Blockchain : A Review

a. Retik Singh, Computer Department, Dhole Patil College of Engineering
b. Vaishnavi Todkar, Computer Department, Dhole Patil College of Engineering
c. Rutuja Varale, Computer Department, Dhole Patil College of Engineering
d. Pratyush Jadoun, Computer Department, Dhole Patil College of Engineering
e. Shailendra Pratap Singh, Dhole Patil College of Engineering

Abstract

With blockchain increasing in popularity and the motto of verify don't trust gaining traction. A detailed introduction to the subject and its many aspects is needed. This paper provides a mathematical and conceptual understanding for anyone getting ready to develop a blockchain application or a new blockchain in itself.

1. Introduction

Blockchain is a technology that has made a paradigm shift possible in the history of the world as we know it. Starting with the 2008 white paper by the pseudonym Satoshi Nakamoto[1]. Money is a concept that has stayed dynamic for ages and remains in contention even today.

During the early part of history, **the barter system** was the way of exchanging goods, however, wanting lettuce in return for a random good which can range from a grain of salt to even a leather shoe. The main problem with this approach was the overwhelming reliance on a coincidence of wants. This simply means person A wants what you have (Want 1) and person B wants what person A has (Want 2). Only, when these conditions are completed on both sides is a good sold in the barter system. Therefore, the creation of a neutral medium had to take place which was initially objects such as seashells then later on coins, and finally evolving to plastic and paper money.

Paper and plastic money was mostly run through the **gold standard** wherein the price of the money was backed by the amount of gold in a country's reserves. This standard was then changed to the Fiat System wherein money was not backed by any commodity but the decree of the government. The trust in government and organisations who decided what the value of their currency would be. Therefore, the money was centralised by a Central Bank such as the RBI in India, Federal Reserve in the USA, BoJ in Japan, etc.

In 2008 when the housing market crisis hit the US nationwide large investment banks such as Lehman Brothers and Bear Sterns collapsed causing large-scale panic, unemployment and apparent distrust in the central bank therefore the currency at least for the time being. This led to the publishing of a white paper by Satoshi Nakamoto for a technology called Bitcoin which was in turn built over another technology called Blockchain.

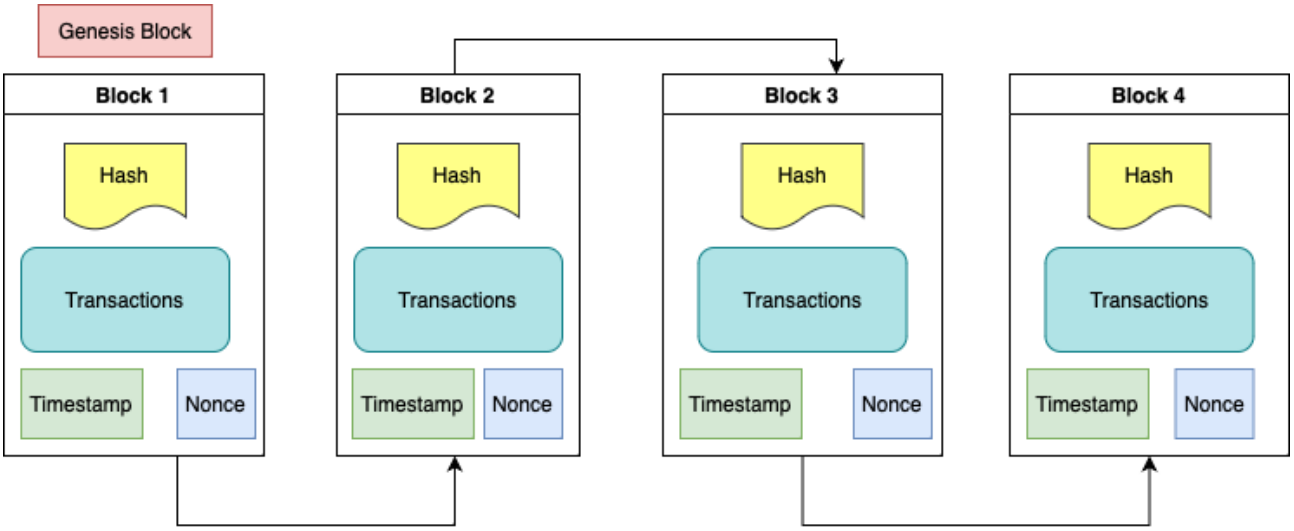
1. 1 Overview of Blockchain

The blockchain allowed the creation of an anonymous, decentralised, persistent authentication of data. From a structural standpoint, a blockchain is a chain of blocks with each block containing a list of transactions that have taken place. When the transactions in a block exceed a certain number a new block is created. Each block has a new hash and each user has two keys a private and a public key. This is also called asymmetric key cryptography. One of the most popular asymmetric key cryptography techniques is the ECC or Elliptic Curve Cryptography. Although the details of every blockchain are different and most after Bitcoin have some additional features to make them better the basic idea and logic generally remain constant[2].

1.2.1 General Blockchain Attributes

Blockchain at its core is a ledger that is public, decentralized and anonymous with a high degree of security with more reliance on proof rather than a trust which is a crucial change from the existing and the more mainstream Fiat Money which has been discussed above.

So, a blockchain is quite literally a chain of elements called blocks these blocks by themselves are ledgers that consist of hundreds and hundreds of lines of records, records of transactions made to people by people in a certain value or good. This can be monetary transactions such as in the case of Bitcoin or transferring goods such as NFTs (Non-Fungible Tokens).



A block by itself consists of a few data points.

- Block Version - Consists of a set of rules mainly used for validation
- Block Hash - A constant size (depending on the algorithm) that points to the previous block
- Timestamp - Time in seconds since, 1st Jan 1970 00:00 UTC or Zulu Time.
- Nonce - A 4-byte field working as an incremental index present in each block

A block can have two distinct parts a **block head** and **a block body** which are divided as metadata about the block is in the head and the transaction data is in the body. This is somewhat similar to how HTML webpages are structured where the head tag has all the metadata and the body tag has the visible data. On a broad scale, all blockchain architectures while taking their distinct features into consideration have three types

- **Open Access Blockchain**
- **Permissioned Blockchain**
- **Consortium Blockchain**

Here is a clear distinction between the three blockchains types

Factor	Open Access	Permissioned	Consortium
Decentralisation	Completely decentralized and open to all. Each node usually having the same rights	Managed by a central authority giving it partial decentralisation. Each node might not have equal rights to perform certain actions	Here decentralisation is more than Permissioned Blockchains but less than Open Access blockchains. It is controlled by multiple entities.
Inmutability	Nearly impossible to tamper with data	Tampering is possible.	Tampering is possible

Factor	Open Access	Permissioned	Consortium
Efficiency	As number of users grow the efficiency reduces	Stays Efficient	Stays Efficient
Record Access	Public	Maybe Restricted	Maybe Restricted

There are a few characteristics that a blockchain mostly relies on and the methodology applied generally has some similarities

- *Consensus Mechanism* - Consensus mechanisms in a blockchain are utilised to create agreements between all the users of the blockchain to authenticate a new record or transaction in the block. Different blockchains use different mechanisms. Since trust is an issue the technology set out to remove the used method is proof. This can be Proof of Work (used in Blockchain) or Proof of Stake (used in the Ethereum blockchain since 2022). Currently, the best-used mechanism is arguably Proof of Elapsed Time (PoET).

- *Decentralisation* - In a conventional system of scale the governing body is limited to a small number of people deciding rules, regulations and regulating the system. Blockchain on the other hand disregards this paradigm by utilising a system wherein every user part of the system holds a copy of the blockchain and therefore, they all can take part in regulating and maintaining the system.

- *Anonymity* - A common phrase in the cybersecurity world is "Either everyone is secure or no one is". The virtue of anonymity provides two primary benefits namely, The identity of the user is secure and secondly, the data of every user is safe from the basic to most sophisticated infiltration attacks such as rainbow table attack, SQL injection and good old brute force.

    - *Record Audits* - Building on the preceding points since records can't be altered due to the immense security and the decentralised nature of blockchain. The records are reliable for reference and fraud detection. And here's why this would work,

Even though a group of users may change their copy of the blockchain to reflect fraudulent data the data wouldn't be considered legit unless and until they have control over the network meaning 51% of all nodes have accepted that data as true. Thus, on a large and well-established blockchain records can be audited with almost no worry of record tampering, especially in the case of Open Access or Public Blockchains.

- *Persistence* - Transactions in a blockchain can only happen between two users that are part of the network already and have the amount or good they are trying to send. Likewise, the receiver should have the accepted monetary value of the good or equivalent amount. The timestamp data point (also called timestamp server) further strengthens this part of a blockchain by providing the **most recent value** that is to be transacted for a successful exchange. Keeping this in mind, and building on the above points the data is nearly impossible to falsify and manipulate in any way, This keeps the data reliability strong thus, the persistence is high as it should be. Thus, read permission in a blockchain allows greater audit capability and Immutability.

The rest of the paper is divided into further sections as such. Section 2 focuses on consensus mechanisms. Section 3 focuses on cryptographic hash functions. Section 4 focuses on applications of blockchain presently found. Section - 5 speculates on the future of blockchain and some developments that might take place or are needed in the field. Section 6 concludes this paper.

2. Consensus Mechanisms

As stated earlier a consensus mechanism is a replacement for trust in the Fiat System. Over time several different proof methods have been developed with various different concepts and utilities in mind. HashCash[3] used in Bitcoin was part of the first generation. Since then many more paradigms have come into existence as verification systems which remove the trust aspect in conventional systems.

2.1 Introduction

There are various such systems however, here are a few that seem to have the most innovative and effective changes to the technology of blockchain. However, applying broad strokes the most important problems that all consensus mechanisms have to solve is "The Double Spending Attack" and "The Byzantine Generals Problem".

There are several flavours of consensus mechanisms such as:

1. PoW(Proof of Work)

Proof of Work (from now on called PoW) is the consensus mechanism used in Bitcoin and works on the principle of

"A problem that is hard to solve but easy to verify"

A simple example which wouldn't be used due to its predictable nature is to find the ten thousandth prime number if the first one is zero. Now, as far as it's known there is no series or a general equation to find the nth-placed prime number. Therefore, it is hard to solve however the answer can be verified easily due to the answer's static nature. Typically, blockchain uses much more sophisticated methods to find a problem which can take 10 mins to solve. Moreover, the first one to solve that problem is rewarded with a fixed value (which is 6.25 BTC in Bitcoin) along with a small transaction fee.

2. PoS(Proof of Stake)

Proof of Stake[4] is an improved system that Ethereum shifted to by ditching its earlier adopted PoW system (discussed above). The system by design works on the following technique. A user has to **stake a minimum amount** (for Ethereum it is 32 ETH) after which they become a "validator". There is no certain upper limit. Thus, all these validators are ready to create a new block after the validation of all non-validated transactions in the previous blocks is done by a specific number of validators. This validator is picked for creating a new block on the blockchain network.

Coming to the stake taken by the network is taken as collateral or security deposit and can be deducted in case of any mishap or any deviation from the accepted protocol by the validator. This method was primarily picked for its effective reduction in energy efficiency over the existing PoW system.

3. DPoS(Delegated Proof of Stake)

DPoS[5] was first conceived by Dan Larimer in his treatise published in 2013, consequently, it was also used in his own cryptocurrency BitShares. The DPoS consensus mechanism is an evolution of the above-discussed PoS system. Herein, the users vote for delegates rather than all voting directly. The delegates then act as validators. Voting can change and others can be roped in as delegates. Delegates when validating a block the transaction fees are distributed among their votes depending on how much they staked.

This system is very alike to how the federalist democracies work in the real world. However, strict transparency allows strict checking of the possible misuse of power. DPoS is used by BitShares, TRON, EOS, etc.

4. PBFT(Practical Byzantine Fault Tolerance)

Earlier a reference to the Byzantine Generals' problem (BG problem) was made. This is a specific problem a consensus mechanism has to solve to allow a blockchain network to thrive without major hiccups or repercussions. The PBFT[6] mechanism works by creating a practical solution for the BG problem on the precondition that it doesn't have more than 1/3rd of nodes compromised. The operation in total has 4 steps

- The client sends a request to the leader node.
- The leader node then proceeds to broadcast the request to all the general nodes.
- The nodes regardless of being leader or general perform the service requested and then send back a response to the client.
- The request is served successfully if and only if the client receives n(leader) + n(general) replies from the different nodes present in the network with a consistent result.

5. PoB(Proof of Burn)

Proof of Burn[7] is another alternative consensus mechanism that solves the problem of excessive power consumption. Herein, the process is to burn coins in exchange for getting a virtual rig which would be used to mine these coins. More, the burn value more chances are for the miners to be selected to mine the next block. This method also avoids the possible misuse by early adopters by changing the burn value after regular intervals and adding a time limit for the virtual rig expiry. Simcoins uses the PoB mechanism for its verification and consensus.

6. PoC(Proof of Capacity)

Proof of Capacity[8] uses storage space as its collateral value. Asking network users to dedicate storage space with the condition of more space dedicated to the chances of creating the next block. BurstCoin, SpaceMint and Chia are some examples of projects using PoC as a consensus and validation mechanism.
The Mechanism works by plotting the Hard disk by repeated hashing of data by using nonce values as indexes. The indexed spaces are clubbed into pairs of two to create a **scoop**. This scoop value is what is calculated by a miner then,

using the scoop value at a given location a deadline value is calculated. After this is done for all the scoops. The miner with the least deadline value creates the new block in return for getting the block reward.

7. PoET(Proof of Elapsed Time)

Where the above PoC system was a biased lottery system PoET[9] uses a fair lottery system methodology to give every node a chance to win the right to mine a new block. It does so by assigning a random time for all nodes to fall asleep. The node to wake up the fastest (or the one with the shortest amount of elapsed time) wins the lottery and mines the new block.

8. RAFT

RAFT[10] is a distinct in nature consensus algorithm with being inspired by the earlier Paxos algorithm. It is easier to understand and that is touted as a major feature in its literature. There are three types of nodes in the RAFT system. 1. Leader, 2. Follower, 3. Candidate.

The client interacts only with the node that has been elected as the leader. Leaders can almost be 1. In case the leader node goes down any follower node can contest an election (becoming the candidate). A predetermined interval of time is what is chosen as a term for a leader.

Another crucial concept in RAFT is the CAP Theorem. It is a list of three major features that a distributed database system can have. However, the conundrum is only 2 of the 3 can be possible in the system.

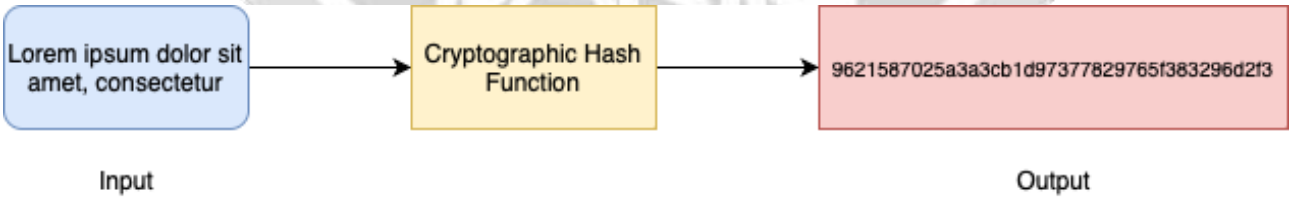
- **Consistency** – The data remains constant in every server nodes be it a leader or a follower thus, implying that the system has nearly instantaneous sync capabilities.
- **Availability** – Every request gets a response of either success or failure). It requires the system to be operational 100% of the time to serve requests.
- **Partition Tolerance** – The system remains responsive, even when a few of the server nodes do fail. Therefore, the system persists all requests/responses functions.

The leader node appends all the records to its ledger and after certain intervals of time, the other nodes update their records on a cyclic basis.

There also exist other less popular methods such as Proof of Activity, Proof of Weight, Proof of Importance, Leased Proof of Stake, etc which we aren't discussing in detail.

3. Cryptographic Hash Functions

A cryptographic hash function[11] is a one-way function that takes in a string of any length and returns a fixed-length string called a digest(called output in the figure).



3.1 Properties of Hash Functions

1. One-Way functions: Once, a block of text has been converted by a hash function it cannot be converted back to its original form. This provides safety from pre-image attacks.
2. Collision resistance: It's hard to find two strings such that hash(x1) = hash(x2). This is called strong collision resistance.
3. It's hard to predict what the original text would have been after the text has been hashed.

3.2 Standards of Hash Functions

SHA or (Secure Hash Algorithm) is a family of hash algorithms. Which currently include six distinct types namely, SHA-0, SHA-1, SHA-224, SHA-256, SHA-384, and SHA-512. These all are

1. SHA0

SHA0[12] was the first standard to be introduced in 1993 (as SHA) but was later withdrawn due to a flaw in the algorithm. It was later reintroduced and rechristened as SHA1. It had a 160-bit hash function

2. SHA1

   SHA1[13] also had a 160-bit hash function which resembled the earlier MD5 standard that was used as a benchmark when it was being developed. SHA1 was developed by the NSA but was declared as compromised in 2011. It was removed from most use cases by 2017. CWI Amsterdam and Google successfully proved their compromised status when they produced two different PDFs with the same hash.

3. SHA2

   SHA2[14] is actually a family of two different algorithms that together form the SHA2 group. These are SHA 256 and SHA 512. The major difference between them is the word size of the digest they produce. There were also some more versions which were also developed by the NSA. However, the next family SHA3 (earlier called Keccak) was developed by non-NSA designers.

4. SHA3

   Introduced in 2015 by NIST as a recognised standard and created as an alternative to SHA2 if any attacks were to be demonstrated. SHA3[15] is a function different from what SHA2 had. This is because it uses a sponge function instead of the usual style that SHA1 and SHA2 have.

Although the standard has not been compromised concerns about quantum attacks are a threat as shown by Grover's Algorithm[16].

Applications of Hashing Algorithms

- Database Indexing
- Password Storage
- Data Compression
- Search Algorithm
- Blockchain
- File Comparison

4. Applications

1. Finance

In 2015, major banks including but not limited to HSBC, Wells Fargo, BoA, and Allianz, joined hands to create a DLT-based permissioned private blockchain called R3 Corda[17].
Similarly, in 2015 again the technology of Hyperledger[18] revolutionised how we looked at Blockchain from a commercial and business-oriented perspective. According to World Bank[19]
countries like China and India already using weChat and UPI responsive and attain an advantage for mass adoption before most other nations across the globe.

The growth of Dapps, systems that accelerate the trust factors become imperative to how the world would be shaped and transformed via Prediction Markets, APIs that tackle the data sharing practices without compromising on data persistence speeds and data integrity thanks to the cryptographic hash function which has been discussed earlier.

2. Security

The biggest corollary of this section is the introduction of Smart Contracts first introduced in the Ethereum blockchain. An open-source project which allows a contract to be created, signed and sealed to be put into use. The contract can then only be changed if and only if 51% or greater users of the chain give a yes to a certain change.

3. Trust Metrics

Algorithmic Trust[21] is a paper that provides a theoretical basis for a framework to provide a basis of a score or numeric value for better trust in businesses thus in theory reducing fraud and cheating. A very similar approach can be seen in reputation-based systems. A very crude and approachable counterpart would be the Elo ranking system used in Chess.

Rep on the block[22] is a paper that presents a generalised reputation system that can be applied to multiple networks. One major improvement the system provides is the resistance to false rating attacks and Sybil attacks. This is done by making the potential hard work required to tamper with the rating way harder than the potential reward. Therefore, in theory killing possible incentives.

5. The Future

The future of blockchain has two separate one is issues we would want to resolve and the other is advances we would want to make.

5.1 Issues

1. Coding Error: It is possible for a user to destroy and forfeit their financial holding due to a glitch or bug on the wallet system or exchange. This can cause major distrust for the user in itself and widespread fear.
2. 51% Attacks: Even though, blockchains are impenetrable the systems their users handle aren't thus, it is possible for mid-sized or small blockchains to be maliciously changed or altered to benefit a group or single entity.
3. Attacks on Exchanges: Exchanges, where transactions take place, are a weakness in the overall architecture of a blockchain. Since they are a centralised environment. One such attack was on Bitfinex a crypto exchange where 120,00 bitcoin were stolen. This was possible due to a vulnerability in the multi-signature wallets. Multi-signature wallets are wallets where three signatures are needed. One public and two private.

5.2 Advances

1. Resisting Centralisation: Major chunks of blockchain are owned by a handful of people[23] this makes the spirit of achieving true decentralisation harder. This is one aspect where more work is needed to improve the current status quo. Newer consensus methods such as Proof of Capacity and Proof of Elapsed Time are techniques aimed at solving this problem as well as energy efficiency.
2. Government Regulation: Due to the anonymous nature of the blockchain technology currencies such as Bitcoin are being used by bad actors to conduct nefarious activities and have a strong bank balance while also evading the watchful eyes of the government. The former drug marketplace Silk Road used Bitcoin as a primary source of transactions. In these few instances, government oversight can help alleviate such issues and allow outlaws to be brought into the court of justice.

3. Mainstream applications: Currently, most applications of Blockchain aren't mainstream. However, this is changing with technologies such as IoT have opened avenues to enable more mainstream utilisation of the technology to happen. Thus realising the dream of Satoshi Nakamoto.

6. Conclusion

Above we have discussed the Consensus mechanism in detail with different techniques. Most 2nd generation methods have focused on energy efficiency which was an issue in the Proof of Work method used in Bitcoin that was first proposed by Satoshi Nakamoto. However, this is a long way back. Today, blockchain has become a part of our lives and lifestyle. With new technologies, products and ideations coming to fruition. Which seemed impossible just two decades ago. It has been a long and arduous journey from the genesis block mined to NFTs being sold for millions.

7. References

[1] Bitcoin: A peer to peer Electronic Cash System, Satoshi Nakamoto, https://bitcoin.org/bitcoin.pdf

[2] Nofer, M., Gomber, P., Hinz, O., & Schiereck, D. (2017). Blockchain. *Business & Information Systems Engineering*, *59*, 183-187.

[3] Back, A. (2002). Hashcash-a denial of service counter-measure.

[4] Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, *151*(2014), 1-32.

[5] https://steemit.com/dpos/@dantheman/dpos-consensus-algorithm-this-missing-white-paper

[6] Castro, M., & Liskov, B. (1999, February). Practical byzantine fault tolerance. In *OsDI* (Vol. 99, No. 1999, pp. 173-186).

[7] Karantias, K., Kiayias, A., & Zindros, D. (2020). Proof-of-burn. In Financial Cryptography and Data Security: 24th International Conference, FC 2020, Kota Kinabalu, Malaysia, February 10–14, 2020 Revised Selected Papers 24 (pp. 523-540). Springer International Publishing.

[8] https://blog.ethereum.org/2014/01/15/slasher-a-punitive-proof-of-stake-algorithm

[9] Buntinx, J. P. (2017). What is proof of elapsed time. The Merkle Hash. Available online: https://themerkle. com/what-is-proof-of-elapsed-time/(accessed on 5 December 2019).

[10] Ongaro, D., & Ousterhout, J. (2015). The raft consensus algorithm. *Lecture Notes CS*, *190*, 2022.

[11] Preneel, B. (1994). Cryptographic hash functions. *European Transactions on Telecommunications*, *5*(4), 431-448.

[12] Naito, Y., Sasaki, Y., Shimoyama, T., Yajima, J., Kunihiro, N., & Ohta, K. (2006). Improved collision search for SHA-0. In Advances in Cryptology–ASIACRYPT 2006: 12th International Conference on the Theory and Application of Cryptology and Information Security, Shanghai, China, December 3-7, 2006. Proceedings 12 (pp. 21-36). Springer Berlin Heidelberg.

[13] Wang, X., Yin, Y. L., & Yu, H. (2005). Finding collisions in the full SHA-1. In Advances in Cryptology–CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005. Proceedings 25 (pp. 17-36). Springer Berlin Heidelberg.

[14] N. Ferguson, et al., The Skein Hash Function Family, Version 1.3, 1 Oct 2010

[15] G. Bertoni, et al., The Keccak reference, Version 3.0, January 14, 2011

[16] Lavor, C., Manssur, L. R. U., & Portugal, R. (2003). Grover's algorithm: Quantum database search. *arXiv preprint quant-ph/0301079*.

[17] Brown, R. G., Carlyle, J., Grigg, I., & Hearn, M. (2016). Corda: an introduction. *R3 CEV, August*, *1*(15), 14.

[18] Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., ... & Yellick, J. (2018, April). Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the thirteenth EuroSys conference* (pp. 1-15).

[19]Niforos, Marina. 2017. Blockchain in Financial Services in Emerging Markets, Part II : Selected Regional Developments. EMCompass,no. 44;. © International Finance Corporation, Washington, DC. https://openknowledge.worldbank.org/entities/publication/28cef00e-9121-5291-b522-72ff8704cfb1 License: CC BY-NC-ND 3.0 IGO.

[20] Swan, M. (2015). Blockchain: Blueprint for a new economy. California, USA: O'Reilly Media, Inc.

[21] Swan, M., & Brunswicker, S. (2018). Blockchain economic networks and algorithmic trust.

[22] Dennis, R., & Owen, G. (2015, December). Rep on the block: A next generation reputation system based on the blockchain. In *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)* (pp. 131-138). IEEE.

[23] "Crypto-currency market capitalizations," 2017. [Online]. Available: https://coinmarketcap.com