Journal Pre-proof

Scalable blockchains — A systematic review

Muhammad Hassan Nasir, Junaid Arshad, Muhammad Mubashir Khan, Mahawish Fatima, Khaled Salah, Raja Jayaraman

PII: S0167-739X(21)00297-1
DOI: https://doi.org/10.1016/j.future.2021.07.035
Reference: FUTURE 6202

To appear in: *Future Generation Computer Systems*

Received date : 30 November 2020
Revised date : 2 May 2021
Accepted date : 30 July 2021

Please cite this article as: M.H. Nasir, J. Arshad, M.M. Khan et al., Scalable blockchains — A systematic review, *Future Generation Computer Systems* (2021), doi: https://doi.org/10.1016/j.future.2021.07.035.

Scalable Blockchains - A Systematic Review

Muhammad Hassan Nasir[a], Junaid Arshad[b], Muhammad Mubashir Khan[a], Mahawish Fatima[c], Khaled Salah[d], Raja Jayaraman[e]

[a]*Department of Computer Science & IT, NED University of Engineering & Technology, Pakistan*
[b]*School of Computing and Digital Technology, Birmingham City University, UK*
[c]*Department of Software Engineering, Bahria University Karachi Campus (BUKC), Pakistan*
[d]*Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi 127788, UAE*
[e]*Department of Industrial & Systems Engineering, Khalifa University, Abu Dhabi 127788, UAE.*

Abstract

Blockchains are a key disruptive technology of recent times with applications across several application domains such as finance, supply chain, healthcare and Industry 4.0. With growing application types, scalability of blockchains has emerged as a key challenge in achieving the potential of this technology. This paper attempts to outline this challenge for blockchains as follows. It highlights blockchain scalability as a composite concept which can be interpreted in various ways i.e. expanding the number of participants of the network, or enhancing capabilities of the participants (such as processing, memory, storage or an efficient consensus strategy) to minimize the scalability overhead. Further, it presents a first thorough systemic study of existing efforts to achieve blockchain scalability with respect to; i) innovative methods and mechanisms to strengthen blockchain's capabilities with respect to scalability, ii) approaches leveraging blockchains to achieve scalable applications, and iii) efforts analyzing blockchain solutions with respect to scalability. Through a critical analysis of existing literature, the paper identifies contributions as well as open challenges which can help research community to advance knowledge within this domain.

Keywords: Blockchain, Scalability, Blockchain Throughput, Systematic Review, Transaction Processing Speed

1. Introduction

Decentralization is a key theme for recent advancements within computing systems with emerging applications in diverse domains, such as Industrial Internet of Things (IIoT), autonomous vehicles, supply chain management, and finance. With cutting-edge computing paradigms such as Cyber-Physical Systems (CPS), interaction among multiple different, geographically dispersed entities in an automated manner is critical to their wide-spread adoption. Distributed Ledger Technologies (DLTs) have recently emerged as enabling platforms to aid the drive towards decentralization within computing systems. A DLT typically involves participants across different networks to record and share data via a synchronized ledger. By providing fundamental properties such as public availability (typically), and immutability (tamper-proof) of the ledgers, DLTs foster trustworthy interactions in untrustworthy settings [1].

A Blockchain system exemplifies distributed ledger technologies enabling a digital transaction's data to be stored in the form of a chained block where each participant stores the transaction records in these connected blocks and the transactions are included with the consensus of verifying nodes also known as *miners*. A blockchain based system has inherent properties of decentralization, transparency, trustless-trust, immutability, and traceability, since it allows each participating node to have equal opportunity to influence the ledger. The transparency and immutability is enforced by allowing the nodes to view and maintain the ledger which can only be altered with the consensus of other mining nodes [2]. Although Bitcoin [2] still represents the most popular application of blockchain, use of blockchain has witnessed significant attention across diverse application domains including healthcare, finance, e-government [3], and supply chain management [4]. Blockchain features such as security, immutability and decentralization have a profound role in this. The widespread adoption of blockchain has been facilitated by research on blockchain beyond cryptocurrencies where significant efforts have been made to devise different business models, its applicability in various domains and to optimize its adaptability for different applications. Furthermore, a number of platforms such as Ethereum [5], Multichain [6], and Hyperledger [7] have been developed to facilitate adoption of blockchain technology beyond cryptocurrency and have had significant impact to achieve its widespread adoption.

In addition to the inherent benefits of blockchain, the increasing use of blockchain within diverse domains has also identified several challenges which have to be address for widespread adoption. For instance, transaction malleability [8] is considered a significant threat for blockchain-based applications and therefore requires appropriate protection mechanisms to mitigate against it.

One of the significant factors in the adoption of blockchain is its ability to facilitate large number of transactions which require a scalable solution. An empirical analysis of Bitcoin reveals that it supports *seven* transactions per second which is significantly low when compared to 2000 transactions per second achieved by VISA. Similarly, applications such as public voting require concurrent processing of large number of transactions [9]. Such scenarios present opportunities for further dedicated research into the ability of blockchain technology to facilitate development of scalable applications and associated challenges. For instance, a common belief is to treat blockchain as a black box which can be used to develop decentralized applications that are inherently efficient. However, in order to use blockchain to achieve efficient decentralized application, a rigorous analysis of characteristics such as block size, transaction processing speed and block generation rate is required. For instance, size of block in Bitcoin was set to 1MB, and it takes on average 10 minutes to generate a block. Enhancing the block size may result in improving throughput by including more transactions in a single block, but leads to increase in the block propagation time. On the other hand reducing the block size leads to decrease in latency but can also lead to a problem of generation of multiple branches of chain known as forks [10].

1.1. Scope of study:

In this paper, we aim to achieve a robust understanding of the challenge of achieving scalable applications using blockchain whilst highlighting the role of specific parameters such as block size, transaction processing speed, and block generation rate.

Through our research we identified existing surveys to review the state of the art with respect to scalable blockchains such as [11], [12], [13] and [14]. However, these studies are limited in that these have a limited scope in terms of coverage of different dimensions of scalability (horizontal and vertical, detailed in section 3) as well as depth and breadth of literature reviewed. For instance, [11] presented a comparison of POW and BFT based protocol families with respect to scalability, whereas [15] and [16] presented a comparative study of selected consensuses approaches. Although consensus algorithm is an important component of the overall blockchain system however other aspects of the system also have an impact on its scalability, which should be taken into account to achieve a comprehensive study. Furthermore, Kim et al. [17] presented an effort limited to only two aspects of vertical scalability i.e. throughput & storage. Additionally, existing reviews of scalable blockchains are also limited with respect to the methodology used to conduct the study with only Fournier et al. [14] representing a systematic approach. However this effort is limited in terms of its coverage as it is only focused at some aspects of scalability. Furthermore, the study has a weak article inclusion criteria with limited paper coverage and scope. Contrary to existing studies, we attempt to present a comprehensive study which adopts a systematic approach and takes into account both horizontal and vertical scalability thereby achieving depth and breadth of existing literature.

In addition to the limitations of existing efforts to review state of the art within scalable blockchains, the widespread use of blockchains has also resulted in increased awareness with regards to scalability as being one of the primary challenges encountered in widespread adoption of blockchains. Therefore, our study aims to present a comprehensive review of state of the art within scalable blockchains addressing the gaps in the existing surveys so as to help research community to contribute towards addressing this challenge.

In particular, we define scalability as a composite concept comprised of horizontal (node and client scalability) and vertical (throughput, block generation rate, latency, and storage) scalability. We conduct a first thorough systematic study of existing efforts to achieve scalable blockchains, focusing at three specific categories i.e. i) innovative methods and mechanisms to strengthen blockchain core (e.g. sharding, on- and off-chain solutions, blockchain redesign), ii) approaches leveraging blockchains to achieve scalable applications, and iii) efforts analyzing blockchain solutions with respect to scalability. Through a critical review of these efforts, we identify contributions from existing literature and highlight limitations to outline open challenges which require further efforts by the research and practice community.

Specifically, we make the following major contributions through this study:

- Definition of scalability within the context of blockchains which encompasses both horizontal and vertical dimensions thereby achieving a broad coverage.
- An in-depth study and analysis of existing literature with respect to blockchain scalability. This review provides a critical insight into the state of the art within scalable blockchains focusing at blockchain enhancement mechanisms, blockchain-based scalable applications, and scalability analysis of blockchains. Therefore, this paper provides a significant snapshot of existing literature in this domain.
- Identification of open challenges to achieve scalable blockchains based on critical review of existing literature aiming to help channel efforts by research community to achieve scalability in blockchain-based applications.

1.2. Organization of the paper:

Rest of the paper is organized as follows. First, an overview of Blockchain technology, its basic functions, types and applications with proper depth and sufficient breadth is presented in section 2. Section 3 discusses various dimensions of scalability, such as throughput, latency, block generation rate etc, which is a key barrier in applying the

blockchain technology in various domains. Following that, various factors that can affect scalability are discussed in section 4. Section 5 contains research methodology which is mainly based on kitchenham et al. strategy [18]. Sections 6 7 discusses the state of the art solutions to improve one or more aspects of scalability and applications respectively. Discussions and analysis on available solutions is presented in section 8 and finally, section 9 contains the conclusions and future directions.

2. Blockchain Technology

Blockchain is a distributed ledger technology which aids decentralized, distributed computing in a trustless environment. The attention gained by blockchain is primarily due to its most popular application i.e. Bitcoin [2] which effectively seeks to conduct financial transactions in a peer-to-peer model without the support of conventional banking system. Although Bitcoin was setup in 2009, the concepts and technologies which underpin blockchain in general and Bitcoin in particular have evolved over the last few decades. For instance, the concept of anonymous transactions (using pseudonyms) which cannot be traced back to the sender, was introduced by David Chaum in 1983 [19] and today serves as one of the elementary concepts within Bitcoin. Similarly, the consensus algorithm used by Bitcoin (Proof of Work (PoW)) is based on the HashCash [20] a countermeasure originally developed to protect against unsolicited email in 1999.

Distributed ledger technologies in general and blockchain in particular can be represented in terms of different layers involved such as infrastructure, data, network, and application. We present a layered view of blockchain in Figure 1 and use this structure to describe different layers and the components within them. In describing these layers and components, we have used language which is agnostic of technical terminologies used within a specific platform such as Ethereum, Hyperledger or Multichain.

2.1. Hardware or infrastructure layer

A typical blockchain network is comprised of multiple physical nodes which are potentially distributed across different geographic locations. These nodes can either be hosted within a cloud or a be part of an organization's corporate network with connections to other resources such as file storage. Blockchain is a peer-to-peer network which means all the participating nodes are connected to each other however this connectivity is achieved through traditional infrastructure governing the Internet. This network of computers computes transactions, validates them, and stores them in a ledger shared across all participating nodes. Data recorded in the blockchain can be stored locally on the participating nodes as part of; *on-chain storage* or in remote locations such as clouds as part of the *off-chain storage*.

2.2. Data layer

Blockchain is a linked-chain of blocks where transactions are stored in an ordered form. The data contained within a block can be divided into two types i.e. *block header* and *block body* as illustrated in Fig 2. Block body contains one or more transactions (data stored in the ledeger as part of a transaction) and the transaction counter. Block header contains metadata about the block such as block version, Merkle tree root hash, time stamp, nonce and parent (previous) block hash. The data within a blockchain is immutable and can not be modified or altered once it is added to the chain.

Since the data is stored in an encrypted form, the identification of each block and the detection of any modification, is done by generating a fixed-length hash of the data via a cryptographic hash algorithm such as SHA 256. These hashes are stored in the form of a binary tree known as Merkle Tree (MT) (Patricia tree in Ethereum) and consist of valuable information of each block typically varying based on the type of blockchain (public or private). Figure 3 presents an in-depth view of a blockchain highlighting different elements of a block and the connection between different blocks. To ensure integrity, security, and immutability, all transactions are digitally signed with a private key using asymmetric cryptography and only the public key holder can verify the signing entity. These signatures secure the identity of the sender (achieving authenticity and non-repudiation) as well as guarantee the integrity of the data by detecting any malicious tampering.

2.3. Network layer

Although the blockchain network is using Internet infrastructure, it is still a peer to peer network. The network layer, also known as the P2P layer or propagation layer, is responsible for overall node-to-node communication, discovery, and synchronization. It also takes care of state propagation to maintain the existing global state of the blockchain. In a public blockchain, any node can participate in the network by implementing the protocol. However, private blockchain is typically governed by a single entity that grants permission to the nodes to take part in the network.

Typically, there are various types of nodes in a blockchain network, and depending upon their roles they perform certain tasks. However, there are two major types; a *full node*, and a *normal node*. A full node, also known as a validator node or a mining node, maintains a copy of all the history of transactions. Moreover, it takes part in the verification and authentication process including validation of the transaction's signature, and mining. It also takes care of the execution and implementation of consensus rules and processes. These nodes are the root of trust in a blockchain network. A mining node must always be a full node since it needs all transaction history to validate a transaction in a

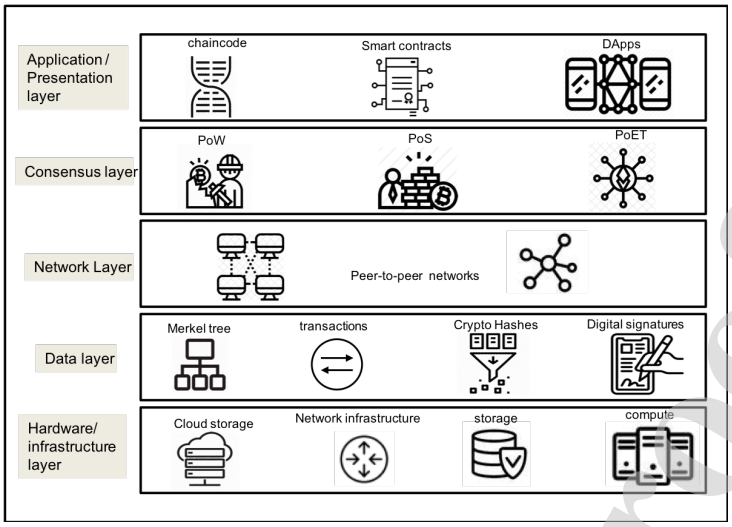


Figure 1: A layered view of blockchain

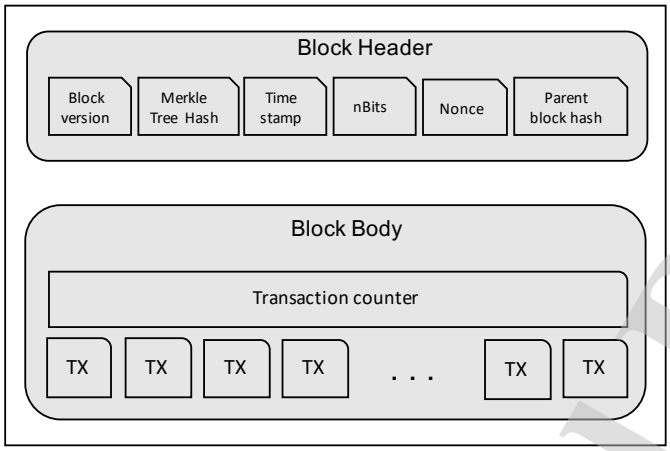


Figure 2: Structure of a block within blockchain

newly generated block. However, a full node need not necessarily be a mining node. It may take care of other issues (discussed above) without creating/mining new blocks.

On the contrary, a normal node, also known as lightweight node and Simplified Payment Verification (SPV) node in Bitcoin, do not perform verification/validation and maintains the storage of complete history of transactions, rather it can only generates and sends transactions while maintaining the header of the blockchain. Although, a lightweight node consumes less space, but these nodes are solely dependent upon full nodes even for a balance query, block synchronization, and correctness of data. For instance, in Bitcoin and Ethereum environment, the storage requirement of a full node is reaching 200GB and 1TB respectively. Whereas, a half node requires less than 50GB and 100GB for Bitcoin and Ethereum respectively.

2.4. Consensus layer

The consensus layer plays a vital role in maintaining the reliability of the blockchain platform. The consensus is a set of rules, enforced by this layer, which must be followed by every participant to ensure smooth generation of valid transactions/blocks. Typically consensus algorithms guarantees the creation of a single continuous chain and prevents the formation of forks (i-e. multiple branches of a single chain). It acts as a core to maintain the agreement and synchronization of each node within a blockchain network. This layer is responsible for managing, maintaining, and administering block generation, validation, and ordering with the consensus of all participants distributed over a P2P network. It has the mechanism to guarantee the distribution of power across the network which consequently prevents tampering by a malicious adversary. It also takes care of the reward mechanism for mining/validator nodes.

There are various types of consensus approaches to ensure the consistency of the blockchain, including probabilistic and deterministic approaches. In the probabilistic approach, various nodes can have different views of the chained blocks which may result in the formation of forks. These forks are usually prevented by the longest chain rule. The examples of probabilistic approach are Bitcoin and Ethereum platforms which employ PoW (permissioned) and PoS(permissioned/permission-less) consensus methods respectively. On the other hand, in the deterministic approaches, the blocks are mined/validated by the special type of nodes known as ordering nodes. Therefore the possibility of fork generation is prevented. The example of this type of approach is Hyperledger [7] which employs PBFT(permissioned) consensus method.

2.5. Application or presentation layer

The application layer consists of three major components; smart contracts, Decentralized Application(DApps),

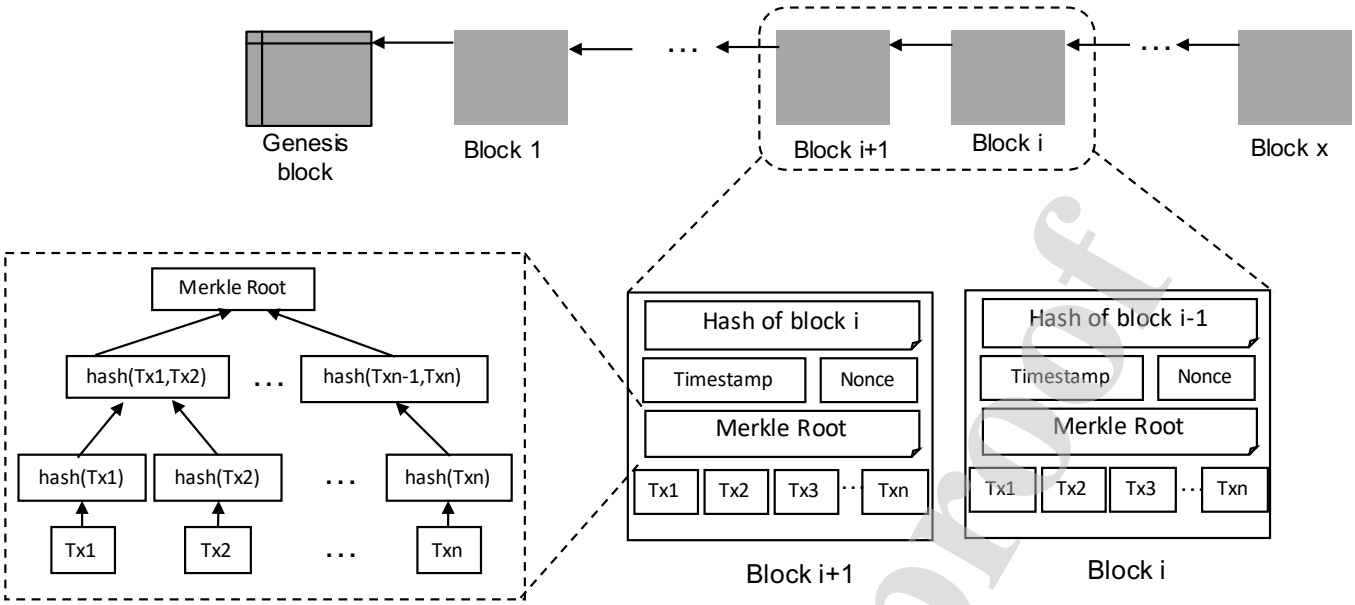


Figure 3: In-depth view of a blockchain

and chaincode. This layer can be sub-divided into application and execution layers. The application layer has various elements including interfaces, APIs, and scripts through which the user can interact with the blockchain core function whereas the execution layer contains the software code that executes the rules.

Smart Contracts (SC) and Chaincode are often used as synonyms and refer to the executable code that is intended to control the actions as per a contract or agreement. Typically, they hold the transaction logic that has the power to control the overall business life-cycle. They consist of methods which are executed when a transaction is committed. The major difference between the SC and chaincode is that the former is typically distributed on-chain having their own address, usually written in solidity/Python language and deployed in Ethereum Virtual Machine (EVM) for compilation/transformation of code into byte-code for further deployment into a blockchain platform. The later, is the term used by Hyperledger Fabric community, usually written in java or Go language and executes separately (off the chain) from the validation process in the form of a secure docker container that is available on each blockchain instance. It also initiates and controls the interaction of applications with blockchain ledger via transactions submitted by DApps.

A **DApp** are user-friendly programs that execute on the top of blockchain and provides interaction with the SC and chaincode, which in turn provides interaction with the blockchain platform. These applications are distributed and are controlled by multiple entities.

2.6. Blockchain process flow

Blockchain introduces a decentralized transaction processing and storage system whereby the contents of the shared ledger depend upon participating validator (permissioned blockchain) or mining (public blockchain) nodes who confirm the transaction legitimacy based on a predefined consensus process. The overall transaction process flow is illustrated in Figure 4.

As illustrated in Figure 3, the root of a blockchain is the *genesis block* also known as *block 0* which is the first block created to form the blockchain. In terms of its contents, genesis block is similar to any other block in the blockchain with the exception that it does not contain the hash of previous block. In a typical blockchain setting, a client (often represented by a process) generates a transaction in the form of a block and broadcasts it to the participants of the peer-to-peer network. Upon receiving the transaction, all the participating validator/mining nodes use predefined consensus algorithm (also called *mining*) to verify the transaction. The mining process varies across blockchains and can include probabilistic (PoW and PoS) and deterministic approaches (PBFT). In a typical mining process such as PoW, several mining nodes compete with each other to perform validation and submit a mathematical proof of their work. Typically an incentive mechanism is in place for attracting the miners to use efficient hardware and by incentivizing the winner mining node. Once verified, the transactions are grouped in the form of block which is subsequently added to the blockchain by linking it with the previous block via a crypto-hash in the immutable main-chain to complete the transaction process flow.

2.7. Permissioned vs. Permissionless blockchain

From the perspective of the participation of nodes, a blockchain can be divided into two broad categories; Permissioned and Permissionless blockchains [21]. The **per-**

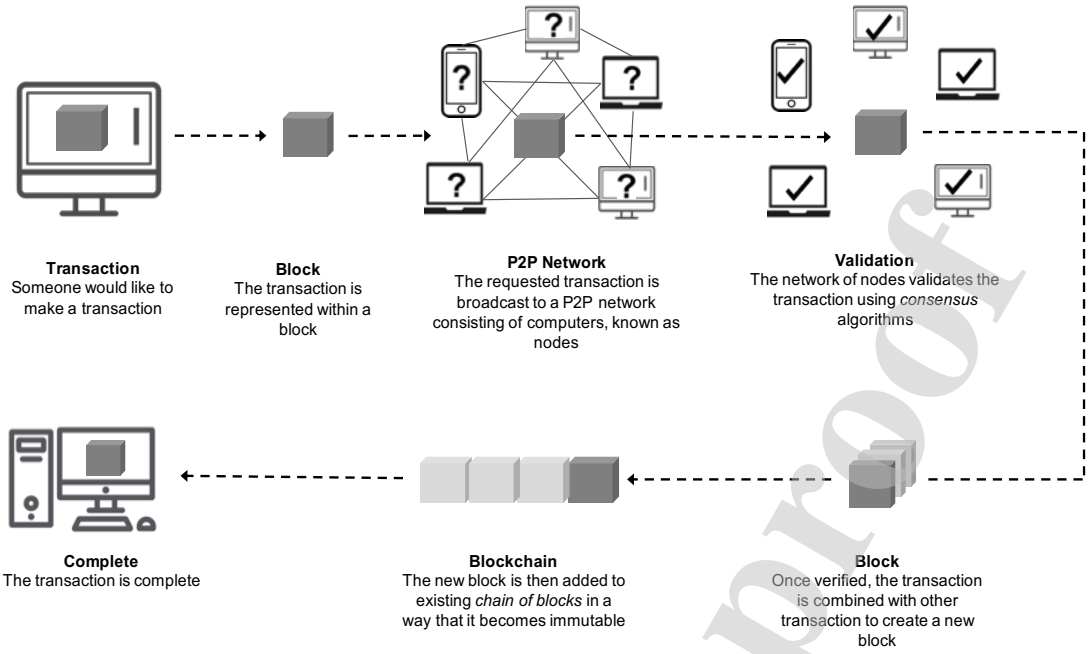


Figure 4: Typical blockchain process flow

missioned blockchains typically have three major characteristics, namely governance structure, private transactions, and an authentication process. They require the approval of the blockchain administrator or owner to join the system and participate in the consensus process [22]. Such type of blockchain are typically private (typical use-cases are Banking and Payments, Food Tracking, Supply chain management, etc.) however, they can be public. They offer extremely fast output, scalability, and energy efficiency but are not truly decentralized and have partial immutability. **Permissionless blockchains**, on the other hand, typically offer true decentralization, transparency, immutability with high security. They require no prior approval of blockchain administrators or owner to join and interact with. They are typically known as public blockchains [23]. They offer slow transaction throughput with limited scalability and are not truly energy efficient. Most of the time, permissionless blockchain is ideal for running and managing digital currencies, public voting, etc.

3. Blockchain Scalability Dimensions

Although blockchain has witnessed significant adoption in recent years, scalability of blockchain-based solutions is one of the major concerns which may limit its role as a disruptive technology [9]. Therefore, our focus in this paper is to investigate and analyze current efforts to enhance the scalability of blockchain. Through our research we have concluded that scalability is a composite term with number of diverse interpretations found in literature. Therefore, in order to define scalability within the context of this study, we leverage definition of scalability in contemporary research where scalability can refer to: i) *Horizontal:* whereby it is achieved by adding/increasing more machines within the existing pool/network, or ii) *Vertical:* in terms of vertical scaling that is achieved by employing more power (processing, storage, memory, efficient strategy etc.) to existing pool of resources [24, 25, 26, 27]. Therefore, we have used these fundamental concepts to define scalability within blockchains.

In addition to the fundamental scalability concepts found in contemporary literature, we have also consulted existing blockchain scalability research such as those discussed in existing surveys within blockchain scalability. Consequently, this section identifies and defines major factors which can be used to describe the scalability of a blockchain system. We define these as different dimensions of blockchain scalability and summarize them in Figure 5.

3.1. Horizontal scalability:

Horizontal scalability refers to the ability of a blockchain to expand the network of participants by adding further nodes and clients. Horizontal scaling is related to improving the core framework of the system that improves the overall throughput of the system by facilitating maximum number of nodes or clients without degradation in efficiency and performance of the blockchain [25, 26]. For example, the PoW based consensus approaches can deliver stable performance with increased number of nodes whereas the performance of Byzantine fault tolerance (BFT) based approaches degrades with increased in number of nodes.

3.1.1. Client scalability:

The clients are the application programs that submit transactions on the behalf of a user and the client scalability refers to the ability of a blockchain system to facilitate an increasing number of clients without affecting the overall performance.

3.1.2. Node scalability:

Node scalability refers to the ability of a blockchain system to facilitate an increasing number of nodes without affecting the overall performance.

3.2. Vertical scalability:

Vertical scalability refers to enhancing the capabilities of participating nodes to achieve efficient processing of transactions [27, 12, 17]. Vertical scalability can be enhanced by adjusting the parameters such as block size, parallel mining, lightening, and sharding etc. Vertical scalability can be further divided into multiple sub-dimensions including throughput, block generation rate, latency, storage scalability which are described below:

3.2.1. Throughput:

Transaction throughput, expressed as transaction per second (TPS) is a major component of scalability. It is the rate at which valid transactions are committed and added to the block after consensus is established among the stakeholders (miners) of the blockchain.

The significant rise in the volume of transactions does not allow current implementations such as Bitcoin, to fulfil the requirements of a transaction hungry environment. The traditional blockchain (Bitcoin) can support a transaction rate of up to seven transactions per second [28]. While other implementations, such as Ethereum [5] and Bitcoin cash [28], perform significantly better and can process up to 20 and 60 TPS respectively however these are still many times less than contemporary digital payment systems such as VISA which can process up to 24000 TPS.[29]

3.2.2. Block generation rate:

In a typical blockchain setting, multiple transactions are grouped together, depending upon the block size, to form a block. In this context, block generation rate (BGR) is the frequency at which a new block is mined, produced and added to the blockchain. The block generation is a resource-hungry process in terms of transaction mining and is dependant upon the size of block and consensus efficiency. For instance, the legacy Nakamoto's algorithm (Bitcoin) has a fixed block side of 1MB and requires each new block to have a unique hashing puzzle solution than previous block to be added to the chain. This results in requisite of 10 minutes time to generate a block [29]. The Ethereum has better BGR which is between 10 to 20 seconds.

3.2.3. Latency:

Within general computer science literature, latency is the time delay between input and output. Low network latency is of paramount importance in blockchain. Within blockchains, latency can refer to two different delays i.e. network latency and transaction's latency. The network latency is the measure of delay between initiating a transaction request and the confirmation of the transaction from the network. A transaction's latency is the measure of consensus efficiency which therein impacts to process and execute large volumes of transactions.

3.2.4. Storage scalability:

Storage scalability refers to the ability of participating nodes to mitigate concerns related to the storage requirements of a blockchain. The storage scalability can further be divided into block size and chain size management sub-dimensions as defined below.

Block size: The block size is the limit of a block to accommodate number of transactions and associated information. For instance, in a typical Bitcoin network, the block size currently stand at 1 megabyte, that can accommodate more than 500 transactions [29]. The increase in block size is an important factor to improve TPS but may have adverse effect on block generation rate. On the other hand, decreasing the block size may lead to formation of forks. Therefore, there is a trade-off between TPS and BGR which impacts selecting the block size.

Chain size: The blockchain is a decentralized ledger in which the number of transactions has witnessed continuous increase. A node, before participating in the network, requires enough storage to download the chain to have a complete view of the network. The Bitcoin blockchain has already exceeded 280 GB [30] which means a miner need to download huge amount of data locally to participate in the network. This factor has significant implications for resource-constrained devices such as Internet of Things (IoT) and therefore requires specific consideration to facilitate adoption of blockchain technology within such systems.

4. Classification of Approaches to Scalable Blockchains

As discussed previously in section 3, the blockchain scalability challenge can be defined through low transaction throughput, increased block/chain/e-signature sizes, network/node/clients storage limitations which limit blockchain's ability to realize its potential as a disruptive technology. Blockchain and cryptocurrency research community have made dedicated efforts to improve blockchain scalability taking into account the trade-off between decentralization, scalability, and security in the blockchain. Based on the findings of our research, we have categorized these scalability solutions into on-chain, off-chain, hardware-assisted solutions, parallel mining and redesigning the blockchain. A comparative analysis of these approaches and their individual limitations is presented in Table 1.

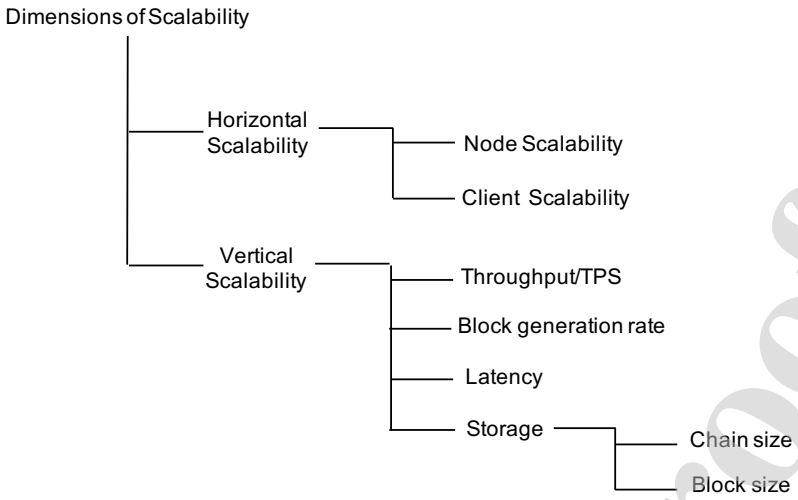


Figure 5: Dimensions of blockchain scalability

Table 1: Comparison of different approaches to blockchain scalability

Scalability Approach	Methodology	Impact on Scalability	Issues/Challenges
Blockchain Pipelining	On-Chain	Throughput	Approaches such as BigchainDB have been known to be susceptible to 33% attack
Blockchain Delivery Network	On-Chain	Throughput & Storage	May Allow Malicious entities to propogate blocks. Relies on Trusted Network Infrastructure
Block Size Adjustment	On-Chain	Throughput, Latency	Application Specific, Forks, Increase BGR,
Payment Channel Networks	Off-Chain	Throughput	Privacy and Security Problem Requires both parties to remain online or lock tokens for committing transactions.
Sharding	Off-Chain	Throughput	Security(Shard Takeover Problem) Communication(Maintaining atomicity and preventing overloading of shard in cross-shard transactions)
Hardware Assisted Approaches	Off-Chain	Throughput, Latency	Increase in number of Nodes degrades TPS/Latency Adds Centralization Requires an attractive incentive mechanism
Parallel Mining	Off-Chain	Throughput, BGR	Requires multiple active miners in an epoch
Redesigning Blockchain	On/Off Chain	Horizontal/Vertical Scalability	Fork avoidance Adversaries avoidance incentive mechanism

4.1. On-Chain solutions:

The on-chain approaches aim to enhance blockchain scalability by adjusting the internal parameters of a blockchain. This can be achieved by improving the network latency or optimizing the transaction or message size. Some of the on-chain solutions are described below:

Blockchain pipelining: Blockchain pipelining, proposed by McConaghy et al.[10], is an on-chain approach to improve transaction throughput by adding blocks in the main-chain without being verified by other nodes. The final decision for the validity of block for the formation of main-chain is done through voting among nodes as a separate layer .

Blockchain delivery network: Through our research we have identified solutions which utilize delivery network including cut-through routing-enabled gateways or cloud delivery network. Such approaches aim to achieve improved transaction throughput, or a collaborative cloud storage mechanism to achieve storage scalability without disturbing the decentralized nature of blockchain.

Block size adjustment: Another approach to attain scalability is to adjust the block size. However, these approaches are application-dependant and require tuning inline with application-specific requirement. For instance, too much increase in block size results in more transactions per block but leads to an increase in propagation time. On the other hand, too much reduction in block size may result in increased BGR with improved latency as well as leading to frequent generation of forks.

4.2. Off-Chain solutions:

The off-chain solutions process transactions outside the main blockchain network to reduce the load from the main

blockchain. These solutions are backed by creating ways that execute on the top of the blockchain. The solutions such as using payment channels (LN or Raiden network) or sharding are the famous off-chain solutions that have significant positive effects on the horizontal as well as vertical scalability. Such solutions particularly present opportunities to develop blockchain-based solutions for settings involving resource-constrained devices such as the IoT.

Payment channel networks: The payment channel allows multiple off-chain transactions between parties by creating a micro-payment channel without having to commit all transactions publicly on the main-chain. It therefore results in increased throughput by reducing workload from the main-chain. In a typical payment channel network, two participants can perform an unlimited number of transactions and only two transactions are required for the main-chain to update a record that can be updated upon completion of all transactions between parties or upon the requirement of an on-chain transaction. Even the parties which are not directly connected can commit the transactions through intermediaries.

Sharding: Typically, a mining node on a blockchain's network is responsible for storing all the states including critical information such as account balance and history of all transactions which reduces the transaction throughput linearly. Sharding is a well-established concept in databases to divide a huge database into smaller and manageable portions to increase its efficiency. In a blockchain, it refers to the horizontal division of the main-chain into multiple independent partitions known as *shards*. Each partition or *shard* is responsible for storing its state. Although sharding is considered to be an off-chain solution, in reality, it does not attempt to move the transactions off the chain but rather divides the main chain into multiple independent groups so that the mining nodes are not responsible for mining every broadcasted transaction on the network. Each shard works separately as an independent blockchain in the network but has a strong cryptographic binding with the main chain appearing as a Merkle tree and therefore can be integrated with the main chain.

4.3. Hardware-assisted approaches:

In addition to the software-based approaches, a number of solutions in literature use specialized trusted hardware devices either to improve consensus or to speed up the transaction process to enhance scalability within the blockchain. The hardware includes high processing machines or Trusted Execution Environment (TEEs) to efficiently handle transactions with accuracy and speed.

4.4. Parallel mining/processing:

Traditional blockchain is based on solo mining and offers limited transaction throughput and scalability. The parallel mining approaches are used to improve blockchain scalability by concurrent mining of multiple blocks without changing the basic structure of the blockchain.

4.5. Redesigning blockchain:

Through our research, we have identified few approaches which are focused on designing an efficient strategy such as a novel consensus framework to address various aspects of scalability. On the other hand, there exists some approaches that presents an alternative scalable DLT solution (other than blockchain), such as Graphchain [31] or HashGraph [32] which uses Directed Acyclic Graphs (DAG) for a non-linear generation of blocks. However these approaches are not in the scope of current study.

5. Existing Surveys

This section presents a critical insight into existing surveys reviewing state of the art within blockchain scalability. In order to perform a rigorous study of such efforts, we conducted a methodical effort to identify and review existing literature through digital libraries such as IEEE Xplore, Google Scholar, and Elsevier. More details of our method to identify relevant literature is presented in section 6. To identify existing surveys, we have focused at studies which review existing efforts to achieve scalable blockchain solutions presented in Table 2. Through our research, we have found some studies such as [34],[35],[36] and [37] which are generic surveys into blockchain and only make reference to scalability as a significant challenge for blockchain. These studies do not particularly explore or address blockchain scalability and therefore have been rendered out of scope of current survey.

Among existing surveys on scalable blockchains, a number of existing surveys are focused on studying different consensus protocols and their impact on blockchain scalability. For instance, Vukoli [11], Mingxiao et al. [15] and Wang et al. [16] are focused at a comparative study of blockchain consensus protocols from the perspective of scalability. These studies consider commonly used consensus protocols such as Proof-of-Work (PoW), Byzantine Fault Tolerance (BFT) and Proof of Stake (PoS) and highlight significance of parameters such as TPS, node and client scalability and other issues such as consensus finality and node identity management. Similarly, [14] has focused on PoW and the impact of block size on the scalability of this algorithm. Although these are early efforts to study existing work within blockchain scalability, these are limited in scope to study role of consensus protocols to achieve scalable blockchains. Furthermore, with the growth of blockchains a number of alternative consensus protocols have emerged since this study which requires a current attempt to review the literature.

Another notable contribution is by Kim et al. [17] which primarily focused on vertical scalability issues including low TPS, increased blockchain size, e-signature size highlighting challenges and identifying potential solutions. The solutions discussed are classified into five categories, i.e. on-chain, off-chain, side chain, child chain, and

Table 2: Comparative summary of existing surveys with this study

Authors	Survey Type	Year	Publisher	Paper coverage		Scalability Coverage						
						Horizontal Scalability		Vertical Scalability				
				From	To	No. of Clients	No. of Nodes	Throughput	Storage (block size)	Latency	Block Gen. Rate	Consensus Efficiency
Marko Vukoli[11]	Traditional	2016	IEEE	2013	2016	✓	✓	-	-	✓	-	✓
Mingxiao et al. [15]	Traditional	2017	IEEE	2015	2016	-	✓	✓	-	-	✓	✓
Kim et al.[17]	Traditional	2018	IEEE	2014	2017	-	-	✓	✓	-	-	-
Rui Wang[16]	Traditional	2019	Springer	2016	2019	-	-	✓	-	-	-	✓
Fournier & Petrillo [14]	SLR	2018	ACM	2016	2018	-	✓	✓	✓	-	✓	✓
Xie et al. [33]	Traditional	2019	IEEE	2015	2019	-	-	✓	✓	✓	✓	-
Hafid et al. [13]	Traditional	2020	IEEE	2014	2020	-	-	✓	✓	✓	-	-
Zhao et al. [12]	Traditional	2020	IEEE	2014	2019	-	-	✓	✓	✓	✓	✓

inter chain. The authors further elaborated the comparison of approaches of each category on three major aspects including throughput, cost capacity.

More recently, Xie et al [33], Hafid et al. [13], and Zhou et al. [12] have focused at reviewing state of the art within blockchain scalability. As highlighted in Table 2, these efforts have focused on the vertical scalability factors i.e. throughput, latency and block size. For instance, Xie et al. analyses distributed storage approaches including DHT, Desema, Inter Planetary file system(IPFS), and BigchainDB which improve storage scalability. Hafid et al. [13] present on-chain and off-chain solutions as two layers of scalability solutions with the main focus on sharding approaches and its security analysis and effects of various aspects such as heterogeneous sharding, shard size in maintaining the security of the blockchain. Zhou et al. [12] present a review of horizontal scalability solutions focusing on two aspects of blockchain scalability i.e. throughput and latency.

In summary, these studies lack a systematic approach to conduct the review as well as limited depth in terms of existing efforts to address blockchain scalability, leading to gaps in terms of coverage of different dimensions of scalability, as well as depth and breadth of the surveys. Furthermore, the challenge of blockchain scalability has recently attracted renewed attention primarily due to increasing size of the Bitcoin chain and developments such as Segregated Witness (SegWit) [38]. Therefore, a current and comprehensive effort is required to conduct a systematic review of state-of-the-art with respect to blockchain scalability to highlight existing work, limitations and future research directions. To the best of our knowledge, this is the first systematic literature survey, based on Kitchenham et al. strategy[18], taking into account existing efforts covering all the aspects of scalability discussed in section 3. In the context of the comparative analysis of existing approaches presented in Table 2, our effort aims to address the shortcomings highlighted in this comparison.

6. Survey Method

In order to achieve a comprehensive and structured study of existing efforts to achieve scalable blockchains, we have adopted a systematic approach developed by Kitchenham et al. [18] to review and analyze existing literature in this domain. Fig 6 presents a graphical representation of the method we adopted whereas details of each phase within our study is explained in this section. As the first step to our study, we conducted a basic study of the prominent recent literature to achieve understanding of the problem domain and key contributors. Having established the suitability of the research, we devised our study into four phases i.e. selection, identification, screening & refinement, and compilation phases as shown in Fig 6 and further elaborated in Table 3.

6.1. Selection phase

In selection phase, we initially selected four scientific databases including IEEE Xplorer, Springer, Science Direct, and ACM to extract relevant papers. In order to identify relevant literature, we used specific keywords such as *Blockchain AND Scalability*, *transaction throughput*, *block generation rate*, and *blockchain storage scalability.* Furthermore, through our preliminary study, we identified early blockchain scalability research in 2015, hence, our study focused on research conducted between 2015 and 2020. Finally, we have included all academic publications (journal, conference and workshop papers) within our study as well as technical reports with proposals for new algorithms.

6.2. Identification phase

Through our initial search within scientific databases, we identified the number of relevant available literature to

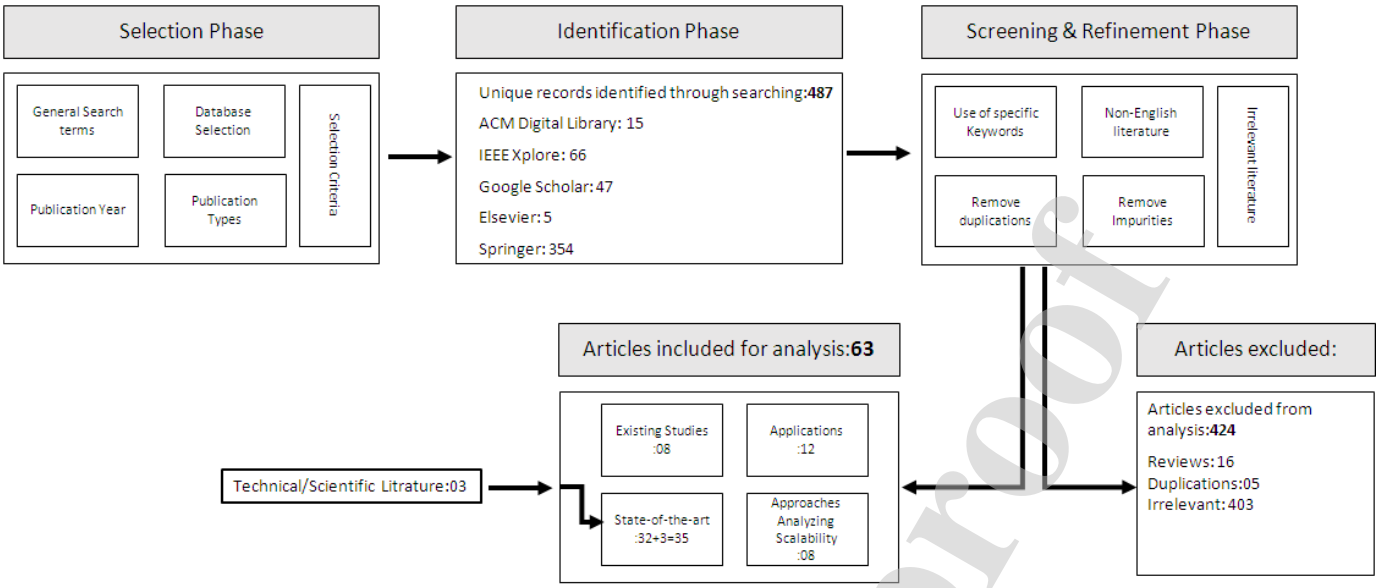


Figure 6: Method used to identify, gather and refine research articles

Table 3: Preliminary search results

Keywords	Type of Articles	IEEE Xplorer	Springer	ACM	Science Direct	Google Scholar
Blockchain Scalability Sub Keywords: Transaction Throughput, Block Size, Chain Size, Storage Scalability, Block Generation Rate, Latency.	Conferences/Chapters	46	85	-	-	47
	Journals	15	269	12	5	
	Early Access Articles	3	-	-	-	
	Magazines	1	-	-	-	
	Tutorials	-	-	2	-	
	Keynotes/Talks	-	-	1	-	
	Total	**66**	**354**	**15**	**5**	**47**
	Grand Total					**487**

be limited, leading to extension in our search to include Google Scholar to expand existing literature. However, literature in Google Scholar presents a significant number of duplicate records i.e. studies overlapping between Google Scholar and other scientific resources. Therefore, a conscious effort was made to remove all such instances in a way that if an article appeared twice is only counted against its publisher database.Consequently, a total of 487 unique research articles were identified across all databases with their breakdown presented in Fig 6.

6.3. Screening & refinement phase

Following the general search conducted in previous phase, we carried out a rigorous process of screening and refinement to filter the literature. One key criteria we used to perform this filtering was to include articles focusing on specific aspect of scalability as identified in section 3 such as transaction throughput, block size, chain size, storage scalability, block generation rate, and latency etc. Furthermore, we refined the articles to remove literature in languages other than English, to remove impurities and to exclude articles where definition of scalability falls out of scope in relation to our definition in section 3.

6.4. Compilation of results

Subsequently after applying the refinements, 63 articles were selected to be included in our study spanning across four categories as follows. The first category consists of 8 articles (Table No.2) that review one or more aspects of scalability. These reviews served as a basis for conducting this study and have been analyzed in section 5. Forward snowballing is also performed in this category of papers to extract more results for the last category. Second category, summarized in Table No.10 consists of the 12 articles which encompasses the applications/implementation of blockchain technology in a specific application domain to improve its scalability. The third category includes 08 articles summarized in Table No.11 which are focused at analysing scalability of blockchain with respect to different dimensions of scalability. Finally, the fourth category of consists of 35 articles summarized in Table 4, 5, 6, 7, 8, 9 which propose novel approaches and mechanisms to

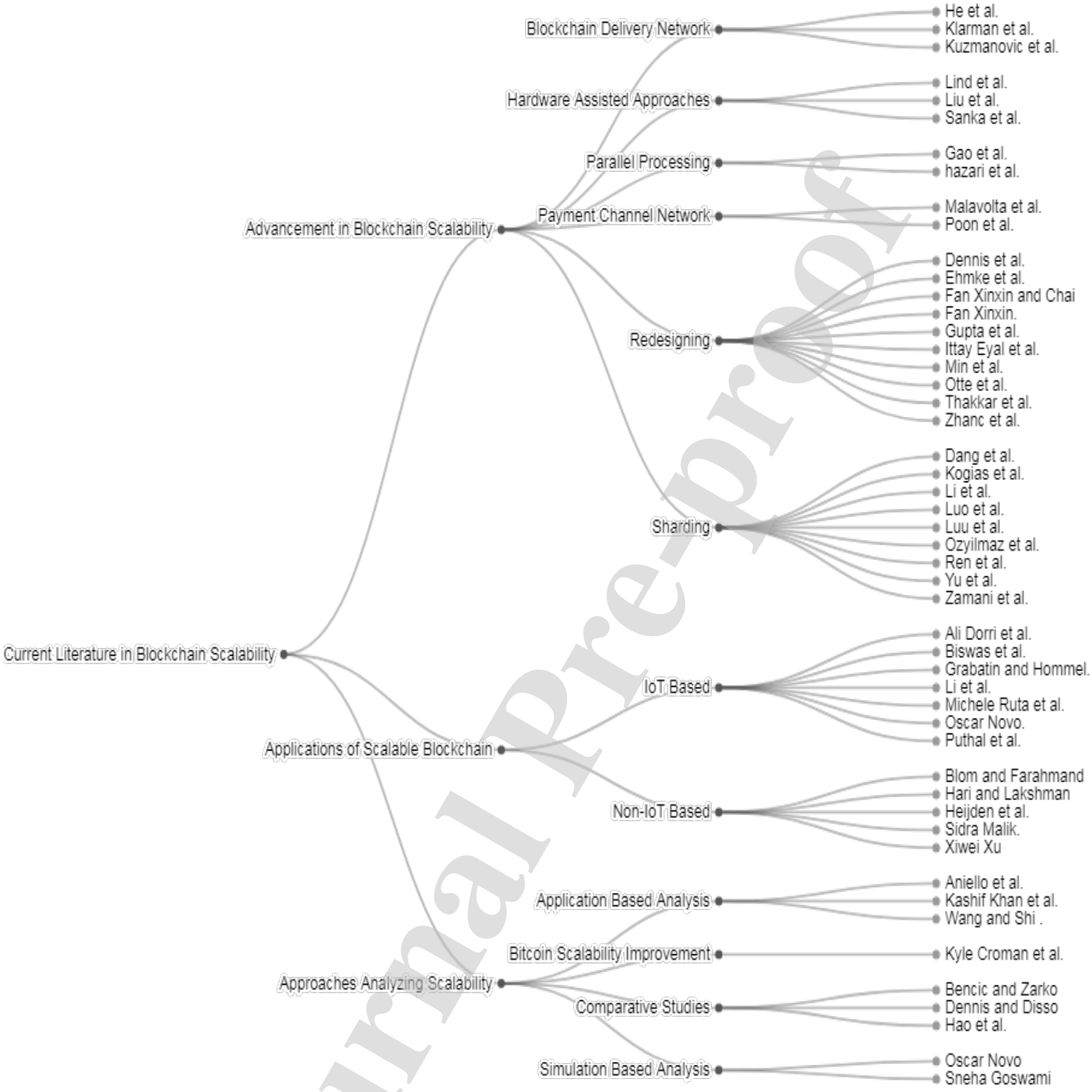


Figure 7: State of the art in scalable blockchains

advance the state-of-the-art with respect to scalability of blockchains.

The articles included in this study along with their categories are presented in Fig 7. In view of the different categories of articles identified through our research and in the interest of preserving focus and enhancing readability of the paper, we have organized the remaining paper as follows:

- Section 7 presents a critical review of approaches focused at advancing state-of-the-art with respect to scalable blockchains.
- Section 8 includes review of existing literature describing application of scalable blockchain within specific domains, and
- Section 9 analyzes literature focused at analysing

scalability of blockchains with respect to different dimensions.

7. Existing Literature to Advance Scalable Blockchains

With increased use of blockchains in diverse application domains, its scalability has emerged as a significant challenge. Consequently, a number of efforts have been made to address this challenge utilizing different approaches. In this section, we present a critical review of such approaches and organize them into different categories to aid readability and focus.

7.1. Payment channel networks

One of the most significant developments in improving blockchain scalability is the use of Payment Channel Networks (PCNs). In this respect, existing efforts can be classified into Lightning Network (LN) and Radian Network (RN).

With respect to LN approaches, Bitcoin Lightning Network [39] allows off-chain execution of the transaction to ensure micropayments via smart contracts without the requirement of trusted-parties reducing the transaction cost and counter-party risk. In particular, clients can rely on local consensus to perform a transaction with the on-chain commit delayed until either of the two clients wishes to do so. Protection mechanisms such as penalties and time-bound transactions are incorporated to mitigate against malpractice. Since the main chain stores opening and closure of the channel, the time taken for the opening of a channel is equivalent to traditional main chain transaction time. However, the fixed duration for a channel activity imposes recurring opening-closure channel problems and charging transaction fees every time two nodes want to commit the transaction after channel closure.

Malavolta et al. [40] is another LN-based effort, however, focusing on the security and privacy challenges within such approaches. The authors presented an improved version of PCN that maintains the scalability with security and privacy, especially focusing at wormhole attacks through the application of ESCDA signatures, one-way homomorphic functions and locking mechanism in a payment path. This is accomplished by using the Anonymous Multi-Hop Lock (AMHL) method that includes cryptography in the basic design of PCNs to enhance security and privacy. It also uses a Hash Time Lock Contract (HTLC) to mitigate with the security issues and requires a single signature per transaction that reduces overall transaction size also leading to several benefits including reduced communication overhead, transaction fees and memory requirements for closing a channel.

An alternative to LN, Radian Network (RN) [41] is used wihtin Ethereum to enhance its scalability. However, these solutions do not provide full-scale security and privacy. On the contrary, lightening approaches such as BOLT proposed by Green et al.[42, 43] and its implementations [44], [45], [46] provide anonymity but it necessitates cryptographic primitives which are only available in *Zcash* and is limited to single-hop transfers.

7.2. Sharding

Sharding is another promising off-chain approach widely used for blockchain scalability. Inspired by similar concepts in database management systems, it adopts a *divide and conquer* philosophy for parallel execution of transactions by sharing the workload amongst various shards to achieve improved throughput and latency with significantly reduced communication and storage overhead. The notion of the shard was first used in cryptocurrency by Danezis and Meiklejoh in 2015 [56] while devising a distributed cryptocurrency framework whose monitory supply is controlled by a central point of trust (such as a central bank) yet relying on other distributed authorities called *mintettes* within a shard to prevent double-spending attacks. However, this approach is still prone to double spending attacks due to the non-BFT nature of two-phase commit in each shard.

Luu et al presented Elastico[47], a first secure sharding protocol, that can tolerate up to 1/4 Byzantine adversaries and scales linearly with the number of miners. Unlike BFT consensus protocols, it does not require a quadratic number of messages as the network grows. This is accomplished by the formation of small parallel working committees which are assigned to a shard containing a unique set of transactions. A byzantine consensus protocol is executed within each committee to prevent double-spending. The management and controlling of these shards is the responsibility of a consensus committee. It also ensures the generation of random bit strings with a bounded bias to be used in subsequent epoch to prevent adversaries to gain a computational advantage by using the observations from a previous epoch. The experiments show an increase in throughput (40TPS) but with high latency of 800sec. This is due to the fact that Elastico requires all committees and their identities to be re-established at every epoch which not only increases latency but also results in a significant increase in communication overhead. Moreover, the small committee size would limit the overhead of BFT protocol but would significantly increase the failure probability up to 0.97 [48].

Kogias et al. [48] extended the work of Luu et al. through an approach *Omniledger*, which guarantees security, consensus finality with relatively high TPS, and reduction in storage up to 1/4x. It also increases shard size to 600 members and shows a significant decrease in latency from the order of hundreds to the order of tens. The Omniledger uses *byzcoinx*, a variant of byzcoin [57], to improve TPS by utilizing a more robust communication pattern that resembles a two-level tree to efficiently process transactions within a shard. It employs a shard prune mechanism by summarizing the state of a shard's ledger to reduce ledger storage and bootstrapping cost. Omniledger

Table 4: Payment channel networks to enhance blockchain scalability

Authors	Issue Addressed	Methodology	Solution Type	Bitcoin/ Blockchain	Implementation	Results	Limitations	Year
Poon and Dryja[39]	Throughput	Lighteninig Network	Off-chain	Bitcoin	-	-	Lacks providing full-scale privacy and security Recurring open-closure channel problem Problems in reward mechanism	2015 (adapted in 2017 via segwit)
Malavolta et al. [40]	Throughput	LN with Security and Privacy mechanism	Off-chain	Bitcoin/ Ethereum	Python, Charm (Library), ECDSA over the elliptic curve secp256k1	Communication Overhead < 500 bytes with time<100ms	-	2019
Raiden Network [41]	Throughput	Raiden Network	Off-chain	Ethereum	-	-	-	-

Table 5: Sharding-based efforts to enhance blockchain scalability

Authors	Issue Addressed	Methodology	Solution Type	Bitcoin/ Blockchain	Implementation	Results	Limitations	Year
Luu et al.[47]	Throughput	Sharding (Elastico)	Off-Chain	Bitcoin	Amazon Elastic Compute Cloud (Amazon EC2)	Throughput 40TPS, Latency 800s with n=1600	Low Byzantine resiliency (t=1/4) Latency scales linearly	2016
Kogias et al. [48]	Throughput, Latency, Storage	Sharding (OmniLedger)	Off-chain	Bitcoin	-	Throughput up to 4000TPS, Latency 63s.	Time-consuming bootstrap process. DoS Attacks Latency is still 63s	2018
Zamani et al. [49]	Throughput, Latency, Storage	Sharding (RapidChain)	Off-Chain	Bitcoin/ Cryptocurrency	GO, 32 x Intel Xeon Phi 7210	Throughput Up to 7380, Latency 8.7s	-	2018
Dang et al. [50]	Throughput	Sharding + TEE	Off-Chain	Blockchain Permissioned-BFT	C, Intel SGX(TEE)	Throughput of 3000TPS	Weak Two-Phase locking mechanism	2019
Ozyilmaz et al.[51]	Throughput	Sharding	Off-Chain	Bitcoin	Theoretical Comparison with Segwit and Bitcoin-NG	-	-	2018
Ren et al.[52]	Scaleout Blockchain	Spontaneous Sharding	Off-Chain	Cryptocurrency	-	-	No Support for Smart Contract. Forks	2018
Kan Luo et al.[53]	Throughput	Sharding Multi- Chain Architecture	Off-chain	Blockchain	-	TPS B/W 899.81 to 1520.56	Less Secure- no encryption and access control mechanism in place	2018
Yu et al.[54]	Performance/ Throughput	Multi Chain Sharding Architecture	Off-chain	Blockchain	Not Implemented (Framework only	-	Not tested	2018
Wenting Li et al.[55]	Node Scalability	Sharding	Off-Chain	Blockchain-Permissioned	Hyperledger Fabric v.6	Results not shown	Lack of load balancing Mechanism	2017

favors security over scalability by choosing large shards at the cost of higher latency but guarantees the finality. The experimental results show that with 1800 hosts a TPS of approx. 4000 is achieved. However, as illustrated by [49], this is achieved at the cost of more bandwidth requirement per node which sometimes crosses 45Mbps. Furthermore, although the basic design is to achieve concurrency with encouraging results in initial rounds however further increase in concurrency results in performance drop due to the constant number of cryptocurrency operations which add overhead in the parallel consensus. Moreover, this approach can use a trust-but-verify methodology to tolerate up to n/4 failures to achieve low latency for micropayments, but with high risk for regular payments, especially due to the absence of liability in the decentralized ledgers. It also requires active participation of users for inter-shard transaction which is hostile especially for lightweight users.

Zamani et al. [49] presented a sharding framework Rapidchain, with further improvements in latency and throughput. The model has improved security, without the requirement of a trusted setup as in Omniledger and Elastico, with the ability to tolerate 1/3 adversaries in a shard containing 250 members and further offers 1/9x reduction in the amount of storage required by each participant after the same number of transactions. It ensures robustness and reduces communication overhead by using a novel gossiping protocol *block pipelining* with a cross shed communication mechanism inspired by methodology devised by Maymounkov and Mazières [58]. Additionally it requires *O(nn)* messages to bootstrap itself and further uses a similar approach by dividing miners into committees and parallelization of consensus and storage to achieve

high throughput. It also offers sub-linear communication which is in the order of O(n) bits overhead per transaction to achieve speedy committee consensus.

The sharding solutions discussed above including Omniledger, Rapidchains, Elastico are limited to permissionless Cryptocurrencies by using unspent transaction output (UTXO) with main focus on security. Dang et al. [50] extended their solution to BFT-based permissioned blockchain mainly focusing on performance to achieve throughput similar to that achieved by VISA for the non-cryptocurrency workload. Since the BFT protocols have poor node scalability, the methodology relies upon optimized trusted execution environment TEE (Intel SGX) to eliminate equivocation in the Byzantine failure model to achieve higher fault tolerance and lower communication overhead. The formation and assignment of nodes to shards with security is done by implementing a trusted randomness beacon inside TEE to generate unbiased random values in distributed settings and the fault tolerance of TEE is utilized to reduce shard size which means higher throughput per shard. Secure distributed transaction is achieved by using a two-phase locking mechanism and a two-phase commit. Compared with UTXO supported Omniledger, Elastico, Rapidchain with the general workload, the evaluation conducted by authors showed that a transaction rate of over 3000 TPS is achieved on Google cloud. However, the two-phase locking mechanism may not extract enough concurrency from workload because of the batching nature of blockchain. Therefore, optimization in concurrency control protocol is necessary.

Ozyilmaz et al. [51] focused on increasing transaction throughput and block generation rate (BGR) in a Bitcoin network without degrading network decentralization. The framework uses the concept of parallel mining by splitting the unspent transaction output (UTXO) space into a tree structure to increase transaction throughput. It also introduces the concept of *half-node* which can independently verify the transaction on the sub chain it is tracking. The core of this model is splitting the main chain into independently operable sub-chains for creating multiple blocks thereby providing the flexibility of nodes having less bandwidth and storage to track only a subset of the chain without losing any verification capability. Miners must create a block for each sub-chain, and a separate block containing these block headers to claim a block reward. The model uses existing Bitcoin mining approach in a multi-chain setup. However, to verify newly added blocks and detect double-spending attacks, a separate chain called *eigen-chain* is added that keeps track of all sub chain blocks by storing header hash. Although the model is a framework and not evaluated in a real environment, a theoretical comparison with other approaches including Segwit and Bitcoin-NG reveals that the approach scales exponentially with split counts. Furthermore, in order to operate as a half node, it requires to store only one sub-chain and an eigenchain.

Ren et al. [52] presented a concept of *spontaneous sharding* to achieve the scale-out throughput and to resolve the problem of double-spending without maintaining a complete set of transactions at all nodes. Spontaneous sharding comes with an assumption that all nodes are motivated and behave rationally. Each node strives to minimize its transmission and storage cost by reducing the proof size and the number of recorded transactions. This minimum cost will be used for the provision of the authenticity of the value in the transaction. However, limiting the number of attempts to recollect the proof in case of failure will prevent malicious nodes to jam the network by spamming invalid proofs. The authors also acknowledge several issues that need to be addressed, including the storage cost per transaction which will not be scaled out since the sharding process may require recording the whole transaction set before sharding takes place. Furthermore, the issues of discrimination and hidden forks may arise due to varying proof size of transactions if a node refuses the values with huge proof size.

Kan et al.[53] and Yu et al. [54] foucsed on Multi-chain approach to sharding. Specificlaly, Kan et al. presented a layered architecture for heterogeneous inter-blockchain communication that provides routing management, message transfer with three-phase commit and escrow transfer to provide atomicity and consistency. Each blockchain system within the multi-chain architecture is considered as a shard to increase the throughput of the system. The core of this architecture is the router nodes, which maintain the details of different blockchains with the help of a unified transaction format for a standard cross-chain transaction, route information between multiple blockchains. The experiments were conducted to evaluate the performance, which shows that the TPS is increased with an increasing number of shards; this is obvious since each blockchain system within the model performs transactions independently. The throughput also decreases with an increase in the number of cross-chain transactions due to the three-phase commit for inter-chain communication that is time-consuming and consequently reduces overall TPS thereby necessitating optimization in confirming process of cross-chain transactions.

Yu et al [54] focus on heterogeneous multi-chain sharding architecture where services are assigned to various sub-chains with sharding to enhance the scalability of blockchain for large scale business applications. The multi-chain architecture is based on various sub-chains controlled by a single main chain and a layer known as the *value swab layer* which enables interaction between them. This architecture allows the parallel functioning of multiple blockchains, which may belong to multiple organizations, managed by the main chain. The sub-chain downloads the relevant information from the main chain and run independently, the outcome of this chain then again transferred to the main chain. The sub-chains may be a part of different organizations/businesses with no restrictions on the use of a particular consensus algorithm. Although the flexible sharding architecture improves the scalability but it can

increase the workload for the main chain by assigning the task and receiving the results from multiple sub-chains.

Li et al. [55] focused on node scalability and presented a sharding architecture for multiple independent but interconnected sidechains to meet industrial standards. The architecture is based on a notion of satellite chains that contain multiple private (permissioned) ledgers running different consensus protocols and independent auditors and regulators. The regulators can enforce policies without participating in the consensus by using smart contracts. The auditor can passively see and audit the transaction whereas a node can join multiple satellite chains simultaneously. To form a satellite chain, all the interested nodes must agree on validators, consensus mechanism, and access policies. The satellite chain can also transfer assets to other satellite chains. However, the issue of load balancing is not addressed in the paper. As in a case where all or most of the nodes join the same satellite chain may result in overloading of that shard or satellite chain.

7.3. Blockchain delivery network

In addition to sharding and LN-based approaches, recent efforts have focused on advancements in blockchain delivery network to achieve scalable blockchains. In this context, Klarman et al. [59] focused on cryptocurrencies and propose a blockchain distribution network (BDN) framework *bloXroute* which leverages the power of cut-through routing and system-wide caching to reduce propagation delays and improve transmission rate which consequently increases the throughput. The authors proposed a mechanism to propagate the block to all its gateways without prior transaction knowledge which can further propagate the blocks through peers in P2P Network to avoid any discrimination. The bloxroute is an effective broadcast primitive to support thousands of transactions per second. It also supports multiple blockchains simultaneously and ensures neutrality by allowing users to directly audit the services of bloXroute. However, the methodology is based on the concept of fast propagation without information of blocks which may allow malicious entities to propagate blocks.

Among other approahces, He et al. [60] and Kuzmanovic et al [61] focus on general blockchain-based scenarios. In this respect, Guobiao He et al. [60] presented an architecture that improves various aspects including resource utilization through QoS, improved BFT algorithm for security and a transaction level load balancing to improve the throughput of the system. Moreover, it uses a collaborative mechanism that interacts with cloud storage to solve the problem of storage scalability. The model is a four-layered architecture in which the first two layers contribute to improving the scalability of the architecture. In particular, control and authentication layer performs transaction-level load balancing to improve QoS whereas the cloud storage layer stores previous transaction blocks to improve storage burden on consensus node as it only stores one epoch of transaction block. However, load balancing overhead where one area share transaction with other area needs to be optimized.

Traditional blockchain networks such as Bitcoin, consist of trustless peer-to-peer network model with scalability challenge in terms of transaction per second. Inspired by cloud delivery networks such as Youtube, Kuzmanovic et al [61] proposed a framework, to optimize the transport layer to implement blockchain in a cloud delivery network without disrupting its decentralized nature to enhance throughput. The cloud distribution network can increase the performance by using a global infrastructure to support distributed blockchain neutrally. This is done by reversing the direction of trust and placing trust in a subset of nodes. The nodes can propagate the block on to the P2P network directly and can communicate it to other geographically apart nodes via Blockchain Distribution Network (BDN) that can be audited by other nodes in the network. The BDN helps to scale in various ways including transaction caching by distributing the transactions in the cloud and utilize indexes instead of a complete transaction. It also uses cut-through routing which propagates received bytes and does not wait for the complete block to be received, this will increase transmission speed especially for larger blocks. It can also solve the transaction in-cast problem in which transaction received at high rates is propagated via a single BDN server. However, the proficiency of design is unable to distinguish between distinct nodes, blocks, and transactions. Its generic nature is not able to identify and prevent malicious behavior and dishonest nodes. Thus the design heavily relies on trusted network infrastructure.

7.4. Hardware-assisted approaches

There are various approaches including [62][63] and [64] which depend upon trusted hardware to improve the chain performance. Among these approahces, Liu et al. [62] and Sanka & Cheung [63] used an on-chain approach whereas Lind et al. [64] and Das et al. [65] present an off-chain approach to scalability.

The BFT-based implementations in Blockchain can achieve higher transaction rates as compared to the traditional Bitcoin's network but it can scale to few tens of nodes due to the requirement of O(n2) message-exchanges to attain consensus on single operation among n servers. Liu et al [62] addressed the issue of node scalability with a fast-BFT protocol and compared it with other variants of BFT protocols including zyzzyva, MinBFT CheapBFT. The core of the Fast-BFT is a novel message aggregation approach, with reduced message complexity from o(n2) to O(n), that combines hardware-based trusted execution environments with lightweight secret sharing. The message aggregation in FastBFT does not require any public-key operations (e.g., multi-signatures), thus incurring considerably lower computation/communication overhead. FastBFT further balances computation and communication load by arranging nodes in a tree topology, so that inter-server communi-

Table 6: Blockchain delivery networks to enhance scalability

Authors	Issue Addressed	Methodology	Solution Type	Bitcoin/ Blockchain	Implementation	Results	Limitations	Year
Uri Klarman et al. [59]	Throughput	Blockchain Delivery Network	On-chain	Cryptocurrency	Framework	N/A	N/A	2018
He et al.[60]	Throughput and Storage Scalability-Chain Size	Cloud storage management with load balancing for efficiency	On-chain	Blockchain	Matlab (Simulation)	Increased performance.	Lack of optimized load balancing mechanism. Single point of failure	2018
Aleksandar Kuzmanovic [61]	Throughput	Trusted Cloud Delivery Network	Indexing in the cloud network and cut-through routing.	Blockchain	Framework only	N/A	Lacks a mechanism to identify dishonest nodes and relies heavily on Trusted network	2019

Table 7: Hardware-assisted approaches to enhance blockchain scalability

Authors /ID	Issue Addressed	Methodology	Solution Type	Bitcoin/ Blockchain	Implementation	Results	Limitations	Year
Liu et al. [62]	Consensus efficiency, Node Scalability, Latency, Throughput	A hardware- Assisted variant of BFT (FastBFT)	On-Chain	Blockchain	Golang, Intel SGX(TEE)	Peak Throughput of 100,000 TPS	Slight decline in TPS when the No. of nodes increases.	2019
Sanka and Cheung [63]	Throughput, Reducing the load on BC Servers, Latency	Caching system for BC Scalability and Throughput Improvement.	On-chain	Bitcoin	C, VB, ML-605 Virtex 6 FPGA, Jannson, Curl Library, SHA-256.	103% increase in performance when cache hits.	Adds centralization	2018
Lind et al. [64]	Throughput, Latency	Improved Version of Bitcoin-NG. (Teechain)	Off-chain	Bitcoin	Intel SGX (TEE), C++, secp256k1 (Library)	Throughput up to 1M TPS	Increasing members in the committee chain increases latency.	2017
Das et al. [65]	Block Validation Time	TEE +SC	Off-chain	Bitcoin	Intel SGX Python	-	SPOF	2019

cation and message aggregation takes place along edges of the tree. FastBFT adopts the optimistic BFT paradigm that only requires a subset of nodes to actively run the protocol. Finally, a simple failure detection mechanism is used that makes it possible for FastBFT to deal with non-primary faults efficiently.

Sanka and Cheung [63] presented a hardware-based implementation of a caching system for Bitcoin's blockchain core using FPGA(Field Programmable gate array) with customized SHA-256 hash to reduce the load on blockchain servers and improve the scalability and throughput. The evaluation results showed an increase in throughput up to 103x when cache hits.

Using an off-chain method, Lind et al. [64] presented Teechain, a variant of Bitcoin-NG with improved throughput and latency. Teechain is based on off-chain replication with threshold secret sharing and uses a trusted execution environment, a hardware security feature in modern CPUs, as a root of trust to asynchronously process off-chain transactions. The Teechain can handle 1 million Bitcoin TPS in a 30 machine deployment which is 33x higher than the lightning network approach.

To execute reactive multi-round contracts in a cryptocurrency environment, Das et al. presented a model named FASTKITTEN [65], to allow execution of arbitrary complex SC off the chain using TEE within bitcoin/cryptocurrency environment at low cost. This prevents the SC to be executed by all users. It also increases off-chain SC execution efficiency by incentivizing honest behavior. The model is implemented using Intel SGX hardware, the result showed a block validation time of 5seconds. FASTKITTEN supports batch execution of private user inputs that are not supported by both bitcoin and ethereum to provide faster execution of SCs with low computation cost. However, this model can run on a single TEE aka operator for the execution of SCs which adds centralization.

7.5. Parallel processing

One of the inherent limitations of blockchain is that the traditional Proof of Work (PoW) algorithm, based on solo mining, offers limited transaction processing speed and scalability. Hazari et al [66] presented a parallel PoW approach that addresses the problem of low throughput by leveraging the benefits of parallel mining in order to accelerate the process of the traditional PoW approach without changing the basic structure of the PoW process. This is achieved by electing a manager node at every epoch to administer and control the miners within a particular block, with the target that no two miners can be assigned the same work. This approach also has a reward system and prevents the formation of forks. The approach prevents a miner with a high processing power for solving all the blocks by restricting them to act as a manager of a subsequent block to receive his reward.

Table 8: Parallel processing approaches to enhance blockchain scalability

Authors /ID	Issue Addressed	Methodology	Solution Type	Bitcoin/ Blockchain	Implementation	Results	Problems/ Issues	Year
Hazari et al. [66]	Throughput	Parallel Mining on traditional PoW	On-Chain	Bitcoin	Go, GX Golang, Docker Container	34% increase in TPS	Biased with miners having low computational power. Delay in case of manager fails.	2019
Gao et al. [67]	Throughput	Parallel running of Smart contract	On-Chain	Blockchain-Permissioned (Pow Based)	Lingo, GUROBI/SCIP	-	Requires honest nodes with at least 50% of the computational power. The efficiency of ILP decrease with the increase in the number of smart contracts	2017

However, the approach is still biased with the miners having low processing power since those who have high processing power will quickly solve more blocks and have more chances to become a manager in subsequent blocks; thereby getting more rewards. Another issue is that the approach is heavily dependent upon managers to solve the current epoch. In case of failure of a manager node, a miner can still create transaction hash and nonce, but with traditional solo fashion which will consequently lead to an increase in mining time. However, the subsequent blocks will again follow the parallel strategy as the manager of the current block has been decided in the previous block. Furthermore, the benefits of parallel mining can only be achieved in the presence of multiple active miners in an epoch otherwise the approach will perform like a traditional solo PoW approach.

Gao et al. [67] focused on increasing throughput by simultaneous execution of smart contracts. Since most distributed ledger systems process smart contracts in a sequential manner therefore all users have to run smart contracts before its results can be accepted by the system limiting the system capable of handling a large number of smart contracts. If the smart contracts are more sophisticated then it can prevent other smart contracts from execution consuming all the computational resources. The authors addressed this issue by executing multiple smart contracts simultaneously which also increase overall throughput. This is accomplished by first dividing the smart contracts into subparts and a Sub Committee (SCom) which contains a group of users and then allowing only a committee to execute the particular smart contract. Consequently, multiple Subcommittees can be executed simultaneously. The security issue is resolved by allowing only those public keys of the user which recently involved in block generation and employing the majority voting criteria to prevent the users from controlling a SCom. The fairness in terms of incentives and workload of the users is ensured by using an ILP (integrated linear Programming.) solver, but the short term fairness depends upon the even distribution (load balancing) of smart contracts on subcommittees.

7.6. Redesigning

In addition to the approaches discussed earlier, a considerable number of efforts have focused on achieving scalable blockchains by strengthening its core components such as the consensus protocols, data storage and management capabilities etc. We present a summary of such approaches below.

Ittayeyal et al. [68] proposed Bitcoin-NG, a variant of Bitcoin consensus algorithm where the leader is selected for every epoch allowed to serialize state machine transaction until the next leader election. Bitcoin-NG's bandwidth is limited only by the capacity of individual nodes (as higher throughput does not generate key block forks) and its latency is limited by network's propagation delay. The leader election is forward-looking which ensures continuous transaction processing in contrast to Bitcoin where the leader is in-charge of serializing only history. The protocol generates two types of blocks: key blocks for choosing a leader and micro blocks for ledger entries. Due to the frequent generation of microblocks, the short fork may occur on almost every leader switch which is resolved by poison transaction that contains first pruned block as proof of fraud invalidating the fee sent to fraudulent leader generating the fork. Although evaluation results show improvements in fairness and latency, key block forks do occur which can be resolved by the heaviest chain policy. Secondly, varying block sizes increases transaction frequency but increases block verification and propagation time. The Bitcoin-NG is still a PoW concept and is susceptible to selfish mining if the attackers are more than 25% of the whole network. Moreover, there are some issues in incentive policy highlighted by proposed by Yin et al. [82] including that the leader can enhance its revenue up to 100% of the fee by trying to mine the key block himself. Furthermore, a malicious elected leader may perform a DoS attack by not publishing micro-blocks.

Zang Wang [83] mainly addressed the issue of fast convergence across a large number of nodes by using a voter model and the Ising-Glauber model. The Majority vOting Consensus Algorithm (MOCA) is based on cellular automata which is a rule-based computational machine to perform an action. It employs a probability theory approach i.e. voter model which improves the consensus scalability by allowing adjustment of a voter's opinion based

Table 9: Existing efforts to redesign blockchain to enhance scalability

Authors /ID	Issue Addressed	Methodology	Solution Type	Bitcoin/ Blockchain	Implementation	Results	Limitations	Year
IttayEyal et al. [68]	Throughput	Leader Election and formation of key and micro blocks. (Bitcoin-NG)	On-Chain	Bitcoin	-	-	Selfish Mining. DoS attack Weak incentive policy	2016
Zhancang Wang [69]	Consensus Efficiency (Node Scalability)	Majority Voting cellular automata (MOCA).	On-Chain	Blockchain	-	With 1M nodes, the protocol converges in few iterations	The assumption that in 1M nodes, each node has only 10 neighbors The presence of adversaries can affect convergence.	2018
Xinxin Fan and Qi Chai [70]	Node Scalability	A variant of DPoS with voting and epoch system with short-lived ESCADA and BLS signatures to add decentralization and extensibility for IoT (Roll-DPoS)	Off-chain	Blockchain for IoT	Framework only	-	Prevents resource constraint devices to actively participate in the consensus process.	2018
Dennis et al. [71]	Storage (Chain Size)	Temporal Rolling (pruning)	On-Chain	Bitcoin	Framework implemented in B Language	Constant 4.5GB chain Size	Delays in transactions during periodic rolling	2016
Min et al. [72]	Throughput	Dynamic Block Size	On-Chain	Blockchain for e-commerce application	10 Intel Core i5 Systems,1000 Nodes	1Million TPS in a trusted trading network, Block Size in KBs, BGR= 10,000 Micro-BPS (1BPM)	Requires all nodes to be trustworthy in the network	2016
Otte et al. [73]	Throughput	Focuses on the agent's trust (Trustchain)	On-Chain	Blockchain (permissionless)	Python	Throughput up to 210TPS	Performance deteriorates upon an increase in database size	2017
Xinxin Fan. [74]	Consensus Efficiency (Node Scalability)	A variant of PBFT with Fast Signature Verification Process-Less communication and computational overhead.	On-Chain	Blockchain	C, Libraries (secp112r1, sect163k1,MNTcurve), ECDSA (Elliptic Curve Digital Signature Algorithm), BLS(Boneh–Lynn–Shacham)	ESCDA is 10 times faster but BLS has 50% less overhead	-	2018
Ehmke et al. [75]	Storage scalability (Chain Size)	Keeping the global state of the system in the current block. (Proof-of-Property)	On-Chain	Blockchain	Not Implemented (Framework only)	NA	Maintains only one state i-e balance The requirement of an efficient key naming mechanism Forks	2018
Parth Thakkar [76]	Throughput and Latency	Optimization in Hyperledger Fabric	On-Chain	Blockchain	Hyperledger Fabric, Kafka ordering service for consensus	Throughput=2250, Latency =30ms	The network may become a bottleneck with n nodes at the diverse geographical location	2018
Gupta and Janakiram [77]	Block Generation Rate	Proof of Win- A knockout tournament for verification nodes. (Colosseum)	On-Chain	Blockchain	Vishwa-a grid Computing Middleware	Results not shown	No mechanism to avoid a Validator's biasness Forks	2019
Gang Wang and Zhijie Jerry Shi [78]	Throughput	Reducing intraplant communication from $O(n^2)$ to $O(n)$	On-chain	Blockchain	-	-	-	2019
Kalodner [79]	Throughput	Incentive-based VM implementation of SC	Off-chain	Blockchain	GO	-	Assumes that manager nodes behave rationally	2018
Teutsch & Reitwiebner [80]	Throughput	Decoupling mining and verification process with verification performed off the chain	Off-chain	Cryptocurrency	Google's Lanai interpreter	-	Relies on attractive incentive mechanism and assumes the presence of trusted verifiers	2017
Gorbunov & Reyzin [81]	Storage	Vector commitment to aggregate small proofs across multiple commitments	On-Chain	Cryptocurrency	Rust	Storage requirement for 108 SC = 4.5GB with 31KB/block overhead.	In a sharding cryptocurrency scenario, Frequent shard switching may introduce delays.	2020

on the opinion of their neighbors and Ising-Glauber model which is based on spin (listening to neighbors). The experiments show that with 1000,000 nodes where each node has only 10 neighbors, the system will converge in a few iterations. However, the state where every node converged to the correct state is not possible if the adversaries are more than the number of the neighbor of each node, since there is a chance that an honest node is surrounded by adversary neighborhoods.

The work of Xinxin Fan and Qi Chai [70] focuses on addressing the issue of sidechains scalability by allowing a significant number of nodes to actively participate in the consensus process for large-scale IoT systems. Traditional DPOS consensus limits block production to limited partially trusted nodes. The Roll DOPS consensus consists of epochs and sub epochs and uses an Ethereum based community voting process for selection of block producers. The potential block producer attracts community members to vote for him via a campaign website offering available resources and rewards. N block producers are randomly chosen through deterministic random bit generator (DRBG) for the first epoch. These block producers then generate short-lived epoch specific private key shares to sign the PBFT process with ECDSA and BLS threshold schemes with Pederson's DKG's approach to speed up the PBFT process. The nodes which are not selected as block producers for the root chain also has an option to be selected as block producers of side chains. This auto-scaling process enables a significant number of nodes to participate in the consensus process. Although, the approach invites nodes to become block producers for root chain and side chains, but in IoT system majority of devices have low processing power and storage capacity, which may prevent the vast majority of devices to remain idle and not actively participate in the consensus process.

The peer to peer nature of the blockchain requires each client to download and maintain blockchain up to date therefore resource constrained devices cannot fully participate except for creating transactions. Dennis et al. [71] focused at addressing this challenge by allowing the data to be stored for only a preset period and empty blocks known as *spent blocks* which can be used as input to new blocks will automatically be deleted. Consequently, a mining node is no longer required to maintain the history of blockchain from the genesis block. All the miners, on a pre-set time every day, will add a checkpoint and remove all the blocks older than 30 days. The authors claim that this technique improves scalability by allowing the consistent size of 4.5GB for Bitcoin's blockchain irrespective of the network size, which is significantly smaller than its current size, without changing the properties of the traditional blockchain. However, the deletion process of spent blocks may add a delay in mining new transactions or halt the transaction at that time.

Min et al. [72] presented the adoption of Blockchain in e-commerce specifically addressing the scalability barrier in terms of throughput in comparison with Bitcoin's network. The authors proposed addressing this problem by using a permissioned Blockchain framework (PBF) to create a public, autonomous e-commerce ecosystem. PBF partitions the network into subcommittees by using a random partition algorithm which can also limit dishonest peers. Each subcommittee has linear membership in the total computational power and runs an inner consensus protocol to process a separate set of transactions and blocks. The special committee agrees on peer's blocks, consisting of a set of valid transactions. Another special committee is designated to write peer's blocks into a global block so that each peer can identify number of transactions or blocks belonging to each peer at any point in history. Peer Inner Blockchain Protocol (PIBP) is used to improve throughput and support instant transactions. E-commerce Blockchain architecture keeps unlimited transactions without sacrificing higher credibility.

Another approach, named Trustchain, presented by Otte et al. [73] is focused at redesigning a sybil-resistant blockchain that is scalable in terms of transaction throughput. The authors use an agent-based approach that mainly focuses on trust and ensures the contribution of every agent which has taken resources with a decent transaction throughput. This is accomplished with the agents that can create their own genesis blocks and works in parallel powered by a Sybil resistant algorithm NetFlow which ensures the contribution of every agent that has taken resources. To increase the resistance against tampering double-spending attacks, each block also refers to a block in the chain of the counterparty. This ensures that each block has two incoming and two outgoing pointers. Evaluation showed an improvement in TPS for first 1000 transactions however with the subsequent increase in capacity of database, the insertion overhead also increased. This is because each insertion necessitates a DB query for fetch evidence about the latest block of a particular agent, which slows down the transaction rate.

With the traditional PBFT, instant consensus finality can be achieved, but this is at the cost of limiting the number of nodes due to the broadcast nature of the three-phase protocol. Fan et al. [74] proposed an improved version of PBFT with short-lived cryptographic keys to speed up the consensus process. PBFT experiences computational overhead which slows down consensus process due to round-based broadcast nature in which all nodes verify multiple received signatures from their peers. The authors addressed this issue by implementing short-lived cryptographic keys and a BC-based distribution approach to periodically update those keys. The efficient signature verification process has a key role in improving the scalability and computational cost of PBFT. This is accomplished by implementing the digital signatures at medium and low-security levels with periodic short length key pair update which has less communication overhead that enables faster verification process to improve the node scalability without compromising the transaction throughput.

In traditional blockchains such as Bitcoin, in order to

participate in the formation of new blocks, a local copy of the global state of the system is required at the beginning. For instance, the space required to store Bitcoin's network is more than 150GB. Consequently, the participation of new members will be delayed. Ehmke et al. [75] presents a framework named Proof of Property (PoP) to address the storage issue by deriving the concept from Ethereum to keep the global state of the system in the current block and include the relevant state of the system in each transaction for efficient verification. Therefore reducing the space required to store the complete system. The proposed framework uses Markle Patricia Tree (MPT) and is based on the proof that in each transaction the input addresses owns enough coins to fulfill the stated transaction. Although the framework saves the chain storage space, the authors illustrated some of the issues that require further research, including the ability of PoP to keep only one state which is unspent transaction balance. Furthermore, the author also highlighted the importance of a missing mechanism to avoid the formation of forks especially in a situation where different clients could be in a different state. The drawback is that transactions already being processed might become invalid if the block they are included in, is in the part of the forked blockchain that will be dismissed.

Parth Thakkar [76] focuses on the optimization of various operations of hyperledger fabric to achieve a high throughput. These optimizations include crypto-operations, serial validation of transaction in a block and multiple REST API calls to CouchDB to achieve maximum block size with low latency. The optimization consists of MSP cache, which uses a hash map as a key to prevent the decentralization of serialized identify and an ARC algorithm for cache replacement. This reduces crypto-operation and call- stack-depth. The optimization in the Validation system chaincode (VSCC) is by introducing parallel VSCC validation of a block through a configurable number of worker threads per channel which validates 1 transaction. This results in a reduction in the latency of VSCC from 300 to 30 ms and TPS of 950 for the block size of 30. The bulk-read/write during MVCC validation commit for CouchDB allows to cut down the number of REST API calls. With all three optimizations combined a TPS of 2250 is achieved. These optimizations have also been adapted in Hyperledger Fabric v1.1. The authors admit that they did not consider the network as a bottleneck, however, the nodes dispersed in diverse geographical locations with different arrival rates may impact the overall transaction rates.

The Cash Fault Tolerance (CFT) or Byzantine Fault Tolerance (BFT) based consensus algorithms usually uses voting-based consensus in permissioned blockchain. They have a message complexity of O(N2) for validation of a transaction and limit the participation of a number of validator nodes. Since the immutability of blockchain is dependent on the phenomenon of excessive replication of data and therefore, limiting the number of nodes reduces the attack surface for an adversary trying to manipulate the ledger. Gupta et al. [77] proposed a framework to increase block generation rate by proposing a consensus algorithm for permissioned blockchains, Proof of win (PoWin), that is based on a knockout tournament between two players. The colosseum formulates a ring network and a two-step process. First, to find the validator node that can issue a PoWin certificate to a winner node in a knockout game. Second, a ring-based DHT keeper to store the results. The winning nodes propose a new block for blockchain and the keeper can verify matches This elimination process amongst two opponents reduces the message complexity and allows the protocol to scale. However, the protocol heavily relies on validator node to control, manage, and verify the knockout results. A biased validator can easily drop a tournament by foreseeing that an ally is losing. Moreover, the framework lacks a mechanism to manage forks i-e simultaneous block generation in the same round.

Furthermore, Wang et al. [78] presented a blockchain-based model for metering in industrial plants. The implemented approach guarantees security, trustworthiness, and immutability in metering systems via blockchain. It also deals with reducing complexity from O(n2) to O(n) by using a novel approach based on BFT. A two-layered hierarchical structure chronologically chained by a hash, an independent local chain for each industrial plant based on BFT protocol which prevents adversaries to change private ledger while enforcing crypto-primitives and threshold signatures to reduce communication messages within the individual plant to improve scalability. This is done by using a short signature scheme, Boneh-Lynnh-Shacham (BLS), based on Diffie-Hellman assumption on certain elliptic and hyper-elliptic curves which significantly reduces the communication complexity in a consensus process. Secondly, A cloud-based state chain is used to aggregate all the data blocks of individual chains in an epoch. However, the design relies on the efficiency and trustworthiness of cloud infrastructure, and the performance of individual chains will directly have an impact on the aggregated state chain.

Kalodner et al. [79] presented an incentive-based model named Arbitrum to execute SC off the chain to cope with the limitations including privacy and scalability of the Ethereum platform. On contrary to traditional systems, the authors propose implementing SC as a virtual machine with a set of managers who maintain the VM state with only one manager chosen to control the VM. The validator nodes can accept and track the hash of the state endorsed by all managers. An incentive/ penalty mechanism is employed by the verifiers, to cope with the disagreement among managers, by verifying the proof submitted by manager nodes. However, the irrational behavior of manager nodes may result in imposing an extra burden on verifier nodes.

Teutsch & ReitwieBner [80] presented a model (TrueBit) to securely execute SC as well as reducing redundant computations of network nodes used in Ethereum by outsourc-

ing computation with the incentive mechanism in place to motivate outsourcing parties and to guarantee the correctness of the solution. The model also consists of a network of trusted verification nodes, called judges, which has been occasionally forced to inspect the task via incentivized interactive verification game and submit incorrect solutions for further resolution of the occurrence of any dispute in the solution. The dissociation of verification protocol for mining nodes enables improvement in the transaction throughput without facing verifiers problem. However, the verification game introduces additional overhead on verifiers as well as requiring trusted verifiers to perform small computations to find the correctness in the solution. Moreover, it requires an attractive incentive mechanism for the verifiers to find the bugs in the solution. Furthermore, it also assumes that the attackers are rational with limited financial resources and cannot subvert the underlying consensus with at least one verifier present for checking each task.

Gorbunov et al. [81] presented an efficient vector commitment scheme for SC which is based on aggregating proofs across multiple commitments. It also improves propagation bandwidth for transmission of blocks by reducing the proof size. The small commitments and proofs generated by users in the point-proofs scheme allow a third party (block proposer) to aggregate multiple proofs concerning different commitments into a single proof. The scheme is especially focused on stateless block validation of SC for reducing its storage requirements. However, the model still requires 4.5GB of storage for 108 accounts which may lead to delays especially in the sharding scenario where frequent shard switching by a node requires downloading part of such a large state each time.

7.7. Summary of existing literature to advance scalable blockchain

This section presented the state-of-the-art approaches that are focused at enhancing different aspects of scalability in the blockchain. It is evident from the literature that due to the adoption of blockchain in diverse scenarios, research community has made considerable efforts to devise solutions to address scalability challenges including throughput, storage, latency, and node/network size. These approaches are either focused specifically for on/off-chain or can be applied in both scenarios. Off-chain approaches can be further categorized into PCNs and Sharding which improve scalability by committing the transactions off the chain. Further, approaches such as BDN and parallel processing propose on-chain improvements whereas hardware-assisted approaches and those redesigning the consensus/data-structure are applicable to both on and off chain solutions to enhance the scalability in various scenarios. Key observations derived from this section are:

- Sharding has emerged as one of the promising solutions to improve chain performance. The increase in the number of shards in the blockchain network scales the transaction throughput linearly without compromising the decentralized nature of the blockchain. These approaches can be applied to permissioned as well as permission-less scenarios. Many sharding solutions including Elastico [47] and Omniledger [48] rely on trusted setup and use UTXO with the main focus on security. They can tolerate up to n/4 malicious adversaries and have a decent transaction throughput but this is achieved on compromising latency. On the contrary, rapid-chain [49] has better throughput whilst significantly lowering the latency and storage. Moreover, it does not rely on trusted setup and can tolerate up to n/3 adversaries. Other sharding solutions, such as [50] used trusted hardware to improve performance under non-cryptocurrency load. However, sharding solutions also have various issues including low byzantine resiliency, increased latency, time-consuming bootstrap process, the formation of forks, security and efficient load balancing mechanism which need further research.

- In recent years, the concept of PCNs was adapted by both bitcoin and the etherum community as Lightning Network [39] and Raiden Network [41] respectively to improve the overall blockchain's throughput. The PCNs are more suitable in scenarios like a bank to bank, bank to customer, and customer to customer within a bank. However, these PCN based micropayment systems have several security and privacy issues which limit their widespread adaption in the diverse application domains. The solution like BOLT [42, 43, 44, 45, 46] provides anonymity but lacks cryptographic primitives for multi-hop transfers. [40] tries to improve the security by using cryptographic primitives but primarily focusing on wormhole attacks. Moreover, the concept of micropayments via PCNs necessitates both parties to always remain online for committing transactions that make them vulnerable to attacks. Furthermore, the prolonged downtime of any party may result in the expiration of the waiting time for channel closure. This may lead to freezing of funds or a situation where a party can fraudulently pocket the funds and close the channel while other is away. This requires a strong mechanism where a party can prove a fraudulent activity to impose a penalty on the other party and an efficient balance planning mechanism to initially predict and freeze a certain amount that is required by the channel.

- As blockchain overlays public network, an efficient network infrastructure is another key to blockchain scalability. The use of cut-through routing, such as proposed in [59], significantly improves the blockchain throughput. However, this requires a trusted setup

due to the inability of segregating legitimate and malicious traffic.

- One of the limitations in wider adoption of blockchain is the storage requirements to join the network. For instance, a node aiming to join the Bitcoin blockchain has to download the existing chain/ledger which currently is more than 260GB. In this context, cloud storage [60] can reduce the burden from mining nodes however it ads centralization thereby introducing single point of failure.

- The use of efficient and/or trusted hardware with higher processing power, storage, and memory can greatly improve the performance in terms of throughput and latency of blockchain [64], [63], [62]. However, its adaption in public blockchain requires an incentive mechanism to attract the miners to use resourceful hardware.

- Although parallel mining is proposed to increase throughput but the efficiency is directly proportional to the number of active minors in an epoch. Such approaches, for instance [66], require an efficient mechanism for managing the task-assignment to remove the bias against the miners with low processing power. Furthermore, the single managing entity in an epoch also increases the mining time in case of failure of a manager node.

- The blockchain that has message complexity in the order of O(n) is considered as scalable. Reduction in message propagation quantity and complexity has significant positive effects in improving the throughput of a network [78].

8. Applications of Scalable Blockchain

In addition to innovative methods to enhance scalability of blockchains, a number efforts in existing literature are focused at application-specific measures to achieve scalable decentralization applications using blockchains. These applications leverage fundamental properties of blockchains such as tamper-proof storage and decentralized architecture to achieve application function however adopt specific measures to improve the scalability required by the applications. For instance. Ruta et al. [84] presented a Service Oriented Architecture (SOA) that adds a semantic blockchain layer with the help of smart contracts for an efficient resource discovery process that includes registration, discovery, selection, and payments implemented in Hyperledger Iroha for IoT based CPS. Similarly, Dorri et al. [85] presented a centrally managed blockchain framework to optimize the energy consumption and distributed trust to increase BGR for a smart home IoT environment.

Oscar Novo et al. [86] also focused on the adoption of blockchain in IoT where scalability is a problem for managing access to billions of constrained devices in the IoT network and centralized access control systems lack the ability to deal with increased load efficiently. For IoT based on centralized models, single centralized access control server might become a bottleneck in blockchain when access control queries and updates are frequent. The approach uses a concept of management hub nodes to increase the latency of the overall IoT-based blockchain network.

Most identity and access management solutions are centralized and require metadata exchanges which is either federation-based or manual. The central server maintains all the metadata of Identity Providers (IdP) and Services Providers (SP) into a centralized database. The copy of this database with signatures, which are usually valid for a few days, is communicated to all Federation Members (FM). Consequently, requiring all FMs to query the server for the updated version of metadata frequently. If the validity of the metadata signature is set too long, this may leave the SPs with the outdated information. Grabatin et al. [87] addressed this challenge by implementing Hyperledger Fabric based decentralized blockchain technology in managing metadata which has an intrinsic feature of instant ledger updates and reliability. The authors achieved Hyperledger-based solution which contains a feature that only allows parties that are part of the transaction to download metadata, consequently improving overall storage efficiency. This is done through a two-step process, including provider registration via chaincode deployed by a federation and micro federation creation. After SP registration, the clients can issue proposals for the creation of micro federation which contains all the required SPs. If all the SPs approve this event, the metadata of all SPs are integrated into the peer's ledger of the network in the form of a byte-array that can only be downloaded by the SP belongs to the same micro federation. This consequently allows different communication endpoints for different federations and allows SP to use less storage and requires less computational power to process the metadata. Although the solution improves efficiency compared to traditional setup, the performance of the system depends upon appropriate configuration of block time (s) in Hyperledger and SP's reaction time (r).

Biswas et al. [96] mainly focused on the integration of blockchain with IoT. Traditional blockchains such as Bitcoin with a transaction rate of up to 7 TPS are primarily not designed for high transaction rates. However, IoTs connect billions of heterogeneous resource-constrained devices with requirement for exponentially higher transaction rate which is not supported by Bitcoin structure. The authors have filled this gap by designing an architecture that uses a concept of associating each IoT device with an organization via local registration process and a certification authority(CA)is used to authenticate each node to achieve better TPS and ledger scalability.

Li et al [89] presented a distributed storage for IoT

Table 10: Applications of scalable blockchains

Author(s)	Application Domain	Scalability Aspect	IoT	Publisher	Year
Michele Ruta et al. [84]	SOA for IoT	Memory usage per node, Processing time (Latency)	✓	RonPub, Lubeck, Germany	2017
Ali Dorri et al. [85]	Optimization of Blockchains	Block Validation, the Processing time	✓	ACM	2017
Oscar Novo [86]	Scalable Access Management IoT	Latency	✓	IEEE	2018
Grabatin and Hommel [87]	DLT in Identity Federations	Storage, Latency	✓	IEEE	2018
Sujit Biswas et al. [88]	simulation-based study to address the issues of privacy, reliability, and efficiency of payment processing in vehicles to the grid network	Throughput, confirmation Time & Storage Requirements	✓	IEEE	2019
Ruinian Li et al. [89]	Data Storage and Protection in IoT	Storage	✓	IEEE	2019
Deepak Puthal et al. [90]	Analysis (Testybed+simulation) of a proposed Consensus Algorithm (Proof of Authentication)	Time Taken for block validation & Latency	✓	IEEE	2019
Hari & Lakshman [91]	Blockchain for Internet Security	Block Size, TPS, Node scalability	-	ACM	2016
Xiwei Xu [92]	The Blockchain as a Software Connector	TPS, Storage	-	IEEE	2016
Heijden et al. [93]	Blockchain in VANET	hierarchical consensus	-	ACM	2017
Sidra Malik [94]	a sharded tiered network architecture to address scalability	Transaction Rate	-	IEEE	2018
Blom & Farahmand [95]	Blockchain for small electricity producers and consumers	Frequency of transaction(TPS) & No of participants	-	IEEE	2018

based systems with security and accountability to reduce the burden from the central cloud server. The scheme consists of edge computing which helps IoT devices performing complex cryptographic computations at network edges. It also helps in collecting and forwarding data to DHT storage. The blockchain authenticates the read-write requests to the DHT by working as a trusted third party. To ensure the protection of the credentials, certificate-less cryptography is used that consists of a Key Generation Center (KGC) to create a partial private key based on the identity of the user which in turn will be used by the user along with a secret value for the establishment of a private key. This also avoids the key escrow problem in identity-based encryption. The proposed model can be applied to various application domains including smart grids, smart metering, healthcare, etc. Data in this scenario can be traded by the owner. For instance, a user can sell ECG data to a research institution by posting a transaction via blockchain. However, the design heavily relies on the scalability (throughput) of the system to efficiently utilize a decentralized storage mechanism. Moreover, the authors assume that communication between the edge servers and IoT devices is secure. Furthermore, the scheme still has security issues as there is no mechanism to cope with majority takeover attacks.

Puthal et al. [90] presented a lightweight trust-based consensus mechanism Proof-of-Authentication (PoAh) for edge computing and resource-constrained IoT devices to reduce the latency up to 3 seconds which is in the order of 10 minutes in traditional PoW approach. The model uses EIGamal encryption for public and private key generation. Initially, the trusted miners are deployed as validators. The others can participate in mining with zero trust value, upon completion of every successful block, the trust value increases. A trusted mining node, based on asymmetric cryptography, authenticates a block signature via the public key of the source. It also checks the MAC address before broadcasting the block to the network with PoAh ID. The framework is implemented in a small scale environment using simulation in Python as well as on testbed using Raspberry Pi. Although the proposed mechanism reduces latency in small scenarios, its applicability in large scale scenarios is still required for the adaption of this mechanism in real-world applications.

Hari and Lakshman [91] presented a blockchain-based framework for securing the Internet BGP and DNS infrastructure. Currently, they both are dependent upon PKI which can be compromised by APTs (Advanced Persistent Threats), a tamper-resistant blockchain-based resource management framework prevents a single malicious

adversary to subvert the system. In this internet blockchain framework, the DNS management and registry entities like Ases act as peers with a private key associated with its internet blockchain address. Multiple genesis blocks are created within each registry that consists of transactions which contain the record of transfer of resource form its registry to internet entities. These records are then published into internet blockchain to provide an initial resource allocation state. These records are continue to be published but can not be revoked by a single entity. However, the blockchain has inherent scalability problems such as BGR and throughput which required to be addressed. Moreover, it also requires an incentive mechanism for global deployment and to enable a tamper-resistant internet.

Xu et al. [92] elucidate how the inherent properties of a blockchain can improve the software quality attributes including security, scalability, and sustainability by employing a distributed architecture as, an alternative to shared centralized data storage, in an untrusted environment. However, the validation (mining) process may increase the overall communication latency. It also requires an efficient strategy to decide which data is required to be stored on or off the chain due to the storage limitation of the current blockchain.

A Blockchain-based collaborative and transparent V2X communication and accountability mechanism for misbehavior detection and revocation without requiring a trust is proposed in [93]. The scalability is achieved through a hierarchical consensus in which clusters are created dynamically to provide consensus decisions to misbehavior authorities via Road Side Units (RSU), which are fixed and have known identities therefore can be employed with a BFT based consensus. However. The performance of BFT based consensus starts degrading upon an increase in the number of nodes. To resolve this issue, the BFT consensus is applied to smaller groups of RSUs that reside in the same geographical location. The Misbehaviour authorities use the decisions of these small clusters to maintain the blockchain state. However, the framework requires vehicles and RSUs to have enough processing resources to implement a consensus mechanism. Moreover, an empirical study is also required to evaluate the performance of hierarchical consensus.

Another application framework, presented by Malik et al. [94] is the applicability of permissioned blockchain in the food supply market to promote food provenance The framework ensures data availability, ensures the security of provenance information, and provides throughput scalability. This is done by creating agreed-upon rules for accessibility of data by a pre-trusted consortium of all business entities and participants and a blockchain technology vendor that consists of a hardware-software solution to manage and control the main-chain. A sharded architecture based on geographical zones is used to provide the transaction scalability. Each shard has a set of participants and dedicated validators to mine and maintain the local chain and lottery-based mechanism to select a cluster head. A global validator node is also employed which acts as a query manager to access the data based on access rules. The framework is generalized and can be applied to any supply chain scenario. However, the framework is based on strong assumption that trust is maintained through a central Certification Authority (CA).

Bloom and Farahmand [95] presented a study of the applicability of the Ethereum platform for small electricity producers in terms of increasing client base and trading frequency(Simulation in Python). The off-chain solution of heterogeneous Multi-Blockchain communication and provides routing management and message transfer with atomicity and consistency parallel running of the transaction to increase throughput(TPS). The model consists of four layers, the basic layer contains network, storage, sandbox, and DB management modules and serves as the foundation of the system. The blockchain layer consists of a block data structure, consensus, and encryption algorithms. The multi-chain communication layer is responsible for overall routing management and a cross-chain protocol to allow secure communication among various blockchains. The topmost layer is the application layer which contains services such as smart contracts, multi-ledger queries, and data opening methodologies. The core of this architecture is the router nodes, which maintain the detail of different blockchains and with the help of a unified transaction format for a standard cross-chain transaction to route information between multiple blockchains. The experiments were conducted to evaluate the performance, which shows that the TPS is increased with an increasing number of shards but decrease with an increase in the number of cross-chain transactions. Moreover, the experiments show that the Ethereum blockchain performs better for 600 participants with a trading frequency of 5 minutes per transaction but it is inversely proportional to the number of participants with higher trading frequencies furthermore the experiments are only based on Ethereum implementation and performance of other Blockchains are not studied.

8.1. Summary of application of scalable blockchain

Blockchain was initially proposed as a decentralized cryptocurrency, but with the evolution of DLTs and the arrival of smart contracts, this technology has witnessed a great expansion and now been applied to non-cryptocurrency domains including the health sector, businesses and finances, identity management, supply chain management, national/consortium databases, asset management, IoTs, and other cyber-physical systems. This section summarized existing literature with regards to applications of blockchain where blockchain is applied to increase the scalability of the existing system, a summary presented in Table 10. Especially with the exponential growth in IoT devices, and wide adoption of IoTs by domestic as well as commercial users, the researchers are attracted towards

adaption of blockchain in the resource-constrained IoT scenarios to achieve enhanced security and decentralization. However, with the continuously increasing number of IoT devices, scalability of blockchain-based solutions for IoT systems becomes a challenge. Key observations found in this section are:

- Blockchain has been applied to diverse domains (see table 10) to improve the scalability of the existing systems. However, blockchains platforms such as Hyperledger, Ethereum, and Multichain offer different strengths and services which may not be suitable for all application types. For instance, the consensus algorithms supported by blockchain platforms vary which has an impact of the type of applications where a consensus algorithm is feasible. Therefore, blockchain platforms should not be considered as a *out of the box* solution and appropriate configurations may be required to achieve optimum performance.

- Although blockchain offers fundamental benefits such as tamper-proof storage and decentralization, scalable blockchains introduce an additional dimension into the estimation of efficiency i.e. the *DSS trilemma* including decentralization, security, and scalability. Therefore, application design should take into account the trade-off between security, decentralization and scalability to achieve optimum solution.

- The integration of IoT with blockchain ([84], [85], [86], [87], [88], [89], [90]) requires adjustments in the blockchain architecture to achieve better scalability. For instance, an efficient gateway, external storage, or delivery network can reduce the workload from resource-constrained IoT devices. However, this comes at the cost of partial or full centralization. Furthermore, recent efforts in adoption of DLTs within IoT have explored alternatives to blockchains such as Tangle [97] with IOTA [98] one of the widely used technologies.

9. Approaches Analyzing Blockchain Scalability

A number of efforts have been made to identify the scalability challenge within blockchains. Such efforts have focused at analyzing blockchains with respect to specific parameters such as block generation rate, transaction processing speed, and block size to understand and highlight their role in achieving scalable solutions using blockchains. Additionally, such efforts also highlight potential of blockchain to achieve scalable solutions through appropriate configuration of blockchain parameters. A summary of efforts considered here is presented in Table 11.

In particular, Croman et al.[99] presented a study to characterize the Bitcoin network performance and suggests improvements in its scalability. The results showed that with a 10-minute BGR of the traditional Bitcoin network, a maximum of 4MB block size is practical to increase the throughput up to 27 TPS compared to the current 7 transactions per second. Furthermore, the latency or block interval must be 12s or higher to achieve full network bandwidth utilization. The authors also suggested that a fundamental protocol redesign is essential to attain the scalability of the decentralized Bitcoin network which can be achieved through use of bespoke consensus algorithms such as Proof of Stake.

Aniello et al. [100] mainly address the challenge of data integrity while also eliminating DoS attacks with BFT implementation and a DHT based implementation to improve scalability. The model is a two-layered architecture implemented on Faas scenario, the first layer consists of permissioned Blockchain and employs a consensus algorithm based on leader rotation approach at each round according to deterministic fair policy and defines the ordering of transactions. Each federation member with one miner having database ledger replicas maintains the integrity and the second layer, with permissionless POW, ensures the integrity of the first layer by storing the hashes of the first layer via an anchor manager. If a miner tries to alter the logs then the difference of hashes of both layers reveals this malicious activity. The scalability of the system is improved by using a DHT-based sharding solution in which a miner only handles specific subsets of transactions thereby making the system scalable.

Khan et al. [9] presented an empirical study for permissioned as well as permissionless blockchain to determine the impact of various scalability aspects including throughput, block size, and BGR for an e-voting system. The authors used Multichain platform as blockchain fabric with additional nodes with java based clients included for permissioned blockchain. In the permissionless scenario, with 10 voters and the BGR and block size is small, the transaction rate of 266.6 TPS is achieved. However, the average transactions per block is recored to be two which is due to the small number of voters and low BGR. Increasing the block size results in increasing the number of transactions to be handled however increase in BGR has lesser impact on overall transaction arrival and confirmation. In the permissioned scenario, with one client, stable throughput is achieved. However, as the number of concurrent clients increases the network latency also increases which results in delays and mining of empty blocks, this is to keep the blockchain live to maintain the pre-set BGR but at the cost of increasing workload on the master node.

Benčić and Žarko[101, 102] conducted a comparative study focused at evaluating two types of DLTs i.e. blockchains and DAGs. Specifically, the study considered Bitcoin and Ethereum as blockchain and nano as DAG DLT implementation comparing them with respect to various properties including data structure, consensus, ledger size, and transaction throughput. The paper also highlighted that the challenge of scalability is not limited to blockchains and is

indeed applicable to other types of DLTs such as DAGs. It also highlighted approaches that are used to achieve scalability in both DLTs including block size increase, PCNs, nested chaining and sharding in blockchain and coupling network usage with transaction verification, so that transactions are managed by each client to be able to use network resources for scalability in DAG-based DLT.

Dennis and Disso [103] presented a comparative study of the scalability of two of the most valuable blockchains; Bitcoin and Ethereum, concerning historical data. On contrary to Ethereum, which requires chain state data, the blockchain requires full chain download in order to participate as a full node in the network making it more towards centralized favoring only resource-full entities to be able to participate. Download time for full blockchain for both network is calculated which results in the need for commercial hardware to participate in the network as a full node. The study also shows that the Bitcoin storage increases almost with the constant rate whereas Ethereum is facing exponential growth in chain size due to the drastic increase in transactions per day. The authors' emphasis on temporal blockchain [71] as a potential solution to reduce the chain size while maintaining the increasing number of average throughput per day.

The paper by Hao et al. [104] proposes BlockP2P that optimizes the P2P network topology to achieve improvements in latency and node scalability. The comparative results of BlockP2P with traditional Bitcoin and Ethereum network in a simulation environment shows significant improvement in the efficiency of the proposed approach. The approach reduces the latency up to 90% and has better node scalability with respect to Bitcoin and Ethereum.

Oscar Novo [105] presented a comparative study of the proposed decentralized Access Management System (AMS) [86] with traditional AMS intending to gain an understanding of the scalability. Since the scalability of traditional access management systems depends on a single server while decentralized management implementation can handle several Management Hubs at the same time and an unlimited number of nodes in the blockchain network. The simulation results favor horizontal scalability with significant scaling advantages of the proposed approach over traditional blockchain scenarios by distributing load among different IoT devices called management hubs, however, the IoT devices need much more transaction rates, this requires optimization in the proposed techniques.

Goswami et al. [29] presented the real-time and simulation-based analysis of various scalability bottlenecks such as block size, block interval, network latency, and transaction cost. The study identified that an increase in the number of transactions will lead to an increase in confirmation time, network overhead as well as increased latency and throughput. Further, the authors identified that the transaction fee plays an important role in deciding a transaction's confirmation time. Specifically, the higher fees will increase the probability of less confirmation time. The simulation results also show that confirmation time decreases when the number of transaction increases but at the cost of increasing latency. Furthermore, increased block size to include more transactions will efficiently increase the transaction load as well as the throughput. In the current Bitcoin system, the block size is fixed at a cap of 1 MB. If an increase in efficiency is desirable, this seems like a reasonable claim but too much increase in block size will increase propagation time.

9.1. Summary of approaches analyzing blockchain scalability

The section presented (See table 11) some of the empirical / theoretical studies analyzing the scalability of blockchain that we found during our survey. various comparative studies including [101], [102], [105], [103], encompassing different blockchain technology aspects are analyzed,

The results from the simulations, real-time and theoretical analysis revealed the following key observations:

- An increase in block size results in increased throughput, however it should be well configured since block size directly affects the propagation time. For instance, the study by Croman et al. [99] revealed that, a Bitcoin network can achieve a TPS of 27 TPS in a scenario where block size does not exceed 4MB with a latency of $\geq$ 12 seconds .
- Number of transactions directly affects the verification process, network overhead, and latency [29]. Consequently, in systems with large number of transactions, appropriate mechanisms should be adopted to address the impact of high transaction numbers on the overall performance of the blockchain.
- Network optimization or the use of efficient network hardware can be used as an approach to increase the latency of a blockchain network. Such delays can also introduce vulnerabilities such as transaction malleability within blockchain with the potential to compromise blockchain state [8].

10. Open Issues and Challenges

The emergence of scalability as one of the primary challenges concerning blockchain has drawn significant attention from the research community. The widespread adoption of blockchain technology in financial as well as non-financial sectors has a profound role in this and necessitates methods and mechanisms to boost horizontal as well as vertical scalability. Increased throughput, efficient storage mechanism, and low latency are the most desirable attributes to scale a blockchain to large number of nodes and clients with minimal performance degradation. Several approaches have been proposed to achieve scalable blockchain applications keeping in view the trade-off

Table 11: Approaches analyzing blockchain scalability

Author(s)	Description	Scalability Issue Discussed/ Addressed	Publisher	Year
Croman et al.[99]	A study for Bitcoin's scalability improvements.	Transaction Throughput, latency	IFCA	2016
Leonardo Aniello et al.[100]	Integrity, Security, and Scalability in Faas Scenario	Transaction load	IEEE	2017
khan et al. [9]	Scalability analysis of permissionless and permissioned blockchain via an e-voting application.	Throughput, Blocksize, BGR	Elsevier	2020
Benčić and Žarko [101] [102]	Comparison of Bitcoin and Ethereum blockchain with Nano DAG.	TPS, Ledger size	IEEE	2018
Richard Dennis and Jules Pagna Disso [103]	Comparison of Bitcoin and Ethereum w r t scalability.	Ledger Size	Springer	2019
Weifeng Hao et al. [104]	Comparison of proposed BlockP2P protocol with Bitcoin and Ethereum	Latency, No of nodes	Springer	2019
Oscar Novo [105]	(Simulation)	Throughput rate, Clients, Latency	IEEE	2019
Sneha Goswami [29]	Simulation-based analysis of BC Scalability	Transaction confirmation time per sec	University of Nevada, Las Vegas	2017

between the most desirable but diverse DSS (Decentralization, Security, Scalability) trilemma. Without completely addressing the scalability challenges, blockchain cannot reach its true disruptive potential. However, our analysis of state-of-the-art efforts highlights several limitations of these efforts which require further efforts from the research community. We summarize the key research challenges to achieve scalable blockchains below.

- Sharding is one of the most widely adopted approaches to enhance blockchain scalability. Various sharding-based approaches including Elastico[47], Omniledger [48], and Rapidchain [31] are proposed to achieve low latency, high throughput, storage scalability, and Byzantine fault tolerance. Of these approaches, Rapidchain emerged as a protocol providing better scalability without compromising security and has a message complexity of O(n). However, these approaches are only for permissionless cryptocurrencies. On the other hand, [50] extended the solution to the general workload but solely relies on trusted hardware to reduce communication overhead.

  Reducing message complexity within blockchain in general and within sharding in particular is the area that requires further exploration. Communication Cost Per Transaction (CCPT) is one of the key to blockchain scalability [78]. A blockchain with CCPT in the order of O(n) is considered as scalable. Existing solutions, in order to achieve O(n) CCPT, either sacrifice reliability, decentralized nature of blockchain, rely on trusted hardware, or assume that all nodes are motivated and behave rationally. Potential areas that can be explored are deployment of decentralized reputation management approaches and load balancing mechanisms with an attractive incentive policy to prevent abnormal behavior of nodes without trusted hardware.

- Maintaining atomicity in a cross-shard transaction is a challenging task [53], [95]. The operations on various shards are required to follow a timeline to validate the time sequence among these operations. In this case, deployment of load balancing mechanism can be explored to deal with miner node exhaustion and DoS situations when a shard has to deal with a large number of legitimate and illegitimate cross-shard transactions.

- The Rapidchain [49] can tolerate up to 1/3 byzantine adversaries but they are only for cryptocurrency workload. A powerful mechanism to deal with byzantine adversaries and to cope with shard take over for non-crypto currency based blockchains must be in place. Since performance (throughput) is directly proportional to the number of shards and an increasing number of shards also divides the resources and computational-power per shard. Consequently, a malicious adversary requires a minuscule amount of resources to have complete control over a shard. This can benefit from further research into the applicability of BFT tolerance mechanisms such as provided by elastico, omniledger, rapidchain etc. under non-cryptocurrency workload to prevent malicious adver-

saries to completely taker over the shard.

- Another area that limits blockchain scalability and requires further attention from the research community is the formation of multiple branches of blocks known as forks. Currently, the longest chain rule [2] [11] is applied to deal with these forks. However, without any fork avoidance mechanism in place, wastage of resources still occurs which can consequently affect the overall blockchain performance. A dimension of further research for this challenge is the formation of fork monitoring committees (consisting of incentivised nodes) that are separate from mining nodes which are used to handle the mining assignments of the mining nodes according to their processing power and to continuously monitor the formation of forks.

- Through our research, we came across several approaches including [59] ,[60], and [61] which rely on efficient network infrastructure and cut-through routing to improve latency. But this requires a complete trusted-setup or a trust management scheme to prevent network becoming a bottleneck or an attack surface. In this context, decentralized and lightweight reputation management approaches can be explored to fill this gap by creating a trusted environment.

- The use of resourceful or trusted hardware devices for mining and validation in a permissioned (consortium) blockchain scenario decrease BGR which has a direct impact on transaction throughput [64], [62], [63], [103], [50]. However, its adoption in permissionless (public) blockchain scenarios requires an attractive incentive mechanism to attract miners to use hardware which has more processing power, storage, and memory.

- Through the analysis, some efforts such as [9, 102, 103] were identified which were focused at investigating the role of parameters such as block size and block generation to achieve scalable blockchains. However, most of these efforts have focused at Bitcoin blockchain and therefore are specific to the settings used within Bitcoin. Further efforts are required to conduct similar analysis at a higher level of abstraction i.e. within a specific platform such as Ethereum or Multichain so as to highlight strengths and limitations especially with respect to scalability. Such efforts will help application developers understand the role of blockchain parameters and choose appropriate blockchain platforms that are suitable for specific application domains.

- A critical component of blockchain software stack is consensus algorithm due to its decentralized operation. Although Bitcoin (first prominent blockchain application) is based on Proof of Work, several other consensus algorithms have been proposed including Proof of Stake, Proof of Authority, and Proof of Weight. However, in order to understand and highlight appropriateness of consensus algorithms for specific application domains, a rigorous empirical analysis of consensus algorithms is required. Current studies such as [106, 107] have made initial efforts by conducting a feature-based comparison of existing consensus algorithms however our search didnot identify any efforts to take an empirical approach to achieve such comparative analysis and therefore is an area of work which requires further attention from the research community.

11. Conclusions

Scalability of blockchains is one of the fundamental challenges which limit their widespread adoption in diverse application domains. This paper has contributed to defining this challenge as well as highlighting existing efforts to address this challenge for blockchains. Specifically, defining blockchain scalability as a composite concept, the paper has identified and described its different dimensions so as to refine the scope of this concept within blockchains. The paper has presented a first systematic effort to identify and collate existing efforts related to blockchain scalability in three aspects i.e. mechanisms and methods aiming to enhance blockchain scalability by contributing to the core blockchain functions, applications leveraging blockchain to achieve scalable applications, and efforts to define the scalability challenges through analysis of blockchain solutions. Although these represent a rich set of efforts, there remain open challenges to achieve scalable blockchain as identified in the paper. Our future work involves converging towards study of consensus algorithms to identify and analyze their role to enhance scalability of blockchains.

Acknowledgment

This publication is based upon work supported by the Khalifa University of Science and Technology under Award No. CIRA-2019-001

References

[1] N. Z. Aitzhan and D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 5, pp. 840–852, 2018.

[2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Cryptography Mailing list at https://metzdowd.com*, 03 2009.

[3] K. M. Khan, J. Arshad, and M. M. Khan, "Secure digital voting system based on blockchain technology," *Int. J. Electron. Gov. Res.*, vol. 14, no. 1, pp. 53–62, Jan. 2018. [Online]. Available: https://doi.org/10.4018/IJEGR.2018010103

[4] N. Nizamuddin, K. Salah, M. A. Azad, J. Arshad, and M. Rehman, "Decentralized document version control using ethereum blockchain and ipfs," *Computers Electrical Engineering*, vol. 76, pp. 183 – 197, 2019. [Online]. Available: https://bit.ly/2NMvKHN

[5] E. Project. Blockchain app platform. [Online]. Available: https://www.ethereum.org/

[6] Multichain. Open platform for blockchain applications. [Online]. Available: www.multichain.com

[7] "Hyperledger Caliper." [Online]. Available: https://hyperledger.github.io/caliper/

[8] K. M. Khan, J. Arshad, and M. M. Khan, "Simulation of transaction malleability attack for blockchain-based e-voting," *Computers and Electrical Engineering*, 2020.

[9] ——, "Investigating performance constraints for blockchain based secure e-voting system," *Future Generation Computer Systems*, vol. 105, pp. 13–26, Apr. 2020.

[10] T. McConaghy, R. Marques, A. M¨uller, D. D. Jonghe, T. McConaghy, G. McMullen, R. Henderson, S. Bellemare, and A. Granzotto, "Bigchaindb: A scalable blockchain database," 2016. [Online]. Available: https://gamma.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf

[11] M. Vukolić, "The quest for scalable blockchain fabric: Proof-of-work vs. bft replication," in *Open Problems in Network Security*, D. Camenisch, Jan andKesdoğan, Ed. Cham: Springer International Publishing, 2016, pp. 112–125.

[12] Q. Zhou, H. Huang, Z. Zheng, and J. Bian, "Solutions to scalability of blockchain: A survey," *IEEE Access*, vol. 8, pp. 16 440–16 455, 2020.

[13] A. Hafid, A. S. Hafid, and M. Samih, "Scaling blockchains: A comprehensive survey," *IEEE Access*, vol. 8, pp. 125 244–125 262, 2020.

[14] G. Fournier and F. Petrillo, "Challenges and solutions on architecting blockchain systems," in *Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering*, ser. CASCON '18. USA: IBM Corp., 2018, p. 293–300.

[15] D. Mingxiao, M. Xiaofeng, Z. Zhe, W. Xiangwei, and C. Qijun, "A review on consensus algorithm of blockchain," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2017, pp. 2567–2572.

[16] R. Wang, K. Ye, and C.-Z. Xu, "Performance benchmarking and optimization for blockchain systems: A survey," in *Blockchain – ICBC 2019*, J. Joshi, S. Nepal, Q. Zhang, and L.-J. Zhang, Eds. Cham: Springer International Publishing, 2019, pp. 171–185.

[17] S. Kim, Y. Kwon, and S. Cho, "A survey of scalability solutions on blockchain," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct 2018, pp. 1204–1207.

[18] K. BA and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," vol. 2, 01 2007.

[19] D. Chaum, "Blind signatures for untraceable payments," in *Advances in cryptology*. Springer, 1983, pp. 199–203.

[20] A. Back *et al.*, "Hashcash-a denial of service counter-measure," 2002.

[21] C. V. Helliar, L. Crawford, L. Rocca, C. Teodori, and M. Veneziani, "Permissionless and permissioned blockchain diffusion," *International Journal of Information Management*, vol. 54, p. 102136, 2020.

[22] S. Pongnumkul, C. Siripanpornchana, and S. Thajchayapong, "Performance analysis of private blockchain platforms in varying workloads," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2017, pp. 1–6.

[23] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *2017 IEEE international congress on big data (BigData congress)*. IEEE, 2017, pp. 557–564.

[24] A. Gupta, R. Christie, and P. Manjula, "Scalability in internet of things: features, techniques and research challenges," *Int. J. Comput. Intell. Res*, vol. 13, no. 7, pp. 1617–1627, 2017.

[25] "Scaling horizontally vs. scaling vertically." [Online]. Available: https://www.educative.io/edpresso/scaling-horizontally-vs-scaling-vertically

[26] "What is Vertical Scaling & Horizontal Scaling? | ESDS," Jul. 2018. [Online]. Available: https://www.esds.co.in/blog/vertical-scaling-horizontal-scaling/

[27] "vaquarkhan/vaquarkhan." [Online]. Available: https://github.com/vaquarkhan/vaquarkhan

[28] C. Li, P. Li, D. Zhou, W. Xu, F. Long, and A. Yao, "Scaling Nakamoto Consensus to Thousands of Transactions per Second," *arXiv:1805.03870 [cs]*, Aug. 2018, arXiv: 1805.03870. [Online]. Available: http://arxiv.org/abs/1805.03870

[29] S. Goswami, "Scalability analysis of blockchains through blockchain simulation.(2017)," 2017.

[30] "blocks-size." [Online]. Available: https://www.blockchain.com/charts/blocks-size

[31] X. Boyen, C. Carr, and T. Haines, "Graphchain: a blockchain-free scalable decentralised ledger," 05 2018, pp. 21–33.

[32] L. Baird, B. Gross, and D. Thibeau, "HashGraph." [Online]. Available: https://www.hedera.com/hh-whitepaper-v1.4-181017.pdf

[33] J. Xie, F. R. Yu, T. Huang, R. Xie, J. Liu, and Y. Liu, "A survey on the scalability of blockchain systems," *IEEE Network*, vol. 33, no. 5, pp. 166–173, 2019.

[34] I. Qasse, M. Talib, and Q. Nasir, "Inter blockchain communication: A survey," 03 2019, pp. 1–6.

[35] C. Bai, "State-of-the-art and future trends of blockchain based on dag structure," in *Structured Object-Oriented Formal Language and Method*, Z. Duan, S. Liu, C. Tian, and F. Nagoya, Eds. Cham: Springer International Publishing, 2019, pp. 183–196.

[36] D. Mechkaroska, V. Dimitrova, and A. Popovska-Mitrovikj, "Analysis of the possibilities for improvement of blockchain technology," in *2018 26th Telecommunications Forum (TELFOR)*, Nov 2018, pp. 1–4.

[37] C. Worley and A. Skjellum, "Blockchain tradeoffs and challenges for current and emerging applications: Generalization, fragmentation, sidechains, and scalability," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, July 2018, pp. 1582–1587.

[38] SegWit, 2019. [Online]. Available: https://segwit.org/

[39] J. Poon and T. Dryja, "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments," Technical Report. [Online]. Available: https://lightning.network/lightning-network-paper.pdf

[40] G. Malavolta, P. Moreno-Sanchez, C. Schneidewind, A. Kate, and M. Maffei, "Anonymous multi-hop locks for blockchain scalability and interoperability." in *NDSS*, 2019.

[41] "Raiden Network." [Online]. Available: https://raiden.network/

[42] M. Green and I. Miers, "Bolt: Anonymous Payment Channels for Decentralized Currencies," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security - CCS '17*. Dallas, Texas, USA: ACM Press, 2017, pp. 473–489. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3133956.3134093

[43] "Basis of Lightning Technology (BOLT)." [Online]. Available: https://wiki.ion.radar.tech/tech/lightning/basics-of-lightning-technology-bolt

[44] "ElementsProject/lightning," Mar. 2020, original-date: 2015-06-13T00:04:22Z. [Online]. Available: https://github.com/ElementsProject/lightning

[45] "lightningnetwork/lightning-rfc," Mar. 2020, original-date: 2016-11-14T19:21:45Z. [Online]. Available: https://github.com/lightningnetwork/lightning-rfc

[46] "lightningnetwork/lnd," Mar. 2020, original-date: 2016-01-16T08:19:33Z. [Online]. Available: https://github.com/lightningnetwork/lnd

[47] L. Luu, V. Narayanan, C. Zheng, K. Baweja, S. Gilbert, and P. Saxena, "A secure sharding protocol for open blockchains,"

10 2016, pp. 17–30.
[48] E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, E. Syta, and B. Ford, "Omniledger: A secure, scale-out, decentralized ledger via sharding," in *2018 IEEE Symposium on Security and Privacy (SP)*, May 2018, pp. 583–598.
[49] M. Zamani, M. Movahedi, and M. Raykova, "Rapidchain: Scaling blockchain via full sharding," 10 2018, pp. 931–948.
[50] H. Dang, T. Dinh, D. Loghin, E.-C. Chang, Q. Lin, and B. Ooi, "Towards scaling blockchain systems via sharding," 06 2019, pp. 123–140.
[51] K. R. Özyılmaz, H. Patel, and A. Malik, "Split-scale: Scaling bitcoin by partitioning the utxo space," in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Nov 2018, pp. 41–45.
[52] Z. Ren and E. Zekeriya, "A scale-out blockchain for value transfer with spontaneous sharding," 01 2018.
[53] L. Kan, Y. Wei, A. Hafiz Muhammad, W. Siyuan, G. Linchao, and H. Kai, "A multiple blockchains architecture on inter-blockchain communication," 07 2018, pp. 139–145.
[54] Y. Yu, R. Liang, and J. Xu, "A scalable and extensible blockchain architecture," 11 2018, pp. 161–163.
[55] W. Li, A. Sforzin, S. Fedorov, and G. O. Karame, "Towards scalable and private industrial blockchains," in *Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts*, ser. BCC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 9–14. [Online]. Available: https://doi.org/10.1145/3055518.3055531
[56] G. Danezis and S. Meiklejohn, "Centrally Banked Cryptocurrencies," *arXiv:1505.06895 [cs]*, Dec. 2015, arXiv: 1505.06895. [Online]. Available: http://arxiv.org/abs/1505.06895
[57] E. K. Kogias, P. Jovanovic, N. Gailly, I. Khoffi, L. Gasser, and B. Ford, "Enhancing bitcoin security and performance with strong consistency via collective signing," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 279–296. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/kogias
[58] P. Maymounkov and D. Mazières, "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric," in *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, ser. IPTPS '01. Berlin, Heidelberg: Springer-Verlag, Mar. 2002, pp. 53–65.
[59] "bloxroute: A scalable trustless blockchain distribution network author=Uri Klarman and Soumya Basu and Aleksandar Kuzmanovic and Emin G¨un Sirer, year=2018, url=https://bloxroute.com/wp-content/uploads/2019/01/whitepaper-V1.1-1.pdf,."
[60] G. He, W. Su, and S. Gao, "Chameleon: A scalable and adaptive permissioned blockchain architecture," 08 2018, pp. 87–93.
[61] A. Kuzmanovic, "Net neutrality: Unexpected solution to blockchain scaling," *Queue*, vol. 17, pp. 20–78, 01 2019.
[62] J. Liu, W. Li, G. O. Karame, and N. Asokan, "Scalable byzantine consensus via hardware-assisted secret sharing," *IEEE Transactions on Computers*, vol. 68, no. 1, pp. 139–151, Jan 2019.
[63] A. Sanka and R. C. Cheung, "Efficient high performance fpga based nosql caching system for blockchain scalability and throughput improvement," 12 2018, pp. 1–8.
[64] J. Lind, O. Naor, I. Eyal, F. Kelbert, P. Pietzuch, and E. G. Sirer, "Teechain: A secure payment network with asynchronous blockchain access," 2017.
[65] P. Das, L. Eckey, T. Frassetto, D. Gens, K. Hostáková, P. Jauernig, S. Faust, and A.-R. Sadeghi, "Fastkitten: Practical smart contracts on bitcoin," in *28th {USENIX} Security Symposium ({USENIX} Security 19)*, 2019, pp. 801–818.
[66] S. S. Hazari and Q. H. Mahmoud, "A parallel proof of work to improve transaction speed and scalability in blockchain systems," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2019, pp. 0916–0921.
[67] Z. Gao, L. Xu, L. Chen, N. Shah, Y. Lu, and W. Shi, "Scalable blockchain based smart contract execution," 12 2017, pp. 352–359.
[68] I. Eyal, A. E. Gencer, E. G. Sirer, and R. V. Renesse, "Bitcoin-ng: A scalable blockchain protocol," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*. Santa Clara, CA: USENIX Association, Mar. 2016, pp. 45–59. [Online]. Available: https://www.usenix.org/conference/nsdi16/technical-sessions/presentation/eyal
[69] K. Zhang and H. Jacobsen, "Towards dependable, scalable, and pervasive distributed ledgers with blockchains," in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, July 2018, pp. 1337–1346.
[70] X. Fan and Q. Chai, "Roll-dpos: A randomized delegated proof of stake scheme for scalable blockchain-based internet of things systems," 11 2018, pp. 482–484.
[71] R. Dennis, G. Owenson, and B. Aziz, "A temporal blockchain: A formal analysis," 10 2016, pp. 430–437.
[72] X. Min, Q. Li, L. Liu, and L. Cui, "A permissioned blockchain framework for supporting instant transaction and dynamic block size," in *2016 IEEE Trustcom/BigDataSE/ISPA*, Aug 2016, pp. 90–96.
[73] P. Otte, M. de Vos, and J. Pouwelse, "Trustchain: A sybil-resistant scalable blockchain," *Future Generation Computer Systems*, 2017. [Online]. Available: https://tinyurl.com/yafu55la
[74] X. Fan, "Scalable practical byzantine fault tolerance with short-lived signature schemes," in *Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering*, ser. CASCON '18. USA: IBM Corp., 2018, p. 245–256.
[75] C. Ehmke, F. Blum, and C. Friedrich, "Proof-of-property - a lightweight and scalable blockchain protocol," 05 2018.
[76] P. Thakkar, S. Nathan, and B. Viswanathan, "Performance benchmarking and optimizing hyperledger fabric blockchain platform," in *2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*. IEEE, 2018, pp. 264–276.
[77] H. Gupta and D. Janakiram, "Colosseum: A scalable permissioned blockchain over structured network," 07 2019, pp. 23–25.
[78] G. Wang, Z. J. Shi, M. Nixon, and S. Han, "Smchain: A scalable blockchain protocol for secure metering systems in distributed industrial plants," in *Proceedings of the International Conference on Internet of Things Design and Implementation*, ser. IoTDI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 249–254. [Online]. Available: https://doi.org/10.1145/3302505.3310086
[79] H. Kalodner, S. Goldfeder, X. Chen, S. M. Weinberg, and E. W. Felten, "Arbitrum: Scalable, private smart contracts," in *27th {USENIX} Security Symposium ({USENIX} Security 18)*, 2018, pp. 1353–1370.
[80] J. Teutsch and C. Reitwießner, "A scalable verification solution for blockchains," *arXiv preprint arXiv:1908.04756*, 2019.
[81] S. Gorbunov, L. Reyzin, H. Wee, and Z. Zhang, "Pointproofs: Aggregating proofs for multiple vector commitments." *IACR Cryptol. ePrint Arch.*, vol. 2020, p. 419, 2020.
[82] J. Yin, C. Wang, Z. Zhang, and J. Liu, *Revisiting the Incentive Mechanism of Bitcoin-NG*, 06 2018, pp. 706–719.
[83] Z. Wang, "Moca: A scalable consensus algorithm based on cellular automata," 11 2018, pp. 314–318.
[84] M. Ruta, F. Scioscia, S. Ieva, G. Capurso, and E. D. Sciascio, "Semantic blockchain to improve scalability in the internet of things," *OJIOT*, vol. 3, pp. 46–61, 2017.
[85] A. Dorri, S. S. Kanhere, and R. Jurdak, "Towards an optimized blockchain for iot," in *2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI)*, 2017, pp. 173–178.
[86] O. Novo, "Blockchain meets iot: An architecture for scalable

access management in iot," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1184–1195, 2018.

[87] M. Grabatin and W. Hommel, "Reliability and scalability improvements to identity federations by managing saml metadata with distributed ledger technology," 04 2018, pp. 1–6.

[88] S. Biswas, K. Sharif, F. Li, B. Nour, and Y. Wang, "A scalable blockchain framework for secure transactions in iot," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4650–4659, 2019.

[89] R. Li, T. Song, B. Mei, H. Li, X. Cheng, and L. Sun, "Blockchain for large-scale internet of things data storage and protection," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 762–771, 2019.

[90] D. Puthal, S. P. Mohanty, P. Nanda, E. Kougianos, and G. Das, "Proof-of-authentication for scalable blockchain in resource-constrained distributed systems," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, 2019, pp. 1–5.

[91] A. Hari and T. Lakshman, "The internet blockchain: A distributed, tamper-resistant transaction framework for the internet," 11 2016, pp. 204–210.

[92] X. Xu, C. Pautasso, L. Zhu, V. Gramoli, A. Ponomarev, A. B. Tran, and S. Chen, "The blockchain as a software connector," in *2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA)*, 2016, pp. 182–191.

[93] R. van der Heijden, F. Engelmann, D. Modinger, F. Schönig, and F. Kargl, "Blackchain: Scalability for resource-constrained accountable vehicle-to-x communication," 10 2017.

[94] S. Malik, S. S. Kanhere, and R. Jurdak, "Productchain: Scalable blockchain framework to support provenance in supply chains," in *2018 IEEE 17th International Symposium on Network Computing and Applications (NCA)*, 2018, pp. 1–10.

[95] F. Blom and H. Farahmand, "On the scalability of blockchain-supported local energy markets," in *2018 International Conference on Smart Energy Systems and Technologies (SEST)*, 2018, pp. 1–6.

[96] S. Biswas, K. Sharif, F. Li, B. Nour, and Y. Wang, "A scalable blockchain framework for secure transactions in iot," 10 2018.

[97] S. Popov, "The tangle," 2015. [Online]. Available: https://tinyurl.com/yd79luyf

[98] D. M and N. Biradar, "Iota-next generation block chain," *International Journal Of Engineering And Computer Science*, vol. 7, pp. 23 823–23 826, 04 2018.

[99] K. Croman, C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, E. Sirer, D. Song, and R. Wattenhofer, "On scaling decentralized blockchains," vol. 9604, 02 2016, pp. 106–125.

[100] L. Aniello, E. Gaetani, F. Lombardi, A. Margheri, and V. Sassone, 09 2017.

[101] F. M. Benčić and I. P. Žarko, "Distributed Ledger Technology: Blockchain Compared to Directed Acyclic Graph," *arXiv:1804.10013 [cs]*, Apr. 2018, arXiv: 1804.10013. [Online]. Available: http://arxiv.org/abs/1804.10013

[102] F. M. Benčić and I. Podnar Žarko, "Distributed ledger technology: Blockchain compared to directed acyclic graph," in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2018, pp. 1569–1570.

[103] R. Dennis and J. Pagna Diss, *An Analysis into the Scalability of Bitcoin and Ethereum: ICICT 2018, London*, 01 2019, pp. 619–627.

[104] W. Hao, J. Zeng, X. Dai, J. Xiao, Q. Hua, H. Chen, K.-C. Li, and H. Jin, *BlockP2P: Enabling Fast Blockchain Broadcast with Scalable Peer-to-Peer Network Topology*, 04 2019, pp. 223–237.

[105] O. Novo, "Scalable access management in iot using blockchain: A performance evaluation," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4694–4701, 2019.

[106] L. Bach, B. Mihaljevic, and M. Zagar, "Comparative analysis of blockchain consensus algorithms," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018, pp. 1545–1550.

[107] N. Chaudhry and M. M. Yousaf, "Consensus algorithms in blockchain: comparative analysis, challenges and opportunities," in *2018 12th International Conference on Open Source Systems and Technologies (ICOSST)*. IEEE, 2018, pp. 54–63.

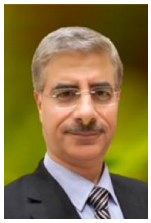
Muhammad Hassan Nasir

Junaid Arshad

Muhammad Mubashir Khan

Mahawish Fatima

Khaled Salah



Raja Jayaraman

Major highlights of the paper are:

- A comprehensive definition of scalability within the context of blockchains has been achieved which encompasses both horizontal and vertical dimensions thereby achieving a broad coverage.
- An in-depth study and critical review of existing literature with respect to blockchain scalability has been presented. This review provides a critical insight into the state of the art within scalable blockchains focusing at blockchain enhancement mechanisms, blockchain-based scalable applications, and scalability analysis of blockchains.
- Identification of open challenges to achieve scalable blockchains based on critical review of existing literature aiming to help channel efforts by research community to achieve scalability in blockchain-based applications.

Muhammad Hassan Nasir is a Ph.D. scholar in the field of cybersecurity at NED UET, Karachi, Pakistan. He received his Bachelors and Masters of Engineering degrees from Mehran UET Jamshoro, Pakistan in the year 2006 and 2012 respectively. His research interests includes Blockchain, Intrusion detection, Swarm Intelligence, and IoT security

Junaid Arshad is an Associate Professor in cybersecurity at the Birmingham City University, UK. Junaid achieved his PhD from the University of Leeds, UK where he investigated the challenge of effective intrusion severity analysis for clouds. His research is focused at challenges within cyber security emphasising impact of novel and emerging technological paradigms, such as blockchain, distributed systems, cloud computing and big data. He has been actively involved in publishing high quality research within this field and has served on Program and Review Committee of a number of journals and conferences.

Muhammad Mubashir Khan is an Associate Professor in the Department of Computer Science and Information Technology at NED University of Engineering and Technology, Karachi Pakistan. He received his Ph.D. degree in Computing from University of Leeds, UK in 2011. His current research interests include Network and Information Security, Cybersecurity and Quantum Cryptography.

Mahawish Fatima is a Senior Lecturer in the Department of Software Engineering at Bahria University, Karachi, Pakistan. She obtained her Master's degree in Information technology in 2014 and Bachelor in Computer system engineering in 2011 from MUET Jamshoro, Pakistan. She is currently pursuing her Ph.D. from Pakistan. Her research interests include the area of Cybersecurity, Internet of Things (IoT) and Machine Learning.

Khaled Salah is a full professor at the Department of Electrical and Computer Engineering, Khalifa University, UAE. He received the Ph.D. degree in Computer Science from the same institution in 2000. Khaled has over 220 publications and 3 US patents, has been giving a number of international keynote speeches, invited talks, tutorials, and research seminars on the subjects of Blockchain, IoT, Fog and Cloud Computing, and Cybersecurity. Khaled served as the chair of the Track Chair of IEEE Globecom 2018 on Cloud Computing. Khaled is an Associate Editor of IEEE Blockchain Tech Briefs, and a member of IEEE Blockchain Education Commitee. Khaled is now leading a number of projects on how to leverage blockchain for Healthcare, 5G Networks, Combating Deepfake Videos, Physical Internet, Digital Twins, Supply Chain Management, and AI.

Raja Jayaraman is an Associate Professor in the Department of Industrial & Systems Engineering at Khalifa University, Abu Dhabi, UAE. He received his Ph.D. in Industrial Engineering from Texas Tech University, a Master of Science degree in Industrial Engineering from New Mexico State University, Masters and Bachelor's degree in Mathematics from India. Raja's research interests are primarily focused in the application of blockchain technology, systems engineering and process optimization techniques to characterize, model and analyze complex systems with applications to supply chains, maintenance planning, and healthcare delivery. He has led several successful research projects and pilot implementations in the area of supply chain data standards in the US healthcare system. His research has appeared in top rated journals including: Annals of Operations Research, IISE Transactions, Computers & Industrial Engineering, IEEE Access, IEEE Network, Energy Policy, Applied Energy, Knowledge Based Systems, Journal of Theoretical Biology, Engineering Management Journal and others.

CRediT author statement:

Muhammad Hassan Nasir: Writing- Original draft preparation, Investigation

Junaid Arshad: Conceptualization, Methodology, Writing- Original draft preparation

Muhammad Mubashir Khan: Conceptualization, Editing, Supervision

Mahawish Fatima: Writing-Reviewing, Editing

Khaled Salah: Writing- Reviewing, Methodology, Funding Acquisition

Raja Jayaraman: Editing, Supervision, Funding Acquisition

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: