An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends

Conference Paper · June 2017

DOI: 10.1109/BigDataCongress.2017.85

5 authors, including:

Zibin Zheng
Sun Yat-Sen University
135 PUBLICATIONS **3,036** CITATIONS
SEE PROFILE

Shaoan Xie
Sun Yat-Sen University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Hong-Ning Dai
Macau University of Science and Technology
64 PUBLICATIONS **426** CITATIONS
SEE PROFILE

An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends

Zibin Zheng[1], Shaoan Xie[1], Hongning Dai[2], Xiangping Chen[4], and Huaimin Wang[3]

[1]School of Data and Computer Science, Sun Yat-sen University Guangzhou, China
[2]Faculty of Information Technology, Macau University of Science and Technology, Macau, SAR
[3]National Laboratory for Parallel & Distributed Processing
National University of Defense Technology, Changsha 410073 China
[4]Institute of Advanced Technology,National Engineering Research Center of Digital Life
Sun Yat-sen University, Guangzhou, China
Email: zhzibin@mail.sysu.edu.cn

***Abstract*—Blockchain, the foundation of Bitcoin, has received extensive attentions recently. Blockchain serves as an immutable ledger which allows transactions take place in a decentralized manner. Blockchain-based applications are springing up, covering numerous fields including financial services, reputation system and Internet of Things (IoT), and so on. However, there are still many challenges of blockchain technology such as scalability and security problems waiting to be overcome. This paper presents a comprehensive overview on blockchain technology. We provide an overview of blockchain architechture firstly and compare some typical consensus algorithms used in different blockchains. Furthermore, technical challenges and recent advances are briefly listed. We also lay out possible future trends for blockchain.**

***Index Terms*—Blockchain, decentralization, consensus, scalability**

I. Introduction

Nowadays *cryptocurrency* has become a buzzword in both industry and academia. As one of the most successful cryptocurrency, Bitcoin has enjoyed a huge success with its capital market reaching 10 billion dollars in 2016 [1]. With a specially designed data storage structure, transactions in Bitcoin network could happen without any third party and the core technology to build Bitcoin is *blockchain*, which was first proposed in 2008 and implemented in 2009 [2]. Blockchain could be regarded as a public ledger and all committed transactions are stored in a list of blocks. This chain grows as new blocks are appended to it continuously. Asymmetric cryptography and distributed consensus algorithms have been implemented for user security and ledger consistency. The blockchain technology generally has key characteristics of decentralization, persistency, anonymity and auditability. With these traits, blockchain can greatly save the cost and improve the efficiency.

Since it allows payment to be finished without any bank or any intermediary, blockchain can be used in various financial services such as digital assets, remittance and online payment [3], [4]. Additionally, it can also be applied into other fields including smart contracts [5], public services [6], Internet of Things (IoT) [7], reputation systems [8] and security services [9]. Those fields favor blockchain in multiple ways. First of all, blockchain is immutable. Transaction cannot be tampered once it is packed into the blockchain. Businesses that require high reliability and honesty can use blockchain to attract customers. Besides, blockchain is distributed and can avoid the single point of failure situation. As for smart contracts, the contract could be executed by miners automatically once the contract has been deployed on the blockchain.

Although the blockchain technology has great potential for the construction of the future Internet systems, it is facing a number of technical challenges. Firstly, scalability is a huge concern. Bitcoin block size is limited to 1 MB now while a block is mined about every ten minutes. Subsequently, the Bitcoin network is restricted to a rate of 7 transactions per second, which is incapable of dealing with high frequency trading. However, larger blocks means larger storage space and slower propagation in the network. This will lead to centralization gradually as less users would like to maintain such a large blockchain. Therefore the tradeoff between block size and security has been a tough challenge. Secondly, it has been proved that miners could achieve larger revenue than their fair share through selfish mining strategy [10]. Miners hide their mined blocks for more revenue in the future. In that way, branches could take place frequently, which hinders blockchain development. Hence some solutions need to be put forward to fix this problem. Moreover, it has been shown that privacy leakage could also happen in blockchain even users only make transactions with their public key and private key [11]. Furthermore, current consensus algorithms like *proof of work* or *proof of stake* are facing some serious problems. For example, proof of work wastes too much electricity energy while the phenomenon that the rich get richer could appear in the proof of stake consensus process.

There is a lot of literature on blockchain from various sources, such as blogs, wikis, forum posts, codes, conference proceedings and journal articles. Tschorsch et al. [12] made a technical survey about decentralized digital currencies

978-1-5386-1996-4/17 $31.00 © 2017 IEEE
DOI 10.1109/BigDataCongress.2017.85

(ii) 默克尔树根哈希：区块中所有交易的哈希值。
(iii) 时间戳：自1970年1月1日以来的通用时间的秒数。
(iv) nBits：有效区块哈希的目标阈值。
(v) 随机数：一个4字节的字段，通常以0开头，并针对每次哈希计算递增（将在第三节中详细解释）

including Bitcoin. Compared to [12], our paper focuses on blockchain technology instead of digital currencies. Nomura Research Institut made a technical report about blockchain [13]. Contrast to [13], our paper focuses on state-of-art blockchain researches including recent advances and future trends.

The rest of this paper is organized as follows. Section II introduces blockchain architecture. Section III shows typical consensus algorithms used in blockchain. Section IV summarizes the technical challenges and the recent advances in this area. Section V discusses some possible future directions and section VI concludes the paper.

II. Blockchain Architecture

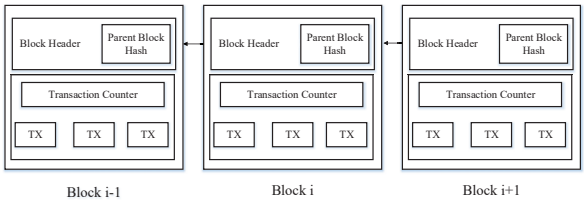

Fig. 1: An example of blockchain which consists of a continuous sequence of blocks.

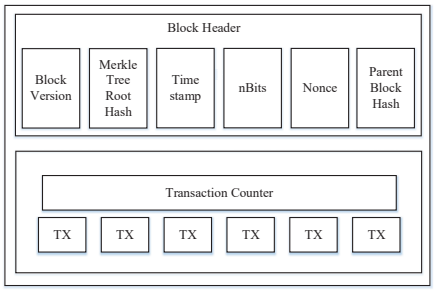

Fig. 2: Block structure

Blockchain is a sequence of blocks, which holds a complete list of transaction records like conventional public ledger [14]. Figure 1 illustrates an example of a blockchain. With a previous block hash contained in the block header, a block has only one *parent block*. It is worth noting that *uncle blocks* (children of the block's ancestors) hashes would also be stored in ethereum blockchain [15]. The first block of a blockchain is called *genesis block* which has no parent block. We then explain the internals of blockchain in details.

A. Block

A block consists of the *block header* and the *block body* as shown in Figure 2. In particular, the block header includes:

(i) Block version: indicates which set of block validation rules to follow.

(ii) Merkle tree root hash: the hash value of all the transactions in the block.

(iii) Timestamp: current time as seconds in universal time since January 1, 1970.

(iv) nBits: target threshold of a valid block hash.

(v) Nonce: an 4-byte field, which usually starts with 0 and increases for every hash calculation (will be explained in details in Section III).

(vi) Parent block hash: a 256-bit hash value that points to the previous block.

The block body is composed of a transaction counter and transactions. The maximum number of transactions that a block can contain depends on the block size and the size of each transaction. Blockchain uses an asymmetric cryptography mechanism to validate the authentication of transactions [13]. Digital signature based on asymmetric cryptography is used in an untrustworthy environment. We next briefly illustrate digital signature.

B. Digital Signature

Each user owns a pair of private key and public key. The private key that shall be kept in confidentiality is used to sign the transactions. The digital signed transactions are broadcasted throughout the whole network. The typical digital signature is involved with two phases: *signing phase* and *verification phase*. For instance, an user Alice wants to send another user Bob a message. (1) In the signing phase, Alice encrypts her data with her private key and sends Bob the encrypted result and original data. (2) In the verification phase, Bob validates the value with Alice's public key. In that way, Bob could easily check if the data has been tampered or not. The typical digital signature algorithm used in blockchains is the elliptic curve digital signature algorithm (ECDSA) [16].

C. Key Characteristics of Blockchain

In summary, blockchain has following key characteristics.

- *Decentralization*. In conventional centralized transaction systems, each transaction needs to be validated through the central trusted agency (e.g., the central bank), inevitably resulting to the cost and the performance bottlenecks at the central servers. Contrast to the centralized mode, third party is no longer needed in blockchain. Consensus algorithms in blockchain are used to maintain data consistency in distributed network.
- *Persistency*. Transactions can be validated quickly and invalid transactions would not be admitted by honest miners. It is nearly impossible to delete or rollback transactions once they are included in the blockchain. Blocks that contain invalid transactions could be discovered immediately.
- *Anonymity*. Each user can interact with the blockchain with a generated address, which does not reveal the real identity of the user. Note that blockchain cannot guarantee the perfect privacy preservation due to the intrinsic constraint (details will be discussed in section IV).

TABLE I: Comparisons among *public blockchain*, *consortium blockchain* and *private blockchain*

Property	Public blockchain	Consortium blockchain	Private blockchain
Consensus determination	All miners	Selected set of nodes	One organization
Read permission	Public	Could be public or restricted	Could be public or restricted
Immutability	Nearly impossible to tamper	Could be tampered	Could be tampered
Efficiency	Low	High	High
Centralized	No	Partial	Yes
Consensus process	Permissionless	Permissioned	Permissioned

- *Auditability*. Bitcoin blockchain stores data about user balances based on the Unspent Transaction Output (UTXO) model [2]: Any transaction has to refer to some previous unspent transactions. Once the current transaction is recorded into the blockchain, the state of those referred unspent transactions switch from unspent to spent. So transactions could be easily verified and tracked.

D. Taxonomy of blockchain systems

Current blockchain systems are categorized roughly into three types: public blockchain, private blockchain and consortium blockchain [17]. In public blockchain, all records are visible to the public and everyone could take part in the consensus process. Differently, only a group of pre-selected nodes would participate in the consensus process of a consortium blockchain. As for private blockchain, only those nodes that come from one specific organization would be allowed to join the consensus process.

A private blockchain is regarded as a centralized network since it is fully controlled by one organization. The consortium blockchain constructed by several organizations is partially decentralized since only a small portion of nodes would be selected to determine the consensus. The comparison among the three types of blockchains is listed in Table I.

- *Consensus determination*. In public blockchain, each node could take part in the consensus process. And only a selected set of nodes are responsible for validating the block in consortium blockchain. As for private chain, it is fully controlled by one organization and the organization could determine the final consensus.
- *Read permission*. Transactions in a public blockchain are visible to the public while it depends when it comes to a private blockchain or a consortium blockchain.
- *Immutability*. Since records are stored on a large number of participants, it is nearly impossible to tamper transactions in a public blockchain. Differently, transactions in a private blockchain or a consortium blockchain could be tampered easily as there are only limited number of participants.
- *Efficiency*. It takes plenty of time to propagate transactions and blocks as there are a large number of nodes on public blockchain network. As a result, transaction throughput is limited and the latency is high. With fewer validators, consortium blockchain and private blockchain could be more efficient.
- *Centralized*. The main difference among the three types of blockchains is that public blockchain is decentralized, consortium blockchain is partially centralized and private blockchain is fully centralized as it is controlled by a single group.
- *Consensus process*. Everyone in the world could join the consensus process of the public blockchain. Different from public blockchain, both consortium blockchain and private blockchain are permissioned.

Since public blockchain is open to the world, it can attract many users and communities are active. Many public blockchains emerge day by day. As for consortium blockchain, it could be applied into many business applications. Currently Hyperledger [18] is developing business consortium blockchain frameworks. Ethereum also has provided tools for building consortium blockchains [19].

III. Consensus Algorithms

In blockchain, how to reach consensus among the untrustworthy nodes is a transformation of the Byzantine Generals (BG) Problem, which was raised in [20]. In BG problem, a group of generals who command a portion of Byzantine army circle the city. Some generals prefer to attack while other generals prefer to retreat. However, the attack would fail if only part of the generals attack the city. Thus, they have to reach an agreement to attack or retreat. How to reach a consensus in distributed environment is a challenge. It is also a challenge for blockchain as the blockchain network is distributed. In blockchain, there is no central node that ensures ledgers on distributed nodes are all the same. Some protocols are needed to ensure ledgers in different nodes are consistent. We next present several common approaches to reach a consensus in blockchain.

A. Approaches to consensus

PoW (Proof of work) is a consensus strategy used in the Bitcoin network [2]. In a decentralized network, someone has to be selected to record the transactions. The easiest way is random selection. However, random selection is vulnerable to attacks. So if a node wants to publish a block of transactions, a lot of work has to be done to prove that the node is not likely to attack the network. Generally the work means computer

TABLE II: Typical Consensus Algorithms Comparison

Property	PoW	PoS	PBFT	DPOS	Ripple	Tendermint
Node identity management	open	open	permissioned	open	open	permissioned
Energy saving	no	partial	yes	partial	yes	yes
Tolerated power of adversary	< 25% computing power	< 51% stake	< 33.3% faulty replicas	< 51% validators	< 20% faulty nodes in UNL	< 33.3% byzantine voting power
Example	Bitcoin [2]	Peercoin [21]	Hyperledger Fabric [18]	Bitshares [22]	Ripple [23]	Tendermint [24]

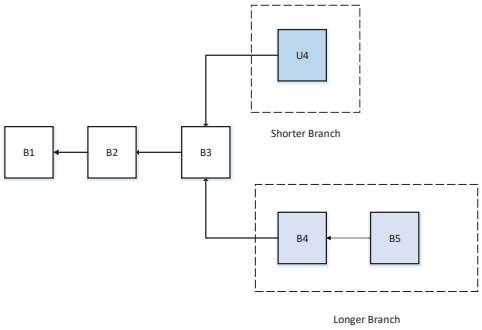

Fig. 3: An scenario of blockchain branches (the longer branch would be admitted as the main chain while the shorter one would be deserted)

矿工：计算哈希值的节点； 挖矿：PoW过程

工作量证明

calculations. In PoW, each node of the network is calculating a hash value of the block header. The block header contains a nonce and miners would change the nonce frequently to get different hash values. The consensus requires that the calculated value must be equal to or smaller than a certain given value. When one node reaches the target value, it would broadcast the block to other nodes and all other nodes must mutually confirm the correctness of the hash value. If the block is validated, other miners would append this new block to their own blockchains. Nodes that calculate the hash values are called *miners* and the PoW procedure is called *mining* in Bitcoin. PoW协议通过选择最长的链来解决区块链分叉问题

In the decentralized network, valid blocks might be generated simultaneously when multiple nodes find the suitable nonce nearly at the same time. As a result, branches may be generated as shown in Figure 3. However, it is unlikely that two competing forks will generate next block simultaneously. In PoW protocol, a chain that becomes longer thereafter is judged as the authentic one. Consider two forks created by simultaneously validated blocks U4 and B4. Miners keep mining their blocks until a longer branch is found. B4,B5 forms a longer chain, so the miners on U4 would switch to the longer branch.

Miners have to do a lot of computer calculations in PoW, yet these works waste too much resources. To mitigate the loss, some PoW protocols in which works could have some side-applications have been designed. For example, Primecoin [25] searches for special prime number chains which can be used for mathematical research.

权益证明

PoS (Proof of stake) is an energy-saving alternative to PoW. Miners in PoS have to prove the ownership of the amount of currency. It is believed that people with more currencies would be less likely to attack the network. The selection based on account balance is quite unfair because the single richest person is bound to be dominant in the network. As a result, many solutions are proposed with the combination of the stake size to decide which one to forge the next block. In particular, Blackcoin [26] uses randomization to predict the next generator. It uses a formula that looks for the lowest hash value in combination with the size of the stake. Peercoin [21] favors coin age based selection. In Peercoin, older and larger sets of coins have a greater probability of mining the next block. Compared to PoW, PoS saves more energy and is more effective. Unfortunately, as the mining cost is nearly zero, attacks might come as a consequence. Many blockchains adopt PoW at the beginning and transform to PoS gradually. For instance, ethereum is planing to move from Ethash (a kind of PoW) [27] to Casper (a kind of PoS) [28].

拜占庭容错

PBFT (Practical byzantine fault tolerance) is a replication algorithm to tolerate byzantine faults [29]. Hyperledger Fabric [18] utilizes the PBFT as its consensus algorithm since PBFT could handle up to 1/3 malicious byzantine replicas. A new block is determined in a round. In each round, a primary would be selected according to some rules. And it is responsible for ordering the transaction. The whole process could be divided into three phase: *pre-prepared*, *prepared* and *commit*. In each phase, a node would enter next phase if it has received votes from over 2/3 of all nodes. So PBFT requires that every node is known to the network. Like PBFT, Stellar Consensus Protocol (SCP) [30] is also a Byzantine agreement protocol. In PBFT, each node has to query other nodes while SCP gives participants the right to choose which set of other participants to believe. Based on PBFT, Antshares [31] has implemented their dBFT (delegated byzantine fault tolerance). In dBFT, some professional nodes are voted to record the transactions.

委任权益证明

DPOS (Delegated proof of stake). The major difference between PoS and DPOS is that PoS is direct democratic while DPOS is representative democratic. Stakeholders elect their delegates to generate and validate blocks. With significantly fewer nodes to validate the block, the block could be confirmed quickly, leading to the quick confirmation of transactions. Meanwhile, the parameters of the network such as *block size* and *block intervals* could be tuned by delegates. Additionally,

users need not to worry about the dishonest delegates as they could be voted out easily. DPOS is the backbone of Bitshares [22].

Ripple [23] is a consensus algorithm that utilizes collectively-trusted subnetworks within the larger network. In the network, nodes are divided into two types: *server* for participating consensus process and *client* for only transferring funds. Each server has an Unique Node List (UNL). UNL is important to the server. When determining whether to put a transaction into the ledger, the server would query the nodes in UNL and if the received agreements have reached 80%, the transaction would be packed into the ledger. For a node, the ledger will remain correct as long as the percentage of faulty nodes in UNL is less than 20%.

Tendermint [24] is a byzantine consensus algorithm. A new block is determined in a round. A proposer would be selected to broadcast an unconfirmed block in this round. It could be divided into three steps: 1) *Prevote step*. Validators choose whether to broadcast a prevote for the proposed block. 2) *Precommit step*. If the node has received more than 2/3 of prevotes on the proposed block, it broadcasts a precommit for that block. If the node has received over 2/3 of precommits, it enters the commit step. 3) *Commit step*. The node validates the block and broadcasts a commit for that block. if the node has received 2/3 of the commits, it accepts the block. Contrast to PBFT, nodes have to lock their coins to become validators. Once a validator is found to be dishonest, it would be punished.

B. Consensus algorithms comparison

Different consensus algorithms have different advantages and disadvantages. Table II gives a comparison between different consensus algorithms and we use the properties given by [32].

- *Node identity management*. PBFT needs to know the identity of each miner in order to select a primary in every round while Tendermint needs to know the validators in order to select a proposer in each round. For PoW, PoS, DPOS and Ripple, nodes could join the network freely.
- *Energy saving*. In PoW, miners hash the block header continuously to reach the target value. As a result, the amount of electricity required to process has reach an immense scale. As for PoS and DPOS, miners still have to hash the block header to search the target value but the work has been largely reduced as the search space is designed to be limited. As for PBFT, Ripple and Tendermint, there is no mining in consensus process. So it saves energy greatly.
- *Tolerated power of adversary*. Generally 51% of hash power is regarded as the threshold for one to gain control of the network. But selfish mining strategy [10] in PoW systems could help miners to gain more revenue by only 25% of the hashing power. PBFT and Tendermint is designed to handle up to 1/3 faulty nodes. Ripple is proved to maintain correctness if the faulty nodes in an UNL is less than 20%.
- *Example*. Bitcoin is based on PoW while Peercoin is a new peer-to-peer PoS cryptocurrency. Further, Hyperledger Fabric utilizes PBFT to reach consensus. Bitshares, a smart contract platform, adopts DPOS as their consensus algorithm. Ripple implements the Ripple protocol while Tendermint devises the Tendermint protocol.

PBFT and Tendermint are permissioned protocols. Node identities are expected to be known to the whole network, so they might be used in commercial mode rather than public. PoW and PoS are suitable for public blockchain. Consortium or private blockchain might has preference for PBFT, Tendermint, DPOS and Ripple.

C. Advances on consensus algorithms

A good consensus algorithm means efficiency, safty and convenience. Recently, a number of endeavors have been made to improve consensus algorithms in blockchain. New consensus algorithms are devised aiming to solve some specific problems of blockchain. The main idea of PeerCensus [33] is to decouple block creation and transaction confirmation so that the consensus speed can be significantly increased. Besides, Kraft [34] proposed a new consensus method to ensure that a block is generated in a relatively stable speed. It is known that high blocks generation rate compromise Bitcoin's security. So the Greedy Heaviest-Observed Sub-Tree (GHOST) chain selection rule [35] is proposed to solve this problem. Instead of the longest branch scheme, GHOST weights the branches and miners could choose the better one to follow. Chepurnoy et al. [36] presented a new consensus algorithm for peer-to-peer blockchain systems where anyone who provides non-interactive proofs of retrievability for the past state snapshots is agreed to generate the block. In such a protocol, miners only have to store old block headers instead of full blocks.

IV. Challenges & Recent Advances

Despite the great potential of blockchain, it faces numerous challenges, which limit the wide usage of blockchain. We enumerate some major challenges and recent advances as follows.

A. Scalability

With the amount of transactions increasing day by day, the blockchain becomes bulky. Each node has to store all transactions to validate them on the blockchain because they have to check if the source of the current transaction is unspent or not. Besides, due to the original restriction of block size and the time interval used to generate a new block, the Bitcoin blockchain can only process nearly 7 transactions per second, which cannot fulfill the requirement of processing millions of transactions in real-time fashion. Meanwhile, as the capacity of blocks is very small, many small transactions might be delayed since miners prefer those transactions with high transaction fee.

There are a number of efforts proposed to address the scalability problem of blockchain, which could be categorized into two types:

叛徒节点的容忍能力。一般来说，51%的哈希算力被认为是获得对网络控制的门槛。但PoW中的自私挖矿策略[可以帮助矿工只拥有25%的算力就获得更多收入。PBFT和Tendermint被设计成可以处理多达1/3的错误节点。Ripple已被证实在UNL中错误节点少于20%时能维持正确性。

- *Storage optimization of blockchain*. Since it is harder for node to operate full copy of ledger, Bruce proposed a novel cryptocurrency scheme, in which the old transaction records are removed (or forgotten) by the network [37]. A database named account tree is used to hold the balance of all non-empty addresses. Besides lightweight client could also help fix this problem. A novel schem named VerSum [38] was proposed to provide another way allowing lightweight clients to exist. VerSum allows lightweight clients to outsource expensive computations over large inputs. It ensures the computation result is correct through comparing results from multiple servers.
- *Redesigning blockchain*. In [39], Bitcoin-NG (Next Generation) was proposed. The main idea of Bitcoin-NG is to decouple conventional block into two parts: key block for leader election and microblock to store transactions. The protocol divides time into epoches. In each epoch, miners have to hash to generate a key block. Once the key block is generated, the node becomes the leader who is responsible for generating microblocks. Bitcoin-NG also extended the heaviest (longest) chain strategy in which microblocks carry no weight. In this way, blockchain is redesigned and the tradeoff between block size and network security has been addressed.

B. Privacy Leakage

Blockchain can preserve a certain amount of privacy through the public key and private key. Users transact with their private key and public key without any real identity exposure. However, it is shown in [40], [5] that blockchain cannot guarantee the *transactional privacy* since the values of all transactions and balances for each public key are publicly visible. Besides, the recent study [41] has shown that a user's Bitcoin transactions can be linked to reveal user's information. Moreover, Biryukov et al. [11] presented an method to link user pseudonyms to IP addresses even when users are behind Network Address Translation (NAT) or firewalls. In [11], each client can be uniquely identified by a set of nodes it connects to. However, this set can be learned and used to find the origin of a transaction. Multiple methods have been proposed to improve anonymity of blockchain, which could be roughly categorized into two types:

- *Mixing* [42]. In blockchain, users addresses are pseudonymous. But it is still possible to link addresses to user real identity as many users make transactions with the same address frequently. Mixing service is a kind of service which provides anonymity by transferring funds from multiple input addresses to multiple output addresses. For example, user Alice with address A wants to send some funds to Bob with address B. If Alice directly makes a transaction with input address A and output address B, relationship between Alice and Bob might be revealed. So Alice could send funds to a trusted intermediary Carol. Then Carol transfer funds to Bob with multiple inputs c1, c2, c3, etc., and multiple output d1, d2, B, d3, etc. Bob's address B is also contained in the output addresses. So it becomes harder to reveal relationship between Alice and Bob. However, the intermediary could be dishonest and reveal Alice and Bob's private information on purpose. It is also possible that Carol transfers Alice's funds to her own address instead of Bob's address. Mixcoin [43] provides a simple method to avoid dishonest behaviours. The intermediary encrypts users' requirements including funds amount and transfer date with its private key. Then if the intermediary did not transfer the money, anybody could verify that the intermediary cheated. However, theft is detected but still not prevented. Coinjoin [44] depends on a central mixing server to shuffle output addresses to prevent theft. And inspired by Coinjoin, CoinShuffle [45] uses decryption mixnets for address shuffling.
- *Anonymous*. In Zerocoin [46], zero-knowledge proof is used. Miners do not have to validate a transaction with digital signature but to validate coins belong to a list of valid coins. Payment's origin are unlinked from transactions to prevent transaction graph analyses. But it still reveals payments' destination and amounts. Zerocash [47] was proposed to address this problem. In Zerocash, zero-knowledge Succinct Non-interactive Arguments of Knowledge (zk-SNARKs) is leveraged. Transaction amounts and the values of coins held by users are hidden.

C. Selfish Mining

Blockchain is susceptible to attacks of colluding selfish miners. In particular, Eyal and Sirer [10] showed that the network is vulnerable even if only a small portion of the hashing power is used to cheat. In selfish mining strategy, selfish miners keep their mined blocks without broadcasting and the private branch would be revealed to the public only if some requirements are satisfied. As the private branch is longer than the current public chain, it would be admitted by all miners. Before the private blockchain publishment, honest miners are wasting their resources on an useless branch while selfish miners are mining their private chain without competitors. So selfish miners tend to get more revenue.

Based on selfish mining, many other attacks have been proposed to show that blockchain is not so secure. In stubborn mining [48], miners could amplify its gain by non-trivially composing mining attacks with network-level eclipse attacks. The trail-stubbornness is one of the stubborn strategy that miners still mine the blocks even if the private chain is left behind. Yet in some cases, it can result in 13% gains in comparison with a non-trail-stubborn counterpart. [49] shows that there are selfish mining strategies that earn more money and are profitable for smaller miners compared to simple selfish mining. But the gains are relatively small. Furthermore, it shows that attackers with less than 25% of the computational resources can still gain from selfish mining. To help fix the selfish mining problem, Heilman [50] presented an novel approach for honest miners to choose which branch to follow. With random beacons and timestamps, honest miners would select more fresh blocks. However, [50] is vulnerable to

在自私挖矿的基础上，人们还提出了许多其他攻击，表明区块链并不那么安全。在顽固挖矿[48]**中，**矿工可以通过将挖矿攻击与网络级日蚀攻击进行非三维组合来扩大收益。 顽固挖矿策略中的"踪迹顽固性"（trail-stubbornness）是一种顽固策略，**即即使私有**链被留下，矿工仍会挖出区块。然而，**在某些情况下，与无"踪迹**顽固"的对应策略相比，它能带来 13% **的收益。**[49]**表明，**与简单的自私挖矿相比，有一些自私挖矿策略能赚更多的钱，而且对小矿工来说也有利可图。但收益相对较小。

forgeable timestamps. ZeroBlock [51] builds on the simple scheme: Each block must be generated and accepted by the network within a maximum time interval. Within ZeroBlock, selfish miners cannot achieve more than its expected reward.

V. Possible future directions

Blockchain has shown its potential in industry and academia. We discuss possible future directions with respect to four areas: *blockchain testing*, *stop the tendency to centralization*, *big data analytics* and *blockchain application*.

A. Blockchain testing

Recently different kinds of blockchains appear and over 700 cryptocurrencies are listed in [52] up to now. However, some developers might falsify their blockchain performance to attract investors driven by the huge profit. Besides that, when users want to combine blockchain into business, they have to know which blockchain fits their requirements. So blockchain testing mechanism needs to be in place to test different blockchains.

Blockchain testing could be separated into two phases: *standardization phase* and *testing phase*. In standardization phase, all criteria have to be made and agreed. When a blockchain is born, it could be tested with the agreed criteria to valid if the blockchain works fine as developers claim. As for testing phase, blockchain testing needs to be performed with different criteria. For example, an user who is in charge of online retail business cares about the throughput of the blockchain, so the examination needs to test the average time from a user send a transaction to the transaction is packed into the blockchain, capacity for a blockchain block and etc.

区块链被设计为一个去中心化的系统。然而，矿工集中在矿池中的趋势正在形成。到目前为止，排名前五的矿池合计拥有比特币网络总哈希算力的 51% **以上**[53]。**除此之外，**自私的挖矿策略[10]**表明，**拥有超过 25% 总算力的矿池可以获得比其公平份额更多的收益。理性的矿工会加入自私的矿池中，最后矿池的算力很容易超过总算**力的** 51%。由于区块链的目的不是为少数组织服务，因此应该提出一些方法来解决这个问题。

B. Stop the tendency to centralization

Blockchain is designed as a decentralized system. However, there is a trend that miners are centralized in the mining pool. Up to now, the top 5 mining pools together owns larger than 51% of the total hash power in the Bitcoin network [53]. Apart from that, selfish mining strategy [10] showed that pools with over 25% of total computing power could get more revenue than fair share. Rational miners would be attracted into the selfish pool and finally the pool could easily exceed 51% of the total power. As the blockchain is not intended to serve a few organizations, some methods should be proposed to solve this problem.

C. Big data analytics

Blockchain could be well combined with big data. Here we roughly categorized the combination into two types: *data management* and *data analytics*. As for data management, blockchain could be used to store important data as it is distributed and secure. Blockchain could also ensure the data is original. For example, if blockchain is used to store patients health information, the information could not be tampered and it is hard to stole those private information. When it comes to data analytics, transactions on blockchain could be used for big data analytics. For example, user trading patterns might

区块链上的交易可用于大数据分析。例如，可以提取用户的交易模式。用户可以通过分析预测潜在合作伙伴的交易行为。

be extracted. Users can predict their potential partners' trading behaviours with the analysis.

D. Blockchain applications

Currently most blockchains are used in the financial domain, more and more applications for different fields are appearing. Traditional industries could take blockchain into consideration and apply blockchain into their fields to enhance their systems. For example, user reputations could be stored on blockchain. At the same time, the up-and-coming industry could make use of blockchain to improve performance. For example, Arcade City [51], a ridesharing startup offers an open marketplace where riders connect directly with drivers by leveraging blockchain technology.

A smart contract is a computerized transaction protocol that executes the terms of a contract [54]. It has been proposed for long time and now this concept can be implemented with blockchain. In blockchain, smart contract is a code fragment that could be executed by miners automatically. Smart contract has transformative potential in various fields like financial services and IoT.

VI. Conclusion

Blockchain has shown its potential for transforming traditional industry with its key characteristics: decentralization, persistency, anonymity and auditability. In this paper, we present a comprehensive overview on blockchain. We first give an overview of blockchain technologies including blockchain architecture and key characteristics of blockchain. We then discuss the typical consensus algorithms used in blockchain. We analyzed and compared these protocols in different respects. Furthermore, we listed some challenges and problems that would hinder blockchain development and summarized some existing approaches for solving these problems. Some possible future directions are also proposed. Nowadays blockchain-based applications are springing up and we plan to conduct in-depth investigations on blockchain-based applications in the future.

Acknowledgement

The work described in this paper was supported by the National Key Research and Development Program (2016YFB1000101), the National Natural Science Foundation of China under (61472338), the Fundamental Research Funds for the Central Universities and Macao Science and Technology Development Fund under Grant No. 096/2013/A3. The authors would like to thank Gordon K.-T. Hon for his constructive comments.

References

[1] "State of blockchain q1 2016: Blockchain funding overtakes bitcoin," 2016. [Online]. Available: http://www.coindesk.com/state-of-blockchain-q1-2016/

[2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[3] G. W. Peters, E. Panayi, and A. Chapelle, "Trends in crypto-currencies and blockchain technologies: A monetary theory and regulation perspective," 2015. [Online]. Available: http://dx.doi.org/10.2139/ssrn.2646618

[4] G. Foroglou and A.-L. Tsilidou, "Further applications of the blockchain," 2015.

[5] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: The blockchain model of cryptography and privacy-preserving smart contracts," in *Proceedings of IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, USA, 2016, pp. 839–858.

[6] B. W. Akins, J. L. Chapman, and J. M. Gordon, "A whole new world: Income tax considerations of the bitcoin economy," 2013. [Online]. Available: https://ssrn.com/abstract=2394738

[7] Y. Zhang and J. Wen, "An iot electric business model based on the protocol of bitcoin," in *Proceedings of 18th International Conference on Intelligence in Next Generation Networks (ICIN)*, Paris, France, 2015, pp. 184–191.

[8] M. Sharples and J. Domingue, "The blockchain and kudos: A distributed system for educational record, reputation and reward," in *Proceedings of 11th European Conference on Technology Enhanced Learning (EC-TEL 2015)*, Lyon, France, 2015, pp. 490–496.

[9] C. Noyes, "Bitav: Fast anti-malware by distributed blockchain consensus and feedforward scanning," *arXiv preprint arXiv:1601.01405*, 2016.

[10] I. Eyal and E. G. Sirer, "Majority is not enough: Bitcoin mining is vulnerable," in *Proceedings of International Conference on Financial Cryptography and Data Security*, Berlin, Heidelberg, 2014, pp. 436–454.

[11] A. Biryukov, D. Khovratovich, and I. Pustogarov, "Deanonymisation of clients in bitcoin p2p network," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, 2014, pp. 15–29.

[12] F. Tschorsch and B. Scheuermann, "Bitcoin and beyond: A technical survey on decentralized digital currencies," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 2084–2123, 2016.

[13] NRI, "Survey on blockchain technologies and related services," Tech. Rep., 2015. [Online]. Available: http://www.meti.go.jp/english/press/2016/pdf/0531_01f.pdf

[14] D. Lee Kuo Chuen, Ed., *Handbook of Digital Currency*, 1st ed. Elsevier, 2015. [Online]. Available: http://EconPapers.repec.org/RePEc:eee:monogr:9780128021170

[15] V. Buterin, "A next-generation smart contract and decentralized application platform," *white paper*, 2014.

[16] D. Johnson, A. Menezes, and S. Vanstone, "The elliptic curve digital signature algorithm (ecdsa)," *International Journal of Information Security*, vol. 1, no. 1, pp. 36–63, 2001.

[17] V. Buterin, "On public and private blockchains," 2015. [Online]. Available: https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/

[18] "Hyperledger project," 2015. [Online]. Available: https://www.hyperledger.org/

[19] "Consortium chain development." [Online]. Available: https://github.com/ethereum/wiki/wiki/Consortium-Chain-Development

[20] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 4, no. 3, pp. 382–401, 1982.

[21] S. King and S. Nadal, "Ppcoin: Peer-to-peer crypto-currency with proof-of-stake," *Self-Published Paper, August*, vol. 19, 2012.

[22] "Bitshares - your share in the decentralized exchange." [Online]. Available: https://bitshares.org/

[23] D. Schwartz, N. Youngs, and A. Britto, "The ripple protocol consensus algorithm," *Ripple Labs Inc White Paper*, vol. 5, 2014.

[24] J. Kwon, "Tendermint: Consensus without mining," *URL http://tendermint. com/docs/tendermint {_} v04. pdf*, 2014.

[25] S. King, "Primecoin: Cryptocurrency with prime number proof-of-work," *July 7th*, 2013.

[26] P. Vasin, "Blackcoins proof-of-stake protocol v2," 2014. [Online]. Available: https://blackcoin.co/blackcoin-pos-protocol-v2-whitepaper.pdf

[27] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, 2014.

[28] V. Zamfir, "Introducing casper the friendly ghost," *Ethereum Blog URL: https://blog. ethereum. org/2015/08/01/introducing-casper-friendly-ghost*, 2015.

[29] C. Miguel and L. Barbara, "Practical byzantine fault tolerance," in *Proceedings of the Third Symposium on Operating Systems Design and Implementation*, vol. 99, New Orleans, USA, 1999, pp. 173–186.

[30] D. Mazieres, "The stellar consensus protocol: A federated model for internet-level consensus," *Stellar Development Foundation*, 2015.

[31] "Antshares digital assets for everyone," 2016. [Online]. Available: https://www.antshares.org

[32] M. Vukolić, "The quest for scalable blockchain fabric: Proof-of-work vs. bft replication," in *International Workshop on Open Problems in Network Security*, Zurich, Switzerland, 2015, pp. 112–125.

[33] C. Decker, J. Seidel, and R. Wattenhofer, "Bitcoin meets strong consistency," in *Proceedings of the 17th International Conference on Distributed Computing and Networking (ICDCN)*. Singapore, Singapore: ACM, 2016, p. 13.

[34] D. Kraft, "Difficulty control for blockchain-based consensus systems," *Peer-to-Peer Networking and Applications*, vol. 9, no. 2, pp. 397–413, 2016.

[35] Y. Sompolinsky and A. Zohar, "Accelerating bitcoin's transaction processing. fast money grows on trees, not chains." *IACR Cryptology ePrint Archive*, vol. 2013, no. 881, 2013.

[36] A. Chepurnoy, M. Larangeira, and A. Ojiganov, "A prunable blockchain consensus protocol based on non-interactive proofs of past states retrievability," *arXiv preprint arXiv:1603.07926*, 2016.

[37] J. Bruce, "The mini-blockchain scheme," July 2014. [Online]. Available: http://cryptonite.info/files/mbc-scheme-rev3.pdf

[38] J. van den Hooff, M. F. Kaashoek, and N. Zeldovich, "Versum: Verifiable computations over large public logs," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, 2014, pp. 1304–1316.

[39] I. Eyal, A. E. Gencer, E. G. Sirer, and R. Van Renesse, "Bitcoin-ng: A scalable blockchain protocol," in *Proceedings of 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, Santa Clara, CA, USA, 2016, pp. 45–59.

[40] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: Characterizing payments among men with no names," in *Proceedings of the 2013 Conference on Internet Measurement Conference (IMC'13)*, New York, NY, USA, 2013.

[41] J. Barcelo, "User privacy in the public bitcoin blockchain," 2014.

[42] M. Möser, "Anonymity of bitcoin transactions: An analysis of mixing services," in *Proceedings of Münster Bitcoin Conference*, Münster, Germany, 2013, pp. 17–18.

[43] J. Bonneau, A. Narayanan, A. Miller, J. Clark, J. A. Kroll, and E. W. Felten, "Mixcoin: Anonymity for bitcoin with accountable mixes," in *Proceedings of International Conference on Financial Cryptography and Data Security*, Berlin, Heidelberg, 2014, pp. 486–504.

[44] G. Maxwell, "Coinjoin: Bitcoin privacy for the real world," in *Post on Bitcoin Forum*, 2013.

[45] T. Ruffing, P. Moreno-Sanchez, and A. Kate, "Coinshuffle: Practical decentralized coin mixing for bitcoin," in *Proceedings of European Symposium on Research in Computer Security*, Cham, 2014, pp. 345–364.

[46] I. Miers, C. Garman, M. Green, and A. D. Rubin, "Zerocoin: Anonymous distributed e-cash from bitcoin," in *Proceedings of IEEE Symposium Security and Privacy (SP)*, Berkeley, CA, USA, 2013, pp. 397–411.

[47] E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Zerocash: Decentralized anonymous payments from bitcoin," in *Proceedings of 2014 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, USA, 2014, pp. 459–474.

[48] K. Nayak, S. Kumar, A. Miller, and E. Shi, "Stubborn mining: Generalizing selfish mining and combining with an eclipse attack," in *Proceedings of 2016 IEEE European Symposium on Security and Privacy (EuroS&P)*, Saarbrucken, Germany, 2016, pp. 305–320.

[49] A. Sapirshtein, Y. Sompolinsky, and A. Zohar, "Optimal selfish mining strategies in bitcoin," *arXiv preprint arXiv:1507.06183*, 2015.

[50] S. Billah, "One weird trick to stop selfish miners: Fresh bitcoins, a solution for the honest miner," 2015.

[51] S. Solat and M. Potop-Butucaru, "ZeroBlock: Timestamp-Free Prevention of Block-Withholding Attack in Bitcoin," Sorbonne Universites, UPMC University of Paris 6, Technical Report, May 2016. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01310088

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

[53] "The biggest mining pools." [Online]. Available: https://bitcoinworldwide.com/mining/pools/

[54] N. Szabo, "The idea of smart contracts," 1997.