On the Use of New Blockchain-based Technologies for Securely Distributing Data

Gustavo Alves, Everton Cavalcante, Thais Batista

Federal University of Rio Grande do Norte (UFRN)
Natal, Brazil

{gustavoalvescc, thaisbatista}@gmail.com, everton@dimap.ufrn.br

Abstract. *Blockchain is a distributed encrypted database that maintains a continuously growing set of data records. Aside its increasing popularity, blockchain presents some technical challenges that encourage the development of new technologies attempting to overcome them. Nonetheless, the literature still lacks more research and details about such technologies and which improvements they bring to overcome current challenges of Blockchain. This work aims at providing an overview on Blockchain, its technical challenges, and the new improving technologies.*

1. Introduction

Blockchain can be defined as a distributed encrypted database that maintains a continuously growing set of data records [Banafa 2016]. Despite being a new technology, it is based on the well-known concepts of *encryption* and *distribution*, relying on a model in which networked computers collaborate to maintain a shared and secure database. In such a database, a string of blocks is connected as a chain and distributed through different computational nodes. Each block encapsulates encrypted data records and contains a unique identifier also known as *hash*. When data need to be added to the database, mining computers (i.e. computers connected to the blockchain network to process transactions) validate the transaction, encapsulate data records into a block, and broadcast the block to the entire network, so that all computers have an up-to-date copy of the database.

Blockchain is best known for its use in digital currencies such as Bitcoin, the first and largest decentralized cryptocurrency [Swan 2015]. A cryptocurrency is a digital currency in which encryption techniques are used to regulate the generation of units of currency and verify the transfer of funds, operating independently from a central bank. Payments using the decentralized virtual currency are recorded in a distributed public *ledger*, a permanent summary of all amounts and transactions (deposits, transfers, withdrawals) within a financial institution. A distributed ledger is a consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites, countries or institutions, thus eliminating the need of a central authority (such as a central bank).

In recent years, the blockchain technology started to be viewed as a more general application beyond digital currencies, towards working as a distributed ledger to track and record exchanges of any form of value [Colchester 2018]. Aside its increasing popularity, blockchain still presents some technical challenges, such as high throughput, latency, size, and energy consumption to maintain the network and process transactions [Swan 2015]. Aiming at overcoming some of these challenges while fostering its general use in different contexts, new technologies have been proposed.

To the best of our knowledge, the literature does not present much research on these new technologies, rather basically presenting the original whitepapers produced by the technology creators [Nakamoto 2008, Grigg 2017, Schiener 2017, Baird et al. 2018, NEO 2018, Popov 2018]. Moreover, there is a lack of a comprehensive overview or comparative analysis on the adoption of new technologies addressing the current challenges of blockchain. It is important that more people be aware that the technology is evolving, limitations are being addressed, and new ways to perform and improve certain process are being proposed, thus contributing to the evolution and implementation of the blockchain technology in other contexts.

Aiming at filling such a gap, this paper presents an overview on the blockchain technology and its technical challenges (Section 2), as well as the new technologies proposed to overcome some of them (Section 3). Finally, we briefly discuss how such blockchain-based technologies could be possibly applied to the Internet of Things (IoT) paradigm, which has some open issues related to data security and privacy (Section 4).

2. Blockchain

One of the main features of the blockchain technology is the elimination of a unique, centralized entity to validate transactions (e.g. adding, editing or removing data from a database), rather depending on a distributed consensus algorithm. For instance, most of the participants in the network need to agree on the validation of an insertion of data into the blockchain database, so that the operation is denied without consensus. What exactly is understood by "valid" is defined by the blockchain system and may vary from one to another, the so-called *protocol*. The integrity of the system is defined by the network, i.e. there is high trustworthiness since all participants in the network must reach a consensus to accept transactions. Once validated, a block is added to the blockchain database as a permanent record. The blockchain is also designed to be immutable: once in the chain, a block cannot be deleted or have its data altered without interfering in other blocks of the chain.

In terms of structure, the blockchain consists of a series of infinite growing blocks, which contain stored data linked to each other. Entries in the database can be made by all computers in the network since each one contains a copy of the database. Whenever data need to be inserted in the database, a new block is generated and given a hash value representing a unique identifier of the data within that block. The generation of the hash consists of encrypting and compressing the block data into a fixed-size string of characters, disregarding the amount of data inside the block [Swan 2015]. Once the user processes the block, the hash value can be recalculated to confirm that data inside the block were not altered. However, the reverse process of generating the block's data from the hash is impossible to perform.

All blocks in the blockchain are formed after the very first block (a.k.a. the *genesis block*), each one containing the hash of the previous block (see Figure 1). As the hash value represents data in a block, an alteration will trigger two possible scenarios: (i) new hash values will be generated for the blocks added after the changed block to fit the new hash, thus increasing the processing cost; or (ii) blocks added after the change will be considered invalid and no longer be part of the database since the hash used to generate its own hash value no longer exists, leading to their removal from the database. For this

reason, data inside the blockchain are hard to change and hence it creates a secure, trusted shared database [Colchester 2018].

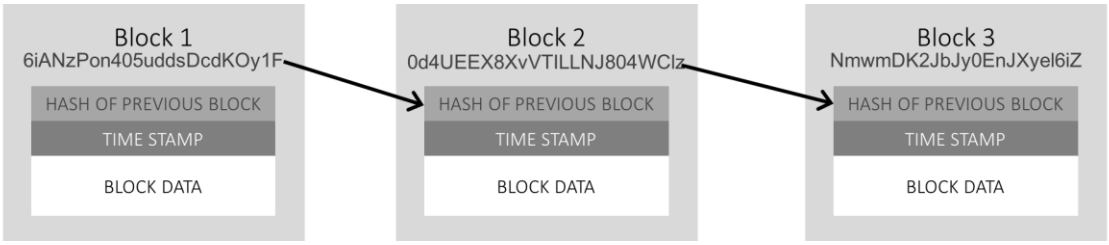

Figure 1. Block structure and connection in a blockchain

To randomize block processing across the mining computers as well as to avoid service abuses, the blockchain technology uses the *proof-of-work* scheme based on *hash-cash*. Hashcash [Back 2002] was originally proposed as a mechanism to throttle systematic abuse of unmetered Internet resources such as e-mail and anonymous remailers. It works by creating a proof-of-work (a cost function that generates a token) to be computed by users to establish a connection with a server. The server will check the value of the token using an evaluation function and it proceeds with the connection only if the token has the required value. As blocks are chained in the database, changing data within a block requires redoing the proofs-of-work in all succeeding blocks [Nakamoto 2008]. Nevertheless, if a node can modify blocks in the chain or it starts accepting invalid transactions, its database would not match the others and it would not be accepted as valid record thanks to the distributed consensus mechanism.

Blockchain was created as the main underlying infrastructure to run the Bitcoin cryptocurrency, which has played an important role as a relatively large-scale proof-of-concept capable of showing limitations of the blockchain technology. A key challenge with the blockchain technology used by Bitcoin refers to scalability in terms of the amount of transactions processed per second, also known as throughput. With cryptocurrency, speed is measured by transactions per second (tps). The current Bitcoin's blockchain technology processes only 7 tps whereas the VISA's credit card processing network handles 2,000 tps and can accommodate peak volumes of 10,000 tps [Swan 2015].

Another significant challenge for blockchain is related to latency. In Bitcoin, each transaction block takes ten minutes to be processed. The scenario is even worse when transferring large amounts of Bitcoins as it must outweigh the cost of a double-spend attack, in which Bitcoins are double-spent in a separate transaction before the confirmation of their reception by the tradesperson [Swan 2015]. In turn, the VISA's system takes around seconds to confirm transactions.

There are also some potential security issues with the Bitcoin's blockchain. Consensus is achieved when 51% of the network agree that the transactions encapsulated in a block are valid. The worst scenario is the possibility of a 51-percent attack, in which one mining entity could grab control of the blockchain and double spend previously transacted coins into his own account [Prashar 2013].

Finally, to support Bitcoin processing, the mining process consumes over US$ 1.5 billion per year in electricity [Colchester 2018]. The worst part of this scenario is that

most of the processing done by miners comes from the competition among them to perform the proof-of-work and be rewarded for processing the transactions in a block. Consequently, resources are wasted on something other than processing transactions.

3. New Technologies for Blockchain

Blockchain is in constant evolution and it has some limitations and challenges to overcome, thereby giving room for the creation of new technologies aimed to address such limitations, present new ways to perform some features or even better suit other contexts. Table 1 presents a comparison among some of these new technologies, which are being used by cryptocurrencies, regarding six criteria: (i) number of transactions performed per second (throughput); (ii) latency, i.e. the average time for confirming a transaction; (iii) the model for structuring data; (iv) the algorithm for achieving consensus among the nodes in the network; (v) the process for replicating data among computer nodes; and (vi) the method for selecting the node to process transactions. It is worth mentioning that the literature does not provide a comparative view on the blockchain technical challenges and the new emerging technologies attempting to overcome them as we try to present in this section.

Table 1. Comparison among new technologies for blockchain-based cryptocurrency

	Throughput (tps)	**Latency (s)**	**Data model**	**Consensus procedure**	**Replication process**	**Node selection**
Bitcoin [Swan 2015]	7	600	Array	Proof of work	Simple	Proof of work
Ethereum [Nakamoto 2008]	15	20	Array	Proof of work	Simple	Ghost Protocol
IOTA [Schiener 2018]	800	10	DAG	Tip algorithm	Assync	Random
Hashgraph [Baird et al. 2018]	500,000	11	DAG	Adapted Byzantine Fault Tolerance	Gossip	Random
NEO [NEO 2018]	10,000	20	Array	Delegated Byzantine Fault Tolerance	Simple	Voters
Lightining [Pool 2016]	oo	-	Array	-	-	-
EOS [Grigg 2017]	1,000,000	0.5	Array	Proof of stake	Simple	21 rounds

Latency. One of the main challenges of blockchain technologies refers to latency, i.e. the average time for to confirm a single transaction and achieving consensus in the network. In Bitcoin, this time is up to 600 seconds (see Table 1). As an attempt to reduce latency, the Lightining technology [Pool 2016] completely removes the process to confirm blocks. In a nutshell, Lightning uses micropayment channels, which create a relationship between two parties outside the global blockchain. These parties send an initial amount of Bitcoin into a multi-signature transaction with a local consensus on the current balance allocated between them. This creates a channel represented as an entry in the Bitcoin public ledger aiming at ensuring that the parties do not spend more funds than they own and hence blocking the funds on the global blockchain. Updates on the current balance of participants can be made only with the cooperation of both parties, enabling

them to create numerous transactions among them while not broadcasting to the global chain until either party wants to redeem their funds. This allows users to perform unlimited transactions between them without overloading the Bitcoin network, but still ensuring that currency was not double spent. Moreover, it increases the number of transactions processed by the network and reduces the time to confirm them since payments do not need block confirmations and network validations.

Data model. In the simplest sense, blockchains are a sequential chain (array) of blocks. Other technologies are proposing different ways to structure the data in the network. For instance, Tangle [Popov 2018] and Hashgraph [Baird et al. 2018] use a directed acyclic graph (DAG) for storing transactions. There are no blocks and transactions issued by the nodes constitute the site set of the tangle graph, which is the ledger for storing transactions. In Tangle, the technology used by the IOTA cryptocurrency [Popov 2018], when a new transaction arrives, it must approve two or more previous transactions, thereby creating a direct relationship between only two transactions, not to the whole array of data (as in traditional blockchain). These approvals are represented by directed edges as shown in Figure 2. If there is no directed edge between transactions A and B, but there is a directed path of length at least two from A to B, it is possible to state that A indirectly approves B and hence the transaction is valid in the tangle network. Even if a transaction in the graph is not direct relate to another, it may still indirect be related to other transactions in the network.

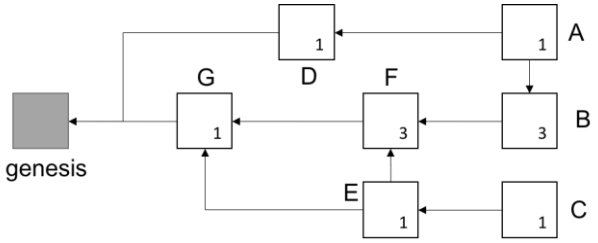

Figure 2. Tangle structure

Replication process. Blockchain uses a simple replication strategy to propagate blocks through the network, that is, whenever a new block is confirmed, it simply gets replicated to the entire network. The blocks are sequentially chained and each new block has a chain of its predecessor (parent) block, thus ensuring that a change in the parent block would invalidate the entire sub-sequential chain. Therefore, only the direct parents of a canonical child block are considered part of the blockchain. However, if miner A mines a block and a miner B mines another block before the A's block be propagated to B, the B's block will be wasted, also known as an uncle block. An uncle block is not part of the blockchain, but it still must process transactions [Swan 2015]. To solve this problem, Ethereum [Nakamoto 2018] introduced the modified GHOST protocol, in which not just the parent and further ancestors of a block participate in the calculation of the proof-of-work, but also the stale descendants of the block's ancestor (uncles). This allows uncles to be referenced/reincluded in the chain, so that each individual block has proportionally less work and hence block processing is accelerated. Furthermore, Ethereum extends the capacity of the technology towards a general platform for running decentralized applications and smart contracts since it provides a decentralized Turing-complete virtual machine, which can execute computer programs using a global network of nodes.

Other technologies present different approaches to propagate data in the network. For instance, the Tangle network is asynchronous, i.e. nodes do not necessarily see the same set of transactions. Hashgraph uses a gossip protocol: a member (e.g. Alice) will choose another member at random (e.g. Bob) and Alice will tell Bob all information that she knows so far. Alice repeats this process with a different random member. Bob repeatedly does the same and all other members as well. Consequently, if a single member becomes aware of new information and has a copy of the hashgraph network, then it will spread exponentially fast through the community until every member is aware of it. The history of gossip is represented through a directed graph, thus creating the data structure of the system [Baird 2018].

Consensus procedure. In Blockchain, consensus is achieved using proof-of-work. Whenever new transactions are broadcasted to the network, miners must find a proof-of-work (a challenge) for that set of transactions to generate a new block and receive their rewards. As miners must solve challenges to mine blocks, this approach can ensure that the entire network will have sufficiently long delays between mining events. It also means that it is necessary to slow down how fast blocks are mined, so that the new blocks can propagate to the entire community and agree if the transactions in a block are valid. On the other hand, EOS uses the Delegated Proof of Stake (DPOS) algorithm [Grigg 2017]. Under this algorithm, users (e.g. companies, bank users, etc.) holding tokens in the blockchain may select block producers through a continuous approval voting system. Anyone may choose to participate in block production and will be given an opportunity to produce blocks if it persuades token holders to vote for it.

NEO uses the Delegated Byzantine Fault Tolerant (dBFT) consensus mechanism, which allows for large-scale participation in consensus through proxy voting while reducing the time to process and validate blocks. The holder of the NEO token (analogue to Bitcoin) can pick the bookkeeper (a node in the network) by voting. Most Byzantine fault tolerance protocols without a leader depend on members sending each other votes [Baird 2018]. On the other hand, a hashgraph-based consensus does not require any votes to be sent as every member has a copy of the hashgraph. If Alice and Bob both have the same hashgraph, then they can calculate a total order on the events according to any deterministic function of that hashgraph and they will both get the same answer. Therefore, consensus is achieved even without sending vote messages [Baird 2018].

Management of conflicting transactions. Tangle uses a different approach that enables the network to contain conflicting transactions. The nodes do not have to achieve consensus on which valid transactions have the right to be in the ledger. However, if there are conflicting transactions, then the nodes need to decide which transactions will become orphaned, i.e. not directly or indirectly approved by incoming transactions. The main rule used to decide between two conflicting transactions is running the tip selection algorithm several times and check which transaction is more likely to be indirectly approved. A tip is an unapproved transaction in the tangle graph, so that a node chooses two or more tips from the tangle to issue a new transaction. A tip can be validated several times by different nodes. For example, if a transaction was selected 97 times out of 100 runs of the tip selection algorithm, then the algorithm will state that the tip is confirmed with confidence of 97%. Therefore, consensus is parallelized in the tangle network and it is done in sequential intervals of batches, so that the network can grow and scale dynamically with the number of transactions [Schiener 2017].

Node selection. Besides latency, throughput, and other concerns, the new Blockchain-based technologies present different ways to choose the node to process transactions. Bitcoin uses proof-of-work: the node that firstly finds a suitable proof-of-work for a set of transactions gets the chance to generate a new block and be rewarded for it. EOS presents an approach in which 21 block producers are selected by token holders in exactly 0.5 seconds (one round). Each producer gets one block per round and it is rewarded for processing blocks. The selected producers are scheduled in an order agreed by at least 15 producers. A block released by one producer is validated by the next producer and so forth; if not validated, then it is not part of the network. A block accepted by a quorum of producers is declared immutable. In Tangle, users must work to approve other transactions randomly assigned to them, so that users issuing a transaction are contributing to the network's security. A given node can both issue and approve transactions, thus eliminating the concept of mining and fees paid to them to process transactions.

4. Conclusion

Blockchain is a new technology that combines encryption and distribution. It is changing the way that we think not only about digital payment, but how computational devices and software communicate, interact, and collaborate with each other. However, it presents technical challenges to overcome towards its implementation in a global scale and in different contexts, such as the Internet of Things (IoT). IoT is an emergent computing paradigm that envisions the active collaboration of smart objects (things) with other physical and virtual resources available in the Internet while providing value-added information and functionalities for end-users and/or applications [Atzori et al. 2010]. Data obtained from IoT devices can contain sensitive information that others may be interested in or transmitted through networks without proper security. Moreover, many devices able to share information and to be controlled via Internet may become vulnerable to several types of attacks. Hackers or malicious users may try to remotely control devices, acquire confidential information or promote changes in the contents of messages while they are transmitted. As IoT devices often have limited processing capabilities, they usually do not have very complex security mechanisms such as data encryption, authentication, etc. Due to such a criticality, it is necessary to investigate mechanisms to protect critical and/or sensitive data from IoT devices and ensure security [Cavalcante et al. 2016].

Several works in the literature already report applications of the blockchain technology in IoT aiming at addressing concerns such as security, data storage, etc. [Conoscenti et al. 2016, Samaniego and Deters 2016, Kshetri 2017]. However, there is still no work detailing the new technologies emerged from Blockchain neither a comparison among them. To fill this gap, this paper presented an overview on these proposals attempting to overcome the Blockchain challenges and as well as representing improvements in both quantitative and qualitative features. With such an overview, it is possible to have a first guidance on which features and technologies would be suitable for securely storing and distributing data using blockchain in scenarios such as IoT.

References

Atzori, L., Iera, A., and Morabito, G. (2010) "The Internet of Things: A survey", Computer Networks 54(15), p. 2787-2805.

Back, A. (2002) Hashcash - A Denial of Service counter-measure. Available at http://www.hashcash.org/papers/hashcash.pdf.

Baird, L. (2018) The Swirlds hashgraph consensus algorithm: Fair, fast, Byzantine fault tolerance. Available at https://www.swirlds.com/downloads/SWIRLDS-TR-2016-01.pdf.

Baird, L., Harmon, M. and Madsen, P. (2018) Hedera: A Governing Council & Public Hashgraph Network. Available at https://s3.amazonaws.com/hedera-hashgraph/hh-whitepaper-v1.1-180518.pdf.

Banafa, A. (2016) A secure model of IoT with Blockchain. Available at https://www.bbvaopenmind.com/en/a-secure-model-of-iot-with-blockchain/.

Cavalcante, E. et al. (2016) "On the interplay of Internet of Things and Cloud Computing: A systematic mapping study", Computer Communications 89-90, p. 17-33.

Colchester, J. (2018) Blockchain: An overview. Available at http://complexitylabs.io/blockchain-overview/

Conoscenti, M., Vetrò, A., Martin, J. C. (2016) "Blockchain for the Internet of Things: A systematic literature review". In: Proceedings of the 13th IEEE/ACS International Conference of Computer Systems and Applications. USA: IEEE, p. 1-6.

Grigg, I. (2017) EOS - An introduction. Available at https://eos.io/documents/EOS_An_Introduction.pdf.

Kshetri, N. (2017) Can Blockchain strengthen the Internet of Things?, IT Professional 19(4), p. 68-72.

Nakamoto, S. (2008) Bitcoin: a peer-to-peer electronic cash system. Available at http://bitcoin.org/bitcoin.pdf.

NEO (2018) NEO White Paper. Available at http://docs.neo.org/en-us/whitepaper.html.

Prashar, V. (2013) What is Bitcoin 51% Attack, should I be worried?. Available at http://www.btcpedia.com/bitcoin-51-attack/.

Pool, K. and Dryja, T. (2016) The Bitcoin Lightning Network: Scalable off-chain instant payments. Available at https://lightning.network/lightning-network-paper.pdf.

Popov, S. (2018) The Tangle. Available at http://untangled.world/iota-whitepaper-tangle/.

Samaniego, M and Deters, R. (2016) "Blockchain as a Service for IoT" In: Proceedings of the 2016 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data. USA: IEEE, p. 433-436.

Schiener, D. (2017) A Primer on IOTA. Available at https://blog.iota.org/a-primer-on-iota-with-presentation-e0a6eb2cc621.

Swan, M. (2015). Blockchain: Blueprint for a new economy. USA: O'Reilly Media.