in-house cryptocurrencies (Faden, 2019). Where transparency, security, and user anonymity are paramount characteristics for a

database, blockchain designs provide the infrastructure to facilitate these needs. Blockchains are not without their flaws, as the reliance on Miners to act honourably expose the chain to serious security concerns. Scalability issues are another major concern that potentially larger ledgers need to address before instantiating the database. However, even with such issues, blockchains have proven to be an innovative technology in recent years with a wide range of potential applications.

References

Crosby, M., Nachiappan, Pattanayak, P., Verma, S., Kalyanaraman, V. (2016). Applied Innovation Review, Issue No. 2. BlockChain Technology: Beyond Bitcoin.

https://www.edureka.co/blog/how-blockchain-works/

Zheng, Z., Dai, H. (2018). Internation Journal of Web and Grid Services. Blockchain Challenges and Opportunities: A Survey.

https://www.researchgate.net/profile/Hong-Ning_Dai/publication/328271018_Blockchain_challenges_and_opportunities_a_survey/links/5bd2706f92851c6b278f31eb/Blockchain-challenges-and-opportunities-a-survey.pdf

Janowicz, K., Regalia, B., Hitzler, P., Mai, G., Delbecque, S., Frohlich, M., Martinent, P., Lazarus, T. (2018). Semantic Web 9. On the Prospects of Blockchain and Distributed Ledger Technologies for Open Science and Academic Publishing.

https://content.iospress.com/download/semantic-web/sw322?id=semantic-web%2Fsw322

Liu, S. (2019). Bitcoin Blockchain Size 2010-2019.

https://www.statista.com/statistics/647523/worldwide-bitcoin-blockchain-size/

Faden, M. (2019). Banks Move into Cryptocurrency Payments.

https://www.americanexpress.com/us/foreign-exchange/articles/us-banks-support-cryptocurrency-payments/