Exhibit 224

Information Processing and Management 58 (2021) 102584

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Taxonomy of centralization in public blockchain systems: A systematic literature review

Ashish Rajendra Sai [a,*], Jim Buckley [a], Brian Fitzgerald [a], Andrew Le Gear [b]

[a] Lero, Tierney Building, University of Limerick, Ireland
[b] Horizon Globex Ireland DAC, T1-017, Tierney Building, Nexus Center, University of Limerick, Ireland

ARTICLE INFO

Keywords:
Decentralized blockchain
Centralization
Classification
Measurement
Taxonomy
Security

ABSTRACT

Bitcoin introduced delegation of control over a monetary system from a select few to all who participate in that system. This delegation is known as the decentralization of controlling power and is a powerful security mechanism for the ecosystem. After the introduction of Bitcoin, the field of cryptocurrency has seen widespread attention from industry and academia, so much so that the original novel contribution of Bitcoin, i.e., decentralization, may be overlooked, due to decentralizations' assumed fundamental existence for the functioning of such crypto-assets. However, recent studies have observed a trend of increased centralization in cryptocurrencies such as Bitcoin and Ethereum. As this increased centralization has an impact the security of the blockchain, it is crucial that it is measured, towards adequate control. This research derives an initial taxonomy of centralization present in decentralized blockchains through rigorous synthesis using a systematic literature review. This is followed by iterative refinement through expert interviews. We systematically analyzed 89 research papers published between 2009 and 2019. Our study contributes to the existing body of knowledge by highlighting the multiple definitions and measurements of centralization in the literature. We identify different aspects of centralization and propose an encompassing taxonomy of centralization concerns. This taxonomy is based on empirically observable and measurable characteristics. It consists of 13 aspects of centralization, classified over six architectural layers: Governance, Network, Consensus, Incentive, Operational, and Application. We also discuss how the implications of centralization can vary depending on the aspects studied. We believe that this review and taxonomy provides a comprehensive overview of centralization in decentralized blockchains involving various conceptualizations and measures.

1. Introduction

Since the introduction of Bitcoin in 2009, blockchain technology has seen a proliferation of scholarly articles investigating the potential and limitations of the technology (Androulaki et al., 2018; Beck, Avital, Rossi, & Thatcher, 2017; Beck, Müller-Bloch, & King, 2018; Davidson, De Filippi, & Potts, 2016; He, Yu, Zhang, & Bao, 2017; Mattila, 2016; Walport, 2016; Wüst & Gervais, 2018; Yli-Huumo, Ko, Choi, Park, & Smolander, 2016; Zheng, Xie, Dai, Chen, & Wang, 2017). Control over the system is a focal point in a significant proportion of these studies, as this either enhances or restricts the usability of blockchain in a wider information system context (Alzahrani & Bulusu, 2018; Azouvi, Maller, & Meiklejohn, 2018; Baliga, 2017; Beck et al., 2017; Beikverdi & Song, 2015; Cong, He, & Li, 2019; Gencer, Basu, Eyal, Van Renesse, & Sirer, 2018; Gervais, Karame, Capkun, & Capkun, 2014; Judmayer, Stifter,

* Corresponding author.
E-mail address: 17053145@studentmail.ul.ie (A.R. Sai).

https://doi.org/10.1016/j.ipm.2021.102584
Received 25 June 2020; Received in revised form 4 February 2021; Accepted 7 March 2021
Available online 31 March 2021

Krombholz and Weippl, 2017; Kwon, Liu, Kim, Song, & Kim, 2019; Mattila, 2016; Sai, Buckley, & Le Gear, 2019a; Wang et al., 2018; Zhang, Xue, & Liu, 2019; Zheng et al., 2017). Indeed, removing central control from the monetary system while continuing to ensure security has been considered as a core novel contribution of Bitcoin (Bonneau et al., 2015). There are three main types of blockchain solutions based on the type of control mechanism used: public, private, and consortium (Zheng et al., 2017).

In public blockchains, such as Bitcoin and Ethereum, every participant in the network contributes to the control mechanism, agreeing on a single state of the data without the need for a trusted third party. All participants can read and write to this single state without any authorization (Guegan, 2017). This consensus is achieved under the assumption of delegation of power of control, and the assumption that the majority of the network participants remain honest i.e., non-malicious. This delegation of power of control is often referred to as decentralization.

Contrary to the decentralized nature of a public blockchain, private and consortium blockchains, such as Hyperledger, tend to impose constraints on participants by including trusted entities in the system (Androulaki et al., 2018; Xu et al., 2021). These constraints can also include limitations on read and write permissions of participants (Guegan, 2017). Based on the sensitivity of the information processed by the blockchain, practitioners may decide to adopt one of these controlling mechanisms (Meijer & Ubacht, 2018; Peck, 2017; Wüst & Gervais, 2018). As reported by Berdik, Otoum, Schmidt, Porter, and Jararweh (2020), the sheer number and complexity of various types of blockchain and their attributes can make it difficult to specifically address the benefits and shortcomings of blockchain as a service for applications within today's information systems. This decision is potentially problematic, e.g., in the case of a practitioner who decides to use a public blockchain for decentralizing the control. As reported by Sai et al. (2019a), decentralization in public blockchain is not a fundamental given by design, but a non-deterministic and probabilistic guarantee provided by clever integration of cryptography, distributed systems, and incentive engineering.

The removal of trusted entities from a distributed system makes a public blockchain attractive to numerous potential users in academia and industry (Mattila, 2016). Public blockchain-based cryptocurrencies have a market capitalization of over $1.05 trillion (Sai, Buckley, & Le Gear, 2019b), making the platform a lucrative target for malicious actors. The majority of these blockchains use decentralization as a security mechanism. In a decentralized system, the malicious actor would need to compromise half of the consensus power before causing significant harm to the system (Karame, Androulaki, & Capkun, 2012). Because of this interplay between decentralization and security, it is highly desirable to have a high degree of decentralization in public blockchains. The security of a public blockchain has been thoroughly investigated in research (Bonneau et al., 2015; Halpin & Piekarska, 2017; Karame, 2016; Karame & Androulaki, 2016). For example, Bitcoin has been reported as secure, subject to its adherence to the honest majority assumption, with notable exceptions such as selfish mining attacks (Sapirshtein, Sompolinsky, & Zohar, 2016) where the attacker only needs to control over 26% of the network.

Even though the initial implementation of Bitcoin was able to circumvent the need for centralization in the system, new avenues of centralization are surfacing (Gervais et al., 2014). Numerous studies have reported various forms of centralization in Bitcoin and other decentralized cryptocurrency systems (Azouvi et al., 2018; Beikverdi & Song, 2015; Gencer et al., 2018; Gervais et al., 2014). These reports of a trend towards centralization have raised security concerns as the security guarantee of a public blockchain is inherently dependent on the honest majority assumption (Sai et al., 2019a). As reported by Gencer et al. (2018), Bitcoin's network is dominated by consortiums of participants working together in groups known as mining pools. We report that the top 4 mining pools constitute 50.36% of controlling power in Bitcoin with our analysis in Section 5.3. This power accumulation trend is also evident in Ethereum, where the top 4 mining pools aggregate 63% of controlling power (Section 5.3). Given that successful attacks on these networks are much more feasible when 50% of the network chose to carry out such an attack, this mining-centralization implies that only a relatively small number of participants (the heads of these mining pools) need to adopt a dishonest approach to threaten these Blockchains. This illustrates the importance of mining-centralization to the cryptocurrency-ecosystems and this research expands that focus to look at the wider implications of centralization in general in Blockchain systems.

Trusting the probabilistic security guarantees of a public blockchain has often been identified as a barrier to entry in the ecosystem (Iansiti & Lakhani, 2017). Security of Blockchain is considered central to the adoption (Akram, Malik, Singh, Anita, & Tanwar, 2020). The security of prominent blockchains seem to depend on the appropriate decentralization (Sai et al., 2019a). Thus deeply understanding the interplay between security and centralization is an important endeavor. The threats of centralization range well beyond security into adoption, and even crypto-economics (Conti, Kumar, Lal, & Ruj, 2018). The decentralized nature of bitcoin permits the uncensored execution of transactions in the payment system irrespective of political or geographical associations. Centralization may threaten the uncensored nature of the decentralized blockchain. Thus, it is crucial for the security and, consequently, the utility of public blockchain systems that they remain adequately decentralized.

Given the significance of decentralization, several studies have analyzed technical aspects (Beikverdi & Song, 2015; Gencer et al., 2018; Gervais et al., 2014) as well as social constructs of decentralization (Azouvi et al., 2018). By far, the most commonly measured aspects of centralization is the consensus power concentration (Azouvi et al., 2018; Beikverdi & Song, 2015; Gencer et al., 2018; Gervais et al., 2014; Kwon et al., 2019). In a Proof-of-Work based blockchain solution, the individual participants' consensus power is defined by their computational power in proportion to the total computational power of the network. However, this measurement mechanism is only useful in determining the present state of the computational power portions of the network. It fails to capture the multitude of factors that may constitute the overall centralization of the system, such as system governance (Beck et al., 2018), wealth concentration (Chohan, 2019), and geographic distribution of participants (Gencer et al., 2018).

To better understand the semantics of decentralization in blockchain, we intend to measure it on all building blocks of the public blockchain. As reported by Wang, Vergne, and Hsieh (2017), the governance structure of the blockchain can have a profound impact on the operations of a public blockchain but is often overlooked as a potential source of centralization. The issues caused by centralization of governance include the long-discussed issue of block size in Bitcoin (Caffyn, 2015) and specific instances

Fig. 1. Methodology.

of unilateral decision making regarding forks in Ethereum (Wirdum, 2016). Bitcoin and other similar cryptocurrencies rely on improvement protocols to dictate the changes in the core system. According to the empirical analysis of Azouvi et al. (2018), the authors report that the vast majority of the improvement proposal in Ethereum are authored by a single user, Vitalik Buterin, the founder of Ethereum. They also report a similar trend for Bitcoin, where a handful of users contribute to the improvement protocol. This observation has been cited as a potential source of centralization in the governance of these cryptocurrencies (Gervais et al., 2014) and may serve to stifle innovation. Alternatively, they may serve to promote high-quality changes from a pool of proposers that know the Bitcoin/Ethereum ecosystems intimately. But the first step in studying this phenomenon is to acknowledge the potential for centralization in improvement protocol and to formulate a measurement to assess it.

In this study, we identify other forms of centralization, including end-user application centralization. According to Böhme, Christin, Edelman, and Moore (2015), 95% of all Bitcoin trades are processed by seven centralized organizations known as exchanges. Thus, the presence of centralized exchanges may be a contributing factor to the wealth concentration on the blockchain. Based on the analysis of Srinivasan (2017), Bitcoin and Ethereum have a wealth inequality greater than the worst real-world economy. This wealth centralization has been linked to severe security threats (Section 4) (see Fig. 1).

Consequently, we reason that we need a vocabulary to discuss and measure centralization in a more holistic manner. To allow for such modular measurement of centralization, we review the generic architecture (Zhang et al., 2019) of blockchain and use it to identify potential avenues of centralization, via a literature review of the field. Focusing on the generic architecture enables us to capture centralization-causing factors that are not implementation-specific, i.e., the same model may be used for both Bitcoin and Ethereum. We also use the generic architecture to partition different centralization concerns into architectural categories such as consensus, network, and application. This abstraction allows us to organize and observe centralization holistically. Thus, in this work, we present the first in-depth analysis of centralization in blockchains to assess the following questions:

RQ1: What are the different aspects of centralization in public blockchains?

RQ2: How can centralization be adequately measured in a decentralized blockchain instance?

To study decentralization in blockchain, we coded and analyzed the content of relevant blockchain literature (see Fig. 1). We chose ten years subsequent to the publication of the original Bitcoin white paper (Nakamoto, 2008). The survey process was primarily driven by the guidelines provided by Kitchenham (2004). In adherence to the guidelines, we conducted a five-step systematic literature review consisting of *Search, Selection, Quality Assessment, Data Extraction, and Analysis*. This systematic literature review produced the final article pool of 89 articles. These final articles, partitioned by architectural components, form the basis of the taxonomy proposed in this review.

Following the development of the taxonomy, we interviewed industrial and academic experts in the blockchain domain to establish the completeness of the taxonomy and to assess any redundant or less relevant components of the taxonomy. This consisted of ten expert interviews: four academic researchers and six industry experts. It resulted in an iterative refinement of the taxonomy.

We believe that the taxonomy presented in this article can assist in better understanding the socio-technical nature of blockchain-based information systems. The taxonomy focuses on reporting the security and performance implications of centralization systematically to reduce the complexities involved in understanding the benefits and shortcomings of blockchain for information systems, as reported by Berdik et al. (2020). We also highlight the issues associated with managing a decentralized blockchain system in the form of governance and protocol improvements. The paper makes the following contributions:

- We systematically review the existing literature to document the different aspects of centralization in public blockchains (Section 3).
- We outline the different techniques employed in the literature to measure centralization (Section 4).
- We manifest the findings of our review in a conceptual taxonomy that encompasses both categorization and measurement of different aspects of centralization in public blockchains (Section 4).
- We illustrate the relevance and utility of this taxonomy by presenting the centralization state of the two most prominent blockchain instances: Bitcoin and Ethereum, based on this taxonomy (Section 5). We also discuss how the adverse impact of centralization varies depending on aspects (Section 6).
- We identify research gaps specifically with regards to the lack of non-Bitcoin-specific centralization investigations. We also report on the lack of objective metrics for some centralization causing factors.

2. Background

The term blockchain is often used as a generic descriptor for the broader field of Distributed Ledger Technologies (Great Britain. Government Office for Science, 2016). Distributed ledger technology refers to the distributed computing networks that record, share, and synchronize data across many participants. More specifically, Blockchain is a type of data structure used to record data on these distributed computing networks. It is a chronologically linked list of data packets received by the participants within a predefined time period. These blocks are connected in a chronological order to form a chain of blocks. The link between these blocks is secured by the use of a computationally hard cryptographic hash function (Nakamoto, 2008). As the chain of blocks grows, the difficulty involved in recalculating the hash value also grows to make any alteration to past data expensive. This growth in difficulty leads to a deterministic guarantee of data immutability.

The participants of the blockchain-based network have to reach consensus on a single state of this append-only structure. Blockchain-based systems utilize a peer-to-peer distributed system with a clever incentive mechanism (Baliga, 2017) to accomplish this consistency of data in an unconstrained distributed environment. Proof-of-work (PoW) and Proof-of-stake (PoS) are two prominent examples of consensus mechanisms used in blockchain-based systems. In PoW, the participants are expected to perform computationally expensive operations to solve a puzzle. The first participant to solve and propagate the solution to a majority of the network is rewarded. PoW is often criticized for the extensive use of electricity (O'Dwyer & Malone, 2014). This issue of electricity usage is addressed in PoS, where the reward distribution is based on the monetary assets of the participants (Nguyen et al., 2019). Other notable consensus algorithms include Proof-of-Authority, Proof of Elapsed Time, and Delegated Proof-of-Stake; we refer the reader to Mingxiao, Xiaofeng, Zhe, Xiangwei, and Qijun (2017) for an in-depth review of consensus algorithms.

As discussed earlier, based on the type of consensus mechanism deployed and the constraints imposed, we can segment blockchain-based systems in three broad categories: Public, Private, and Consortium. In private and consortium-based blockchain systems, the participation in consensus is limited to users approved by a trusted authority. However, in Public blockchain systems, the participation in consensus is open to any individual with appropriate computing and networking capabilities. This unconstrained access to controlling power for all participants in the network is referred to as decentralization. Bitcoin and other public blockchains establish consensus on the blockchain through a decentralized, pseudonymous protocol. This protocol can be considered a core innovation and possibly the most crucial ingredient to the success of public blockchains (Bonneau et al., 2015).

The possibility of decentralized control over a computing network without oversight has resulted in many novel applications of the blockchain technology in information systems to improve efficiency or increase the security of the operations (Hileman & Rauchs, 2017). The blockchain technology provides a general-purpose approach to managing information in a non-trusted computing environment enabling a plethora of information systems use cases such as auditing of big data (Li, Wu, Jiang, & Srikanthan, 2020), secure information management (Putz, Dietz, Empl, & Pernul, 2021), countering fake news (Chen, Srivastava, Parizi, Aloqaily, & Ridhawi, 2020), cloud computing (Baniata, Anaqreh, & Kertesz, 2021), health data management (Hardin & Kotz, 2021), copyright management (Jing, Liu, & Sugumaran, 2021), IoT management (Chen et al., 2020; Zhao, Chen, Liu, Baker, & Zhang, 2020) and assisting autonomous vehicles in reaching consensus on events (Esposito, Ficco, & Gupta, 2021; Khalid et al., 2021; Oham, Michelin, Jurdak, Kanhere, & Jha, 2021).

2.1. Decentralization and public blockchain

Decentralization is an essential property of public Blockchain systems where participants can read, write data, and contribute to consensus without authorization (Davidson et al., 2016). In this subsection, we review the existing discussion around decentralization in the blockchain.

Consensus on the state of data in a public blockchain is attained by the acceptance of a valid block by the network in a time interval determined by a stochastic process to maintain a predefined expected time interval. To deter malicious participants from accepting fraudulent blocks, the majority of the control must be decentralized. This decentralization of control ensures that the blockchain is secure from malicious participants as long as the majority of the network remains honest. This interplay of security and decentralization makes it fundamental that the system remains decentralized.

A survey paper by He et al. (2017) identifies decentralization, among other features, as a prominent reason to adopt blockchain technology for business applications. This view is supported by numerous studies which demonstrate the application of decentralized Blockchains to the liberalization of financial asset management (Guo & Liang, 2016), the Internet of Things (Panarello, Tapas, Merlino, Longo, & Puliafito, 2018; Zhu, Loke, Trujillo-Rasua, Jiang, & Xiang, 2019), healthcare (Dwivedi, Srivastava, Dhar, & Singh, 2019) and smart cities (Xie et al., 2019). The extent of literature surveyed by these review articles demonstrates the significance of decentralization in blockchain applications.

As decentralization is core to the secure functioning of public blockchains, it may be taken as a fundamental given. This assumed association between decentralization and public blockchains may be a vulnerability that malicious actors attack. Security research on the blockchain has focused on the assumption of an honest majority. A survey paper by Li, Jiang, Chen, Luo, and Wen (2017) identifies the centralization of consensus power as a significant security threat to that network. Centralization of consensus power is intrinsic to attacks on the public blockchain, such as the 51% attack (Bradbury, 2013) and Selfish Mining (Sapirshtein et al., 2016).

In the 51% attack, the attacker is assumed to have gained control of more than half of the consensus power, which can then be used to enter fraudulent transactions in the blockchain. Unlike the 51% attack, in selfish mining, the attacker only needs to control 26% consensus power to cause harm to the network (Sai et al., 2019a). More detail on the security of blockchains is provided in Zhang et al. (2019).

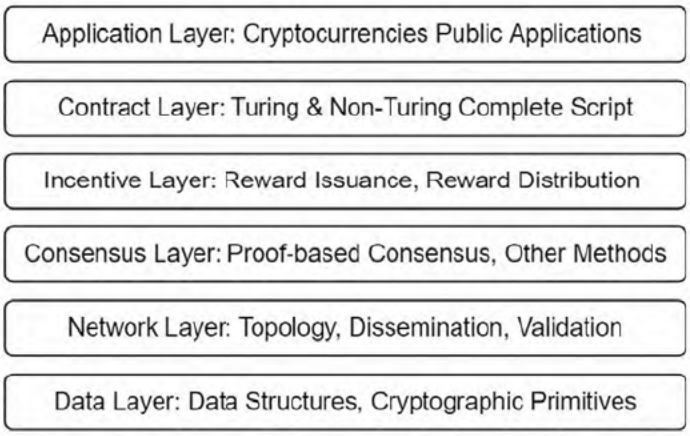

Fig. 2. Architecture of public blockchain.

Studying blockchain as from solely a technical perspective may be misleading due to the inherent socio-technical nature of the blockchain (De Domenico & Baronchelli, 2019). As the study of centralization in public blockchain is still fragmented, current conceptual models, such as security and privacy models, do not provide adequate insights. To overcome this limitation, we devise a novel centralization taxonomy focusing on the different architectural layers of blockchain to categorize centralization concerns. We employ a two-step research approach, first conducting a systematic literature review to construct a taxonomy of centralization, and refine this further through expert interviews.

2.2. Architecture of public blockchains

The first public blockchain, Bitcoin, incorporated the blockchain data structure and consensus mechanism in-depth, but omitted any formalization of the networking structure (Nakamoto, 2008). Since the introduction of Bitcoin, numerous attempts have been made to describe the structure of public blockchains more formally.

Some of these attempts have been aspect-specific with a microscopic focus on one or a few components of the blockchain. For example, Garay, Kiayias, and Leonardos (2015) describe the architecture of blockchain in terms of consensus mechanisms and participants of the network. Another notable description of blockchain architecture is given by Gervais et al. (2016), who focus on security and scalability by describing consensus and a peer-to-peer network.

Since the aim of our review is to analyze public blockchains more holistically to capture the factors causing centralization, we adhere to a more generic description of blockchain used by Zhang et al. (2019) and Zhu et al. (2019). In this generic description, the authors propose a layered architecture of blockchain. As a blockchain is a peer-to-peer distributed network, it is intuitive that blockchain systems will share many similarities with a generic, distributed computing architecture, such as the traditional OSI layered model of a network (Briscoe, 2000).

This layered architecture, illustrated in Fig. 2, describes how the data is stored (Data Layer) and shared (Network Layer) between different participants of the network. Once the data is shared with peers in the network, the network is tasked with agreeing a single view of the data (Consensus Layer). Public blockchains attain consensus in the network by incentivizing non-malicious participants using an incentive mechanism (Incentive Layer). Incentive and consensus operations are performed by the execution of computational scripts (Contract Layer). The computational capabilities of a blockchain are not just limited to these two operations; many different applications can be built on top of the blockchain such as cryptocurrencies and decentralized applications (DAPPS) (Application Layer) (Antonopoulos & Wood, 2018).

In the following subsection, we describe these layers in-depth:

2.2.1. Data layer

The data layer contains the definition of the data structure used by the system, including how transactions are stored, thus encompassing the transactions component proposed by Bonneau et al. (2015). Other data layer components include the cryptographic primitives employed on the blockchain. The network participants must adhere to the data layer specifications to participate in the network, i.e., use the same protocol to communicate. Application layer blockchain clients implement these specifications for the end-user.

2.2.2. Network layer

The network layer specifies the behavior of the nodes (network participants) in a distributed network. This behavior includes the network connection establishment and intercommunication mechanism. The network layer is responsible for the discovery of other nodes on the network and for efficient communication among nodes. The network layer serves as the information dissemination mechanism of the system. This network layer is identical to the network subsystem in the structure proposed by Judmayer, Stifter et al. (2017).

2.2.3. Consensus layer

Once the participating nodes are connected in a predefined topology, the next step is to generate blocks to contribute to the growing ledger. As all the participating nodes are tasked with the creation of the next block, it is crucial that the network can agree on a single state of the ledger. The aim of the blockchain network is to deterministically agree on a single state of the data. The consensus layer assures that the network reaches a consensus with a certain degree of assurance.

2.2.4. Incentive layer

This deterministic assurance in prominent consensus algorithms such as Proof-of-work, and Proof-of-Stake, is based on the assumption of an honest majority, i.e., the network has greater than 50% non-malicious participants. Blockchain systems use incentive engineering to ensure that the majority of the network is honest (Sai et al., 2019a). This incentive is often in the form of a block reward which is assigned to the node that successfully adds a new block to the blockchain. The incentive layer describes the mechanism used for issuance of reward and the distribution of reward. This layer acts as an interface between the user-facing layers and the technical implementation layers.

2.2.5. Contract layer

To process transactions in the network, Bitcoin uses a scripting language called *script* (Antonopoulos, 2017). This scripting language is significantly limited in terms of functionality as it lacks Turing completeness (Buterin et al., 2013). One example of this is the lack of loops in *Script*. Despite the lack of such functionality, the scripting language serves as the building block of Bitcoin cryptocurrency, enabling complex financial transaction processing.

The limitations on the scripting language of Bitcoin served as a motivation for Ethereum's developers (Wood et al., 2014). Ethereum implements a Turing complete computing engine on top of a distributed blockchain. Applications on top of the blockchain exploit this programmable nature of blockchain. The contract layer also acts as the interface between information systems and the blockchain (Beck et al., 2017).

2.2.6. Application layer

Public blockchains provide a mechanism that can be used to interact with and run user-defined code on the computing engine provided by the contract layer. JSON HTTP API is an example of one such public API provided by Ethereum (Lee, 2019). These public APIs serve as an interface between different Broker–Dealer services such as Wallets and Exchanges and the blockchain. These services are primarily used by end-users to interact with the blockchain (Chu, 2018).

2.3. Taxonomy development methodology

Classification of logically related objects is a fundamental problem in many disciplines. Taxonomies are considered an important tool to logically classify objects to better understand complex domains (Guerra García, Espinosa Torre, & García Gómez, 2008). The concept of taxonomy was initially proposed by Carolus Linnaeus (Lindley, 1836) to group organisms in Biology. Since then, taxonomies have been used in different knowledge domains such as social science (Bailey, 1994), computer science (Buckley & Exton, 2003), and information systems (Oberländer, Lösser, & Rau, 2019).

Due to the emerging nature of blockchain technologies, the state taxonomies in the field is preliminary. The most prominent of taxonomies in blockchain have been architecture (Xu et al., 2017) and security-specific (Zheng, Xie, Dai, Chen, & Wang, 2018). However, these taxonomies often treat blockchain as a single-dimensional computer science artifact whereas, as discussed earlier, the secure functioning of the blockchain-based assets is the result of the socio-technical nature of information systems. In the following subsection, we describe the state of the art in information system specific taxonomy formation and how our methodology aligns with this existing research.

Information system researchers have recognized the importance of taxonomies in knowledge organization. Specifically, Nickerson, Varshney, and Muntermann (2013) observed that despite the significance of taxonomies in information systems, the taxonomy development process remained largely ad hoc. To address this research gap, Nickerson et al. (2013) proposed a taxonomy development method specific to information systems. The development method proposed by Nickerson et al. (2013) has served as the guidelines followed by many information systems taxonomies (Oberländer et al., 2019). In this article, we follow the seven-step method proposed by Nickerson et al. (2013). We have illustrated the seven-step method in Fig. 3.

The first step of taxonomy construction is the determination of meta-characteristics. According to Nickerson et al. (2013), meta-characteristic is the most comprehensive characteristic that will serve as the basis for the choice of characteristics in the taxonomy. For example, if the researcher wants to classify a computer platform based on performance, the meta-characteristics are the hardware and software characteristics such as processing power, storage, and software optimization. Nickerson et al. (2013) also highlight the evolving nature of the meta-characteristic as many characteristics only become apparent through the taxonomy construction. In our taxonomy, we ground our meta-characteristic in the generic architecture described in Section 2.2.2.

After establishing the meta-characteristic, Nickerson et al. (2013) suggest the determination of ending conditions. As the taxonomy construction process is iterative in nature, it is crucial to establish end conditions. In the Nickerson et al. (2013) model, there are two types of end conditions: objective and subjective. For our taxonomy construction, we establish one objective and one subjective end condition. The objective end condition is the exhaustive examination and classification of all survey objects (aspects of centralization). The subjective end condition for our taxonomy is the determination of comprehensiveness through expert interviews.

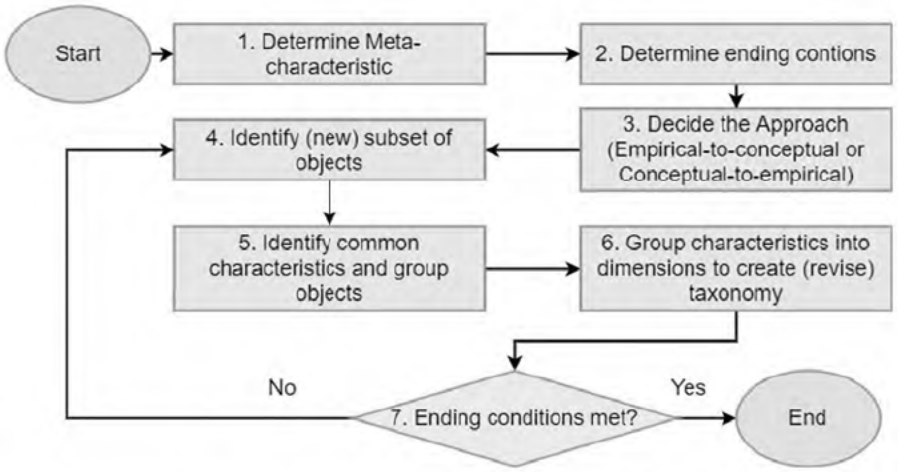

Fig. 3. Taxonomy construction methodology (Nickerson et al., 2013).

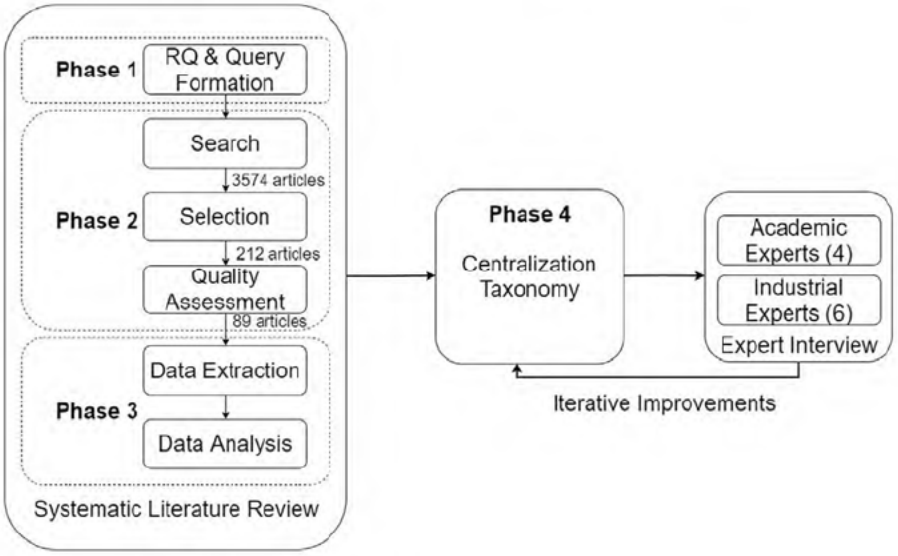

Fig. 4. Methodology.

Once we have established the meta-characteristic and ending conditions, Nickerson et al. (2013) propose two taxonomy construction approaches. In the first approach, conceptual-to-empirical, the researcher attempts to conceptualize the taxonomy dimensions without an exhaustive analysis of objects. The second approach, empirical-to-conceptual, relies on a review of the objects before the taxonomy constructions. This is often done in the form of a literature review. In our taxonomy construction, we follow the empirical-to-conceptual approach.

Within the empirical-to-conceptual model of taxonomy construction in the information system, there are three distinctive steps. In the first step, we identify a subset of objects. In our taxonomy, we identify new objects through a systematic literature review: Phase 1 and Phase 2 in Fig. 4. The second step is to identify common characteristics and group objects. We perform data extraction and analysis of the objects shortlisted through the systematic literature review to construct these grouped objects. This is done in Phase 3 of our methodology, as illustrated in Fig. 4. In the last step of the empirical-to-conceptual model, we group characteristics into dimensions to create or revise the taxonomy. For our taxonomy construction, Phase 4 attempts to construct a conceptual taxonomy that is refined through iterative cycles and structured via the architectural-layers lens.

In the following section, we describe our research methodology in detail.

3. Methodology

In this section, we describe the research methodology employed for our systematic literature review (SLR) of blockchain through which we sought to provide a more cohesive overview of centralization in public blockchains. We follow the SLR guidelines proposed by Kitchenham (2004) to identify the factors associated with centralization. We then use a classification scheme based on the generic

architecture presented in Section 2.1 to map the identified factors and associated measurement techniques. This mapping is loosely based on the approach proposed by Petersen, Feldt, Mujtaba, and Mattsson (2008). The mapping of obtained data to the generic architecture produces an initial taxonomy, which we then refined by conducting ten expert interviews to improve the taxonomy. This process is graphically illustrated in Fig. 4.

3.1. Systematic literature review

The systematic literature review guidelines suggested by Kitchenham (2004) span four phases:

- In the first phase, we define the two primary research questions for the review and produce relevant keywords for the subsequent search.
- In phase two, we systematically extract relevant articles from leading research repositories. We filter the resultant articles through a manual review of titles and abstracts.
- In phase three, the shortlisted articles are then used for data extraction, which is driven by an extraction protocol.
- In phase four, we perform the mapping of the data extracted from phase three to the generic architecture presented in Section 2.1, leading towards an initial taxonomy of centralization in public blockchains.

Fig. 5 illustrates the literature review employed in the study in more detail.

3.1.1. Phase 1: Research questions and query formation

The primary aim of our review is to provide richer insight into the different types of centralization present in public blockchain. We also identify techniques used to measure these aspects of centralization quantifiably. This will inform the development of our initial centralization taxonomy of public blockchains. We define the research questions of our study as follows:

- **RQ1:** What are the different aspects of centralization in public blockchains?
- **RQ2:** What techniques are employed to measure these centralization aspects?

Regarding RQ1, if a paper presented a novel centralization-causing factor, it is mapped to the architecture. If our generic architecture cannot accommodate the identified factor, we modify the architecture. This process is repeated for every novel factor identified. If a paper identified a factor already present in our taxonomy, we retain the reference to the article, using number-of-articles to define a proxy for the significance of that particular factor.

For every identified factor, we also recorded any measurement technique used to quantify the factor. If multiple papers employ different measurement techniques for a single factor, we retained all measurement techniques.

These research questions form the basis of article identification and selection, as they define the relevance of a particular article to our review. As we aim to capture factors from different socio-technical aspects of the blockchain, we conducted an exhaustive search on the following leading digital repositories: **Google Scholar, ACM Digital Library, IEEE Digital Library, ISI Web of Science, Science Direct, Scopus and Springer Link**. These repositories provided us with access to a wealth of articles, including gray literature.

Having identified the search repositories, we formed the search query. We adopted a systematic approach to keyword generation to form the search query:

1. **Initial set of keywords**: We formulated an initial set of keywords for the search consisting of "Blockchain" and "Centralization" with the following synonyms and alternate words:
   Blockchain: *bitcoin, ethereum, blockchain, cryptocurrencies, cryptocurrency, distributed ledger, DLT, Merkel tree, smart contract platform, tokenized asset.*
   Centralization: *centralization, centralism, consolidation, decentralisation, decentralization, devolution, dominating, domination, managed, monopolization, monopolization, monopoly, singular, unipolar.*
2. **Text Corpus Creation**: Complementary to the initial set of keywords, we also reviewed existing studies on centralization to extract more relevant keywords. We selected the two most cited relevant studies from Google Scholar (Gencer et al., 2018; Gervais et al., 2014). We performed forward, and backward snowballing on these two articles and generated a list of the most used keywords from this set. We selected the top 5 keywords from this set. This leads to the inclusion of *"digital currency"* and *"oligopoly"* to our initial set of keywords.

The resultant queries from query formation step are present in Appendix A.

3.1.2. Phase 2: Article search and selection

Given that decentralization is fundamental to a public blockchain, we expect that the search will return a high number of articles. We implement a filtering process to limit the search to relevant articles. We restrict our search to articles published in English after the introduction of Bitcoin in 2009. We refrain from treating citations as a proxy for quality to filter articles, as it has been questioned in the past (Galster, Weyns, Tofan, Michalik, & Avgeriou, 2013).

After the execution of a search query, Google Scholar returned the highest number of articles with 4380 results. However, due to the restrictions imposed by Google Scholar, we can only retrieve the first 1000 most relevant articles (Razzaq, Wasala, Exton, & Buckley, 2018). After applying the language and publication date constraints, we retrieved 982 articles from Google Scholar. We also

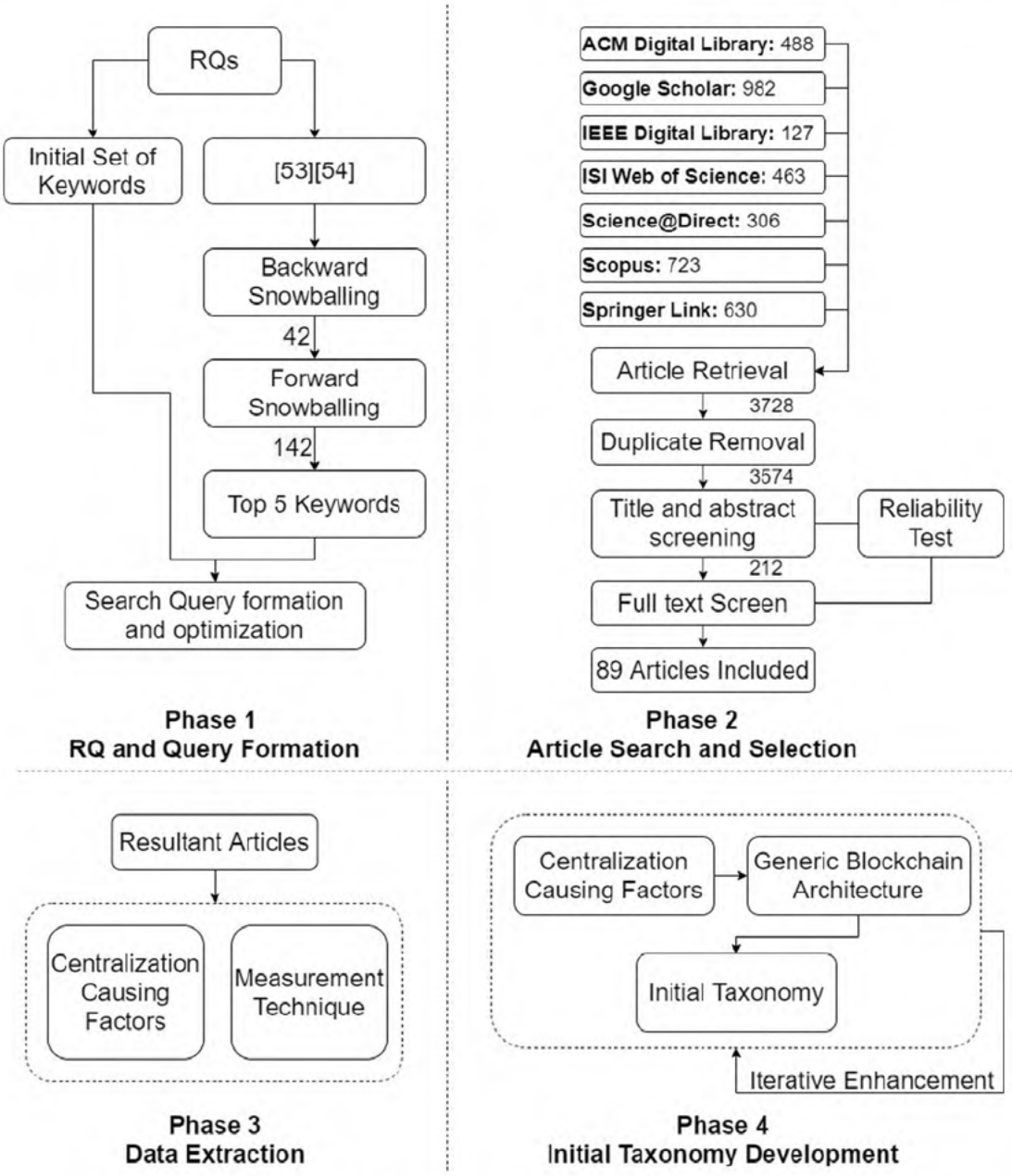

Fig. 5. Overview of systematic literature review.

Table 1
Quality assignment matrix.

Attribute	No	Yes
1. Centralization factor identified	0.0	1.0
2. Factor measurement technique proposed	0.0	1.0

retrieved additional 2737 articles from all other sources resulting in a total of 3728 articles. All of these articles were cross-checked to identify duplicate entries. After the removal of duplicate articles, the final set contained 3572 articles.[1]

Due to the high number of articles, we first analyzed the title and abstract to establish relevance. This was based on explicit inclusion criteria. The shortlisted, relevant articles were then scanned further to assign a quality score. These shortlisted articles

[1] A list of selected articles is available at www.github.com/ashishrsai/centralization.

were assessed for quality with regards to our research questions. To ensure that the assessment process is reliable, we followed the inclusion criteria for titling, abstraction, and full-text screening. This process obeyed the following inclusion criteria:

1. The paper's title mentions centralization, or any of the synonyms mentioned above, or is potentially relevant to the study of centralization.
2. The abstract is relevant to the identification or measurement of centralization-causing factors.

During the review of the title, we tried to avoid eliminating articles that might have some relevance to the topic of centralization. This relevance was evaluated by the review of the abstract. We excluded articles that did not pass both criteria.

The first author conducted this analysis. To test for reliability, we performed cross-validation by following Fleiss and Cohen (1973). We specifically use the guidelines proposed by Sim and Wright (2005) for the calculation of sample size. We select 89 articles with a confidence level of **95%** and a margin of error of **10%**. This sampling contained an equal number of accepted and rejected articles by the first author to eliminate the possibility of only sampling accepted or rejected articles. The second author was then tasked with the evaluation of these 89 articles based on the guidelines provided above. Results from the cross-validation suggest that both the reviewers were in almost perfect agreement over the acceptance and rejection of the articles with the Cohen's Kappa.[2] exceeding 0.8 (Landis & Koch, 1977).

Using this process, we retrieved 212 relevant articles for our study. Subsequently, we performed quality assessment of these articles by conducting full-text review. We assigned a quality score between 0 to 2 based on the relevance of the article to our research question. Table 1 outlines the assignment matrix employed for quality assessment.

We reviewed each article on two attributes - (1) factor identification and (2) measurement techniques used. If an article identifies a novel centralization-causing factor, we assign a score of 1.0 for Attribute 1. Articles that do not identify a novel centralization or refer to already identified factors are assigned a score of 0.0 for Attribute 1.[3]

We follow a similar quality assignment scheme for Attribute 2, where we assign a score of 1 for the identification of a novel measurement technique. Articles not proposing or using any existing measurement techniques are assigned a score of 0.0 for attribute 2.

To ensure that the quality assignment process is reliable, we again perform a similar reliability test but with a smaller data set of 9 articles. We observe that both the reviewers (first and fourth authors) agree on eight score assignments with one score difference for the ninth article. This disagreement is resolved when the article is reviewed by the third author.

This filtering process resulted in a set of 89 articles. These articles are used in the third phase of our study: Data Extraction.

3.1.3. Phase 3: Data extraction

Having identified relevant studies, the next step is to extract relevant data from them. For this purpose, we design a protocol to analyze the articles towards the development of an initial taxonomy of centralization. In this context, we focused on the factors identified and measurement techniques proposed or used. We reviewed all of the shortlisted articles to create a list of factors and associated measurement techniques. The extracted data from this step serves as a building block for our taxonomy.

3.1.4. Phase 4: Development of initial taxonomy

As we aim to structure the findings of the review in an initial taxonomy, we use the data extracted in Phase 3 and map it to appropriate layers in the generic blockchain architecture. We repeat this process for all identified factors; if a factor cannot reasonably be mapped to the existing layers, we typically refine the architecture by including an additional layer. This iterative refinement results in a blockchain architecture specific to the study of centralization. Results from this mapping analysis are illustrated in Fig. 6. Out of all shortlisted articles, 63 considered the consensus layer as prone to centralization, the highest reported count for any layer in our survey: This is represented in Fig. 6 by the size of the bubble, but we discuss these results in more depth in Section 4.

To further validate the initial taxonomy and refined architecture, we conducted interviews with industry and academic experts.

3.2. Interview with experts

The initial taxonomy, as referred to in Section 3.1, is based on the review of existing literature. To raise confidence that the initial taxonomy proposed by the study provides relevant coverage and is accurate, we further refine and validate it by interviewing experts.

To identify experts in the blockchain field, we relied on the epicenters of the bibliographic map generated by Ramona, Cristina, Raluca, et al. (2019). We approached 112 active researchers based on their prominence determined by their location on the bibliographic map. Out of 112 researchers approached for the study, we received a response from 10 and subsequently interviewed them. We interviewed four academic experts (I_1 to I_4) and six experts from industry (I_5 to I_{10}). Interviews were typically one hour in duration and involved open-ended questions[4] These open-ended questions were designed to:

[2] Cohen's kappa is a statistic measure of the agreement between two raters based on the classification of items in mutually exclusive categories.

[3] Although we do record the paper, as this helps us identify the significance of that centralization aspect in the literature.

[4] The interview script is available at www.github.com/ashishrsai/centralization.

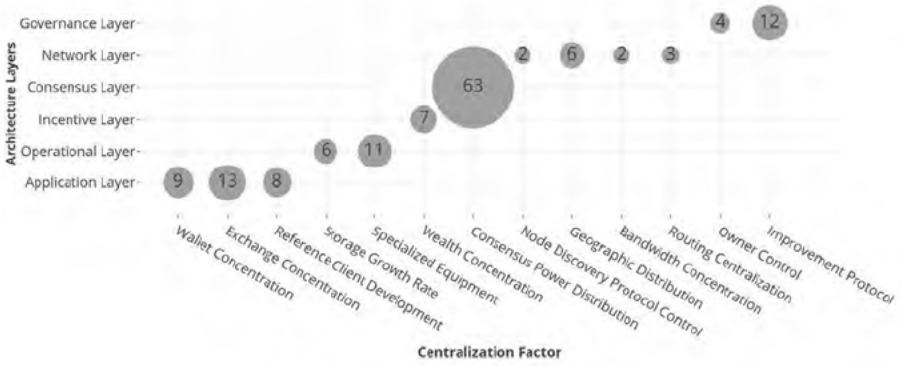

Fig. 6. Article titling and abstraction process.

1. Extract the view of the expert on centralization and the significance of it in their respective field, i.e., security, economics, information systems, and industrial application.
2. If needed, refine the taxonomy and/or the architecture.
3. Validate the generic architecture of the blockchain used in this study (Section 2).
4. Assess the accuracy of the initial centralization taxonomy.

The transcripts of these interviews are available in anonymized form[5] These transcripts are color-coded based on the relevance of the conversation to factor identification and measurement.[6]

3.3. *Illustrative walk-through of the four phases*

Thus far in this Section, we have described the four phases used for taxonomy development and refinement. In this subsection, we present an explanatory walk-through of 2 articles through these phases. For this illustration, we select the following two articles: Gencer et al. (2018) and Peck (2017).

In phase one, we formulate the search query through an initial set of keywords and snowballing on seminal work in centralization. Gencer et al. (2018) is one of the two articles used for snowballing and keyword formulation due to the high citation count. After constructing the query, we move to phase 2: executing the query and shortlisting the appropriate articles.

During title screening in phase 2, after performing the search across the academic databases, Gencer et al. (2018) is included for abstract screening as the title points to the state of decentralization. Likewise, the second illustrative article, Peck (2017), is also shortlisted as the title refers to the difference between different blockchain forms.

In the abstract screening step, the article Gencer et al. (2018) is considered relevant because the abstract makes direct reference to the state of decentralization. The second article, Peck (2017), is also shortlisted as the abstract points to the risk of limiting controlling power to a select few participants.

Both the articles are now evaluated for quality by conducting a full-text review. In the first article, Gencer et al. (2018) describe the fundamentals of centralization on the network layer in blockchain, identifying novel centralization causing factors, and suggesting novel measurement techniques. Following the quality assessment matrix in Table 1, we assign a quality score of 2 to Gencer et al. (2018).

The full-text analysis of Peck (2017) reveals that the article does not identify or measure any centralization causing factor, therefore obtaining a quality score of 0. This article is henceforth excluded from taxonomy formulation.

Having identified the relevant articles, we perform data extraction in Phase 3. In data extraction, we first extract all the factors identified by Gencer et al. (2018): Consensus Power Distribution (Section 4.3), Geographic Distribution (Section 4.2.2), Bandwidth Concentration (Section 4.2.3) and Routing Centralization (Section 4.2.4). After identifying the centralization causing factors, we extract the measurement techniques used or suggested in the article: The authors proposed using a percentage-based value for Consensus Power Distribution, a latency-based measurement for identifying the geographic location for participating nodes, clustering for bandwidth concentration, and using AS (autonomous systems) coverage as a metric for routing centralization (Section 4.2.4).

After extracting the centralization causing factors and measurement techniques, we move to phase 4, constructing the initial taxonomy. In our representative example article, Gencer et al. (2018) have identified four centralization-causing factors. In this step, we venture to map these four factors to the generic blockchain architecture described in Section 2. The first centralization causing factor, consensus power distribution, is mapped to the consensus layer as this factor is within the layer's scope. The remaining three centralizations causing factors are all related to the networking aspects of the blockchain network. The geographic distribution

[5] The transcripts can be obtained from www.github.com/ashishrsai/centralization.

[6] More details on the coding scheme provided in Appendix C.

Table 2
Taxonomy of centralization in public blockchains.

Layer	Centralization factor	Measurement techniques
Application layer	Wallet concentration	Not found
	Exchange concentration	Centrality & Percentage value
	Reference client concentration	Satoshi index
Operational layer	Storage constraint	Ratio of growth
	Specialized equipment concentration	Not found
Incentive layer	Wealth concentration	Gini coefficient & Percentage value
Consensus layer	Consensus power distribution	Percentage value & Gini coefficient & Theil index & Centralization factor
Network layer	Node discovery protocol control	Not found
	Geographic distribution	Gini coefficient & Latency
	Bandwidth concentration	Clustering of provisioned bandwidth
	Routing centralization	AS-Level coverage
Governance layer	Owner control	Fractional measurement
	Improvement protocol	Centrality metrics

results from the open peer-to-peer network topology, whereas the routing, and bandwidth centralization target the network layer's information dissemination aspect.

After mapping these factors and their measurement techniques to the architecture, we construct an intermediate form of the centralization taxonomy. This intermediate form is iteratively refined as we process more articles through the four-phased approach. The resultant taxonomy is described in-depth in the following Section.

4. Taxonomy of centralization of public blockchain

In this Section, we map the results of the systematic review, and the interviews with experts, to the initial taxonomy of centralization outlined in Table 2.

As discussed in Section 3.1, this generic architecture is refined to reflect the centralization-related aspects of the blockchain better. To this end, we refined the generic architecture by removing the Data and Contract layers as none of the surveyed articles suggested any centralization aspects for either of these layers. As can be seen from Table 2, on average two centralization factors were identified for each resultant layer. As is also presented in the table, there are some factors for which there are no proposed measurement techniques (for example 'Wallet Concentration'). We also note that the existing generic architecture was unable to capture governance-related aspects of the blockchain system. For example, as blockchain systems evolve, it is crucial to have a mechanism to handle improvements such as security patches of the system. We account for the governance-related aspects of centralization by including a Governance Layer.

Another set of centralization causing issues that the generic architecture does not capture are associated with the operation of a node on the network. These issues include the computational requirements for participation, such as proprietary hardware and storage. In accordance with the recommendation of interviewee I_{10}, we include an Operational layer to represent the centralization associated with operating as a node on the blockchain.

Table 3 considers the factors identified in Table 2 from the perspective of 'prevalence-of-occurrence' in the literature and the interviews, where prevalence is considered as a proxy for whether the factor is "established" or not. The literature references in the table identify that particular factor as a potential source of centralization.[7] The interviewer identifiers are used to indicate explicit recognition of the factor as a contributor to centralization in the associated interview. Interestingly, based on the data presented in this table, most of the factors can be considered well established, with the possible exception of Bandwidth Concentration and Routing Centralization. Even though Node Discovery Protocol Control was only referred to by one academic article, the majority of interviewees perceived it as a relevant factor.

Based on our taxonomy, we define centralization of public Blockchains as *the process by which one or more architectural dimensions (aspects) of the Blockchain are restrictive to the majority of participants by direct or indirect economic, social, or technical constraints.* We report a total of 13 aspects spread over six architectural layers. The governance layer aims to capture the social constructs of building and maintaining a public blockchain, specifically reporting on the incentives to build (Owner Control) and maintain a public blockchain (Improvement Protocol). The governance layer feeds into the economic aspects of the Blockchain in forms of incentives, this is captured by the Incentive layer, where we review the wealth inequality (Wealth Concentration). This inequality is in part caused by the technical constraints of participation ranging from Networking aspects such as bandwidth and routing requirements to operational requirements such as storage and specialized pieces of equipment for participation. These higher storage and specialized equipment requirements restrict participation in the consensus, which is observable in the consensus layer. We also report on the centralization of end-user applications such as wallets and exchanges. The following subsections discuss the taxonomy in detail.

[7] A complete list of articles is available in Appendix B.

Table 3
Centralization causing factors found in literature and interviews.

Centralization factor	Refereed articles	Interviews
Wallet concentration	R11, R13, R36, R40, R76, R78, R84, R86, R88	I_4, I_5, I_7, I_8
Exchange concentration	R11, R13, R27, R34, R37, R40, R57, R64, R73, R78, R84, R86, R89	I_1, I_3, I_4, I_5, I_7, I_8
Reference client concentration	R4, R6, R8, R26, R36, R50, R67, R83	I_2, I_5, I_8, I_9, I_{10}
Storage growth rate	R9, R24, R38, R39, R63, R80	I_2, I_{10}
Specialized equipment concentration	R23, R51–R53, R55, R62, R67	I_4, I_5, I_7, I_8, I_9, I_{10}
Wealth concentration	R16, R51, R52, R55, R62, R67	I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_9
Consensus power distribution	R1–R3, R5, R7, R9, R11–R17, R19–R22, R25, R26, R28–R33, R35, R36, R39, R40, R42–R47, R49, R52–R56, R58, R60, R61, R65–R72, R74–R79, R81, R82, R87	I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9, I_{10}
Node discovery protocol control	R59	I_1, I_2, I_3, I_5, I_{10}
Geographic distribution	R5, R30, R40, R47, R50, R76	I_1, I_2, I_3, I_4, I_5, I_6, I_7
Bandwidth concentration	R35, R87	I_2, I_{10}
Routing centralization	R3, R20, R35	I_2
Owner control	R14, R18, R26, R41, R48	I_1, I_4, I_5, I_7, I_8, I_9
Improvement protocol	R4–R6, R10, R26, R36, R41, R48, R76, R83, R85	I_1, I_2, I_3, I_4, I_5, I_7

Table 4
Categories of centralization in governance layer.

Ref	Owner control		Improvement protocol	
	Identification	Measurement	Identification	Measurement
R4	×	×	✓	×
R5	×	×	✓	×
R6	×	×	✓	Centrality metrics
R10	×	×	✓	×
R14	✓	×	×	×
R18	✓	Fractional measurement	×	×
R26	✓	×	✓	×
R36	×	×	✓	×
R41	✓	×	✓	×
R48	✓	×	✓	×
R76	×	×	✓	×
R83	×	×	✓	×
R85	×	×	✓	×

4.1. Governance

Blockchain, like any other information system, is subject to evolutionary changes that are governed by a governance structure. These evolutionary changes may include security patches, scalability provisions, and improvement proposals. Wang et al. (2017) theorizes the relationship between the value proposition of blockchain and the governance structure in place. They reason that the core value proposition of blockchain is rooted in decentralization. This property of decentralization is considered valuable by investors.

Decentralized governance was also indicted as a vital component of public blockchains by our interview participants. 80% mentioned governance as a significant centralization threat (I_1, I_2, I_3, I_4, I_5, I_7, I_8, I_9). This is best illustrated by a quote from I_1, with respect to the implication of centralized governance structure: *"if you are talking about the centralization of governance, that for me is the prime example of a private permissioned Blockchain"*.

Despite the significance of decentralization for blockchain, Wang et al. (2017) argue that a high level of decentralization may slow down the strategic decision-making process. Contrary to the proposition in favor of some centralization by Gervais et al. (2014) and Wang et al. (2017) argue against the concentration of decision making power by pointing out instances of unilateral decision making by core developers in the short history of bitcoin; for example, when the core developers unilaterally decided to lower the minimum transaction fee. This criticism of governance centralization is shared by Roubini (2018a) who criticizes the centrality of control over governance as it may concentrate the decision power to a few entities involved in governance of the blockchain. Atzori (2015) expands the analysis of blockchain governance issues towards the emergence of blockchain governance oligarchy. Azouvi et al. (2018) conducts an empirical analysis of two of the most prominent blockchain projects, Bitcoin and Ethereum, by comparing the state of governance to other major open-source projects. They conclude that control governance is usually concentrated in a handful of people in Bitcoin and Ethereum, which is a big centralization factor.

As reported by Wang et al. (2017), the centralization on the governance layer may not be detrimental due to the advantages of rapid strategic decision-making. We expand on the argument in favor of some centralization (Wang et al., 2017) in Section 6, where we discuss how the adverse impact of centralization varies across the different layers of the taxonomy.

Based on the literature review and subsequent interviews, we further divide the issue of governance into *owner control* and *improvement protocol*. These results are presented in Table 4.

4.1.1. Owner control

As described by Wang et al. (2017), the developers of the blockchain often retain some control over the implementation on the governance level. This can be in the form of, for example, the native cryptocurrency owned by the developers. Wang et al. (2017) describes this as *Owner Control*.

Measurement Technique : This type of owner control can be measured by examining the total cryptocurrency accumulated by the owners in the early adoption period (Wolfson, 2015). This early adoption period also includes the pre-mined[8] cryptocurrency (Wang et al., 2017). We report studies such as (Chohan, 2019; Wolfson, 2015) that have implemented a proportional measure to quantify owner control. Owner control can be measured as the fraction of the total allowed cryptocurrency if the supply is capped, as measured by Eq. (1), where $C_{OwnerControl}$ represents the fraction of total cryptocurrency that the owner controls.

$$C_{OwnerControl} = V_{OwnerBalance}/V_{CappedSupply} \tag{1}$$

If the supply is uncapped, owner control is measured as the fraction of total currency in circulation, as illustrated in Eq. (2).

$$C_{OwnerControl} = V_{OwnerBalance}/V_{CurrentSupply} \tag{2}$$

Most interview participants indicated that the use of fractional measurement for owner control was appropriate. However, I_9 suggested a refinement: *"The fractional calculation of the owner control varies with the supply; a simpler approach might be to use a metric such as how much power over the network can be achieved with the money in the owner control. How much hardware can you afford, and what hash power can you get with it. Relating the cryptocurrency to the hashing power would be more informative"*.

Implication of high owner control : Depending on the consensus mechanism used, the owner control has severe impacts on the network. This adverse impact is particularly worrying in the case of Proof-of-stake based cryptocurrency, where the consensus power is determined by the quantity of native cryptocurrency owned by the participant. Having a large amount of pre-mined or early adoption period accumulated cryptocurrency will give the owner a significant advantage over others, resulting in a more centralized network. This high consensus power pose a security threat as an owner with over 50% consensus power can conduct a double spending attacks. Ethereum is a prime example of such wealth concentration due to pre-mined cryptocurrency.

The Ethereum platform was crowdfunded by investors who were rewarded in the form of ETH[9] during the creation of the first block in Ethereum. An estimated 60 Million ETH were distributed among the early investors; another 12 Million were distributed among the developers of Ethereum (Etherscan, 2019a). We calculate the value of $C_{OwnerControl}$ by considering the 12 Million pre-mined ETH that developers control and the total current supply of ETH obtained (from Etherscan, 2019b):

$$C_{OwnerControl} = 12,000,000/106,514,407.78 = 0.11 \tag{3}$$

It should be noted that the value of $C_{OwnerControl}$ feeds into the issue of Wealth Concentration, which is a significant cause of economic centralization. A high wealth concentration in a cryptocurrency is against the founding principle and premise of cryptocurrency providing a more even monetary system. This can consequently disincentivize the adoption.

4.1.2. Improvement protocol

As discussed earlier, evolutionary changes require blockchains to have a robust governance structure in place. As decentralized blockchains do not have any authorized entities moderating the changes, the process of moderation is delegated to the participants. Bitcoin improvement protocol (BIP) is a prime example of such an improvement system (Anceaume, Lajoie-Mazenc, Ludinard, & Sericola, 2016). The formal voting protocol, such as that in BIP, is used to establish consensus over proposed changes, often through voting. 60% of interview participants (I_1, I_2, I_3, I_4, I_5, I_7) mentioned that the improvement protocol performs an essential function in the network with I_7 suggesting: *"Whoever controls the improvements will inevitably shape the future of the network"*.

The literature review points out the similarities between the Python Enhancement Proposals and BIPs, both of which heavily draw from the *"canonical"* approach to consensus (De Filippi & Loveluck, 2016). In the *"canonical"* based BIP, all the suggested changes have to be made available to the public for open discussion. However, the final decision as to how proposed changes will be implemented is taken by the core developers (Gervais et al., 2014).

Measurement Technique: The centralization in a formal voting protocol is measured by analyzing the moderation control. If specific developers or owners can moderate the voting, the moderation may jeopardize changes these that developers or owners disagree with. Thus the determination of the control level is done by examining the voting protocol in place and the controls imposed on it.

As public blockchains often have an open platform for proposing improvements, such as BIP for Bitcoin, and EIP for Ethereum, Azouvi et al. (2018) suggests reviewing the number of improvement proposals made by each author and the respective states of those proposals (i.e., approved, rejected or under review). The authors also suggest reviewing the comments on each proposal to examine the discussions. Based on the data obtained from the author/number of proposals, complemented by comments per author on the proposal, Azouvi et al. (2018) suggests calculating metrics for centralization measurement. Fig. 7 illustrates this measurement technique graphically.

[8] Pre-mined cryptocurrency refers to the native cryptocurrency issued with the creation of the first block in the blockchain.

[9] ETH is the ticker mark for Ether, the cryptocurrency used by Ethereum platform.

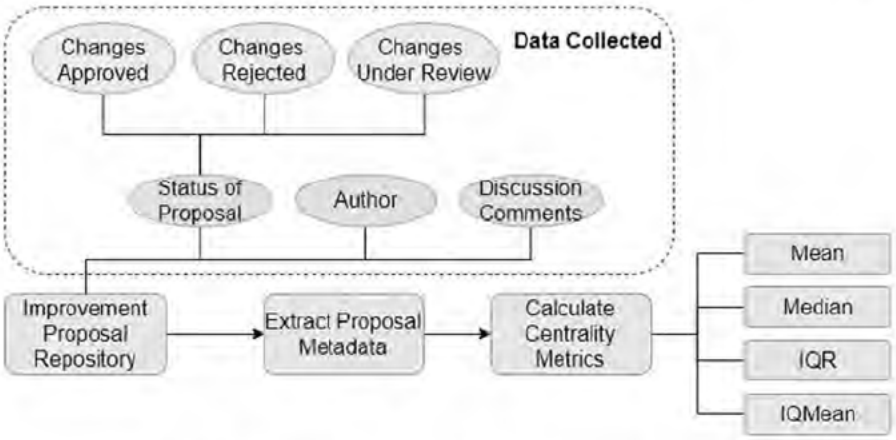

Fig. 7. Improvement protocol centralization measurement technique.

These centrality metrics include Mean, Median, interquartile range (IQR), and interquartile mean (IQMean). IQR is a measure of variability that assists in locating where the majority of values lie in the data sample. It is calculated as the difference between 75th and 25th percentiles of the data. However, IQR is sensitive to noisy outliers, which can impact the overall result. This can be overcome by using the IQMean, which allows us to eliminate the outliers from our data set by calculating the median of IQR.

Implication of control over improvement protocol: If a subset of all participants moderate the improvement protocol, it will result in control over improvements or modifications to the network. The debate over block size in Bitcoin an example of an issue arising due to this type of control over the network (Bitcoin, 2019; De Filippi & Loveluck, 2016). Other significant control implications over the improvement protocol include the unilateral decision making in both Bitcoin and Ethereum, where the governance structure implemented a change not widely supported by the community. This includes the notable transaction fee reduction in Bitcoin (Gervais et al., 2014) and Ethereum hard fork due to DAO attack which led to the subsequent creation of Ethereum classic (Wirdum, 2016). More incidents of unilateral decision making include the changes to the Ethereum consensus algorithm in 2018, where developers decided to modify the algorithm to disable newer mining hardware (Kim & Zetlin-Jones, 2019). These incidents not only represent the lack of a systematic governance model in terms of improvement but also present a challenge in terms of newer participation and updates. This type of centralization impacts the presumed open nature of the Blockchain, which is one of the core contributions of Blockchain to the field of financial technologies.

4.1.3. State-of-the-art for centralization on the governance layer

Based on the literature review, we report that there are two distinctive approaches to centralization in governance. In the first approach, pioneered by Wang et al. (2017), governance centralization is essential for rapid strategic decisions. This approach is countered by the empirical analysis of Azouvi et al. (2018). Those authors report that other non-cryptocurrency open source projects have attained a higher level of decentralization in governance, specifically in the form of improvement protocol without the need for centralized control. These two contrasting approaches highlight the need for a more in-depth analysis of the importance of rapid strategic decision making in the context of blockchain various other open-source projects.

4.2. Network

The network layer acts as the information dissemination mechanism for the blockchain instance. As the decentralized network cannot have centralized nodes that act as relay points to transmit messages between the participants, the network is largely a peer-to-peer system. The network layer acts as the information dissemination mechanism for the blockchain instance. This peer-to-peer network serves as an essential security and usability measure as pointed out by I_8: *"In this peer to peer network, there is no single point of failure and participants can join and leave the network without risking interruption or degradation of the network"*.

Network connectivity of a node is an important aspect of performing the mining operation (Sapirshtein et al., 2016). Higher network connectivity results in a higher likelihood of adding the next block on the longest chain as the miner can propagate the block to a large number of nodes in the network. This interplay between the reward from adding a block to the blockchain and network connectivity has resulted in networking phenomena such as strategizing networking resource concentration in the form of bandwidth (Gencer et al., 2018) and strategizing geographic distribution of nodes in the network (Kim et al., 2018; Roubini, 2018b).

Based on the literature review, we identify another source of centralization on the network layer as the topology formation of the network. This formation includes the node discovery protocol for finding peers in the network (Neudecker & Hartenstein, 2018) and the routing structure of the network (Apostolaki, Zohar, & Vanbever, 2017). The relevant studies identified by our review are presented in Table 5. We describe each of the outlined factors in detail in the following Subsections.

Table 5
Categories of centralization in network layer.

Ref	Node discovery		Geographic distribution		Bandwidth		Routing	
	Identification	Measurement	Identification	Measurement	Identification	Measurement	Identification	Measurement
R3	×	×	×	×	×	×	✓	AS coverage
R5	×	×	✓	×	×	×	×	×
R20	×	×	×	×	×	×	✓	×
R30	×	×	✓	×	×	×	×	×
R35	×	×	✓	Latency	✓	Clustering	✓	AS coverage
R40	×	×	✓	×	×	×	×	×
R47	×	×	✓	×	×	×	×	×
R50	×	×	✓	×	×	×	×	×
R59	✓	×	×	×	×	×	×	×
R76	×	×	✓	×	×	×	×	×
R87	×	×	×	×	✓	×	×	×

4.2.1. Node discovery protocol control

In a peer-to-peer topology, participating nodes directly communicate with other participants to transmit data packets. A node discovery protocol is used to discover nodes in the network with which to communicate (Miller et al., 2015). The node discovery protocol often relies on a set of seed DNS nodes that distribute the address of other active nodes on the network. These predefined DNS nodes may be a potential source of security threat, as demonstrated by Jin, Zhang, Liu, and Lei (2017) and Tapsell, Akram, and Markantonakis (2018). If one of the seed nodes becomes inaccessible, it may result in many participants of the network becoming undiscoverable. As the new nodes in the network discover others by querying these predefined seed DNS nodes, the literature identifies seed nodes as a contributor to centralization on the network layer (Neudecker & Hartenstein, 2018).

Measurement Technique : After the review of all relevant articles in our study, we conclude that no measurement technique focuses on the Node Discovery protocol. Studies such as (Jin et al., 2017; Tapsell et al., 2018) investigate the issue of seed DNS nodes from a security perspective, specifically focusing on the single point of failure issue. We reason that further investigation into centralization in node discovery level is warranted due to the significant security threats that it poses.

Implication of control over DNS: Centralized DNS services are linked to security threats in the network (Jin et al., 2017). They also allow the DNS owners to observe the participants of the network. These centralized DNS services can also act as a single point of failure, which is of particular concern in the case of a Denial of Service attack (Dietrich, Long, & Dittrich, 2000). As core developers select these DNS nodes, the issue of node discovery protocol also feeds into that of trust in the core developers (Tapsell et al., 2018). A malicious developer can also change the DNS seed nodes to conduct an eclipse attack. Several Monte Carlo simulations have shown the effectiveness of such eclipse attacks on Bitcoin and Ethereum (Heilman, Kendler, Zohar, & Goldberg, 2015).

4.2.2. Geographic distribution

Bitcoin and similar cryptocurrencies have been able to gain significant attention from governments around the world due to their decentralized uncensored nature. This has prompted many to argue that a significant concentration of the nodes in any geographic area may be a threat to the network (Roubini, 2018b). This type of geographic concentration may lead to centralization on the network layer as the nodes become prone to geopolitical manipulation. 70% of interview participants indicated that geographic concentration is harmful to the network. I_6 suggested that geographic centralization may be disadvantageous for miners who are not centrally located: *"I fear that in a geographically-focused network, people within the same geographic location will have an edge over others, they will receive and send transactions first"*.

The nodes are distributed over the participating countries in the network. In an ideal case, the distribution of nodes should be equal in all participating countries so as to be able to withstand a geopolitical blockade. Findings from our review suggest there is a trend towards geographic concentration of nodes in both Bitcoin and Ethereum (Gencer et al., 2018; Khairuddin & Sas, 2019; Kim et al., 2018; Roubini, 2018b).

Measurement Technique : Our review suggests that the geographic location measurement in blockchain can be done by measuring latency in the peer-to-peer network (Gencer et al., 2018; Kim et al., 2018). This approach draws heavily from Saroiu, Gummadi, and Gribble (2001), where the authors proposed using latency as a measurement tool in Gnutella. Gencer et al. (2018) first proposed measuring the distance between their geographically distributed nodes and other peers in the network by sending a data packet and measuring the round-trip time. Based on the round-trip time, Gencer et al. (2018) calculated upper and lower bounds between two remote peers in the network. If two nodes take a similar time to respond to the data packet sent by their nodes, it is reasoned that these two nodes are likely geographically close. This approach is further refined by Kim et al. (2018), who consider the average of bounds for final latency estimation.

Fig. 8 illustrates this measurement technique graphically using a toy example. In this example, we have two geographic regions A and B. To identify a blockchain network participant's relative geographical locations, we deploy two measurement nodes that send a data packet to the network participant and wait for the response. Upon the receipt of a response, we can calculate the network latency. In this toy example, the measurement node in geographic area A returns a lower latency; thus, we can assume that the blockchain participant is geographically closer to area A than area B. In their analysis (Gencer et al., 2018), the authors conduct this experiment with a large number of measurement nodes spread out geographically.

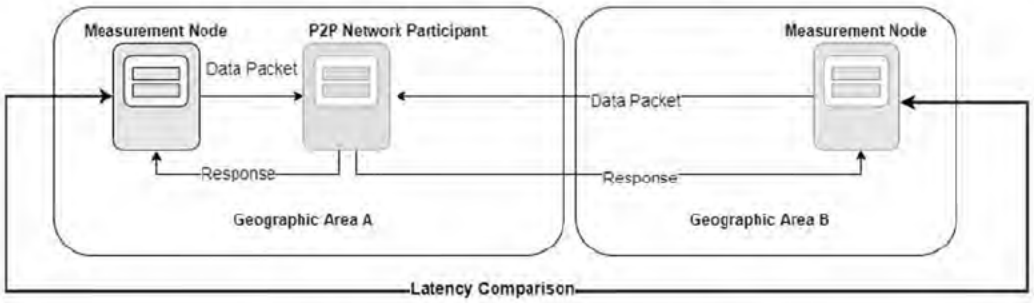

Fig. 8. Geographic distribution measurement technique.

Implications of geographical centralization: The most prominent issue with geographic centralization is the potential for geopolitical manipulation of the network (Roubini, 2018b). Other issues with geographic clustering include the possibility of faster transmission of packets to nearby nodes promoting faster network propagation. This can lead to more clustering, since participant must propagate the solution to the majority of the network in order to get rewarded in Proof-of-work based blockchains. If the majority is located in a geographical cluster away from the participant, that may translate to a loss of revenue. As suggested by Gencer et al. (2018), a low number of geographic clusters are considered good for the decentralization of the network. This is due to the association of potentially high block rewards due to faster network propagation. As shown in Sapirshtein et al. (2016), network connectivity is directly related to the ability to successfully conduct selfish mining attacks, which can support a double spending attack.

4.2.3. Bandwidth concentration

In a public blockchain's peer-to-peer network, the network bandwidth often acts as a crucial factor in the successful propagation of data packets. In Proof-of-Work based blockchain, every consensus cycle acts as a race to first calculate the solution to the cryptographic puzzle followed by dissemination of the solution to a majority of the network. Dissemination requires a large number of network connections with peers in the network, thus increasing the bandwidth requirements. This arms race to attain higher bandwidth may lead to the centralization of mining equipment to services like a centralized data center with high bandwidth (Gencer et al., 2018).

Measurement Technique : Gencer et al. (2018) proposed measuring the bandwidth of each peer by requesting a large amount of data and estimating the speed by observing the time taken for the transmission. Once they estimate the speed of each accessible peer, they calculate and cluster the provisioned bandwidth in groups.

Implication of bandwidth concentration: A high bandwidth requirement may limit the participation to only the participants with significant bandwidth (Zheng et al., 2018). It may also result in a high concentration of networking devices in centralized spaces such as data centers (Gencer et al., 2018). This potential increment in bandwidth requirement may limit the participation to only those entities with high network capabilities making the consensus participation not viable in a domestic setting. The inability to participate in the network violates the open nature of the public blockchain preventing a widespread adoption of the technology.

4.2.4. Routing centralization

As public blockchain networks run over the existing networking stack, they rely on the networking structure used by IP (Internet Protocol). Centralization present in the networking structure of IP transfers to the blockchain as well. Our review reports that this centralization has been studied in blockchain from the privacy (Feld, Schönfeld, & Werner, 2014) and security (Apostolaki et al., 2017) perspectives. Gencer at al. (2018) reports that concentration on AS-Level[10] as a source of centralization for a public blockchain (Gencer et al., 2018). Interestingly, none of the industrial participants mentioned this concern unprompted, suggesting that it might be more of an academic concern than a real-world one. However, when the concern was mentioned, one industry participant agreed.

Measurement Technique : Our review suggests that there is a common network traversing strategy used to determine the network structure from the AS-Level perspective (Apostolaki et al., 2017; Feld et al., 2014; Gencer et al., 2018). To measure the number of ASes in a peer to peer network, the observer node traverses the network by recursively collecting IP addresses of each peer and querying every reachable address. This process is repeated until no new reachable nodes are available in the IP list. For the determination of AS of each IP, Feld et al. (2014) recommend using Maxmind's free Geo API.[11]

Implication of control over ASes: Centralization on AS-Level is reported to have privacy implications for blockchain users as it allows more traceability on a network level (Feld et al., 2014). This concentration of IP addresses under a few ASes is directly linked with potential network security issues in Bitcoin (Apostolaki et al., 2017) and Ethereum (Gencer et al., 2018). However,

[10] Autonomous systems (AS) in computer networks refers to the collection of connected IP routing prefixes under the authority of one or more networking entities.

[11] https://dev.maxmind.com/geoip/geoip2/geolite2/.

Table 6
Categories of centralization in consensus layer.

Consensus power distribution	Selected studies
Identification	R1, R2, R3, R5, R7, R9, R11, R12, R13, R14, R15, R16, R17, R19, R20, R21, R22, R25, R26, R28, R29, R30, R31, R32, R33, R35, R36, R39, R40, R42, R43, R44, R45, R46, R47, R49, R52, R53, R54, R55, R56, R58, R60, R61, R65, R66, R67, R68, R69, R70, R71, R72, R73, R74, R75, R76, R77, R78, R79, R80, R81, R82, R87
Factor measurement	
Percentage based measure	R1, R7, R12, R14, R21, R26, R29, R31, R33, R35, R36, R43, R46, R47, R49, R49, R53, R55, R56, R60, R61, R67, R71, R72, R73, R77, R78, R80
Gini	R15, R16

these privacy and security threats remain largely academic with no real world incident reports in our sample set of articles. This is further evident through our interviews, where no academic or industrial experts pointed to control over ASes as a centralization threat unprompted.

4.2.5. State-of-the-art for centralization on the network layer

In a peer-to-peer, network-based blockchain system, both the network connectivity and the network capabilities have an impact on the likelihood of profit (Sapirshtein et al., 2016). Our survey reports on four types of network-based centralization: node discovery, geographic distribution, bandwidth, and routing. Among the reported centralization avenues, there are no measurement techniques to quantify node discovery protocol centralization despite the security threats associated with centralization (Jin et al., 2017). We reason that a further investigation into the centralization of node discovery protocol is warranted. We also report that the research into the geographic distribution and bandwidth centralization is primarily focused on Bitcoin and Ethereum. Due to the association of the monetary reward and the network connectivity and capabilities, we reason that a further empirical investigation into the network layer for more cryptocurrencies may assist in better understanding network participation's profitability.

4.3. Consensus

The consensus layer establishes an agreement on a single state of the data in the public blockchain. As described in Section 2.2, in the case of Proof of Work, it is attained by inducing a race to solve a mathematical problem. The first person to solve and propagate receives a monetary reward as an incentive. The likelihood of finding the solution to the mathematical problem depends on the computational power devoted to the solution. Thus a high concentration of computational power is a direct signifier of centralization in the blockchain. As identified by articles in Table 6, the **consensus power distribution** is a key contributor to the centralization of the Proof-of-Work based blockchain. Eight interviewees mentioned this aspect unprompted, suggesting that this is a prevalent concern. In this subsection, we review how the literature defines and measures the consensus power centralization.

4.3.1. Consensus power distribution

In the case of a Proof-of-Work based blockchain, the Consensus power is also known as the hash power of the miner (participating node). The centralization of hash power can pose a significant security threat to blockchain solutions such as Bitcoin and Ethereum. One key contributing factor to centralization is commercial mining pools. The income from mining operations depends on the probability of finding and propagating the solution of the puzzle before everyone else. The probability of successfully calculating the solution depends on the hash power of the computing device used for the calculation. Lower probability leads to a lack of stable income and may prompt users to mine as a group and share the profit. This group mining is also known as pooled mining (Lewenberg, Bachrach, Sompolinsky, Zohar, & Rosenschein, 2015). Based on the analysis of the shortlisted literature, we report that the concept of pooled mining in itself is not considered a threat to the decentralization of the network; however, the literature is in agreement over the harms of a centrally run commercialized mining pool. In these centrally run mining pools, the pool manager decides which transactions to include in a block and subsequently distributes the workload among participants of the pool. This type of structure requires trusting the manager of the pool thus limiting the decentralization in the blockchain (Chesterman, 2018).

Measurement Technique : Studies including (Beikverdi & Song, 2015; Gencer et al., 2018; Judmayer, Zamyatin, Stifter, Voyiatzis and Weippl, 2017; Sai et al., 2019a), have deployed an experimental setup to measure consensus centralization. Judmayer, Zamyatin et al. (2017) refer to this approach as a "block attribution scheme". In this experimental set-up, a participating node is connected to the blockchain that actively sniffs the network to extract mined blocks and coinbase addresses.[12] The coin base address is then used to query public blockchain explorers to determine if it belongs to a known mining pool. Based on the results, a list of the mining pools and the proportion of the blocks mined by each respective public mining pool is constructed. Using this approach, we can calculate the proportion of total computational power that each mining pool controls. Fig. 9 illustrates the block attribution scheme graphically.

This proportion can be represented as a percentage value as suggested by referred articles in Table 6 or by using the Gini values, based on the Lorenz Curve (Bruschi, Rana, Gentile, & Sciuto, 2019; Caccioli, Livan, & Aste, 2016).

[12] The coin base address refers to the address of the node that gets the reward for successfully mining a block.

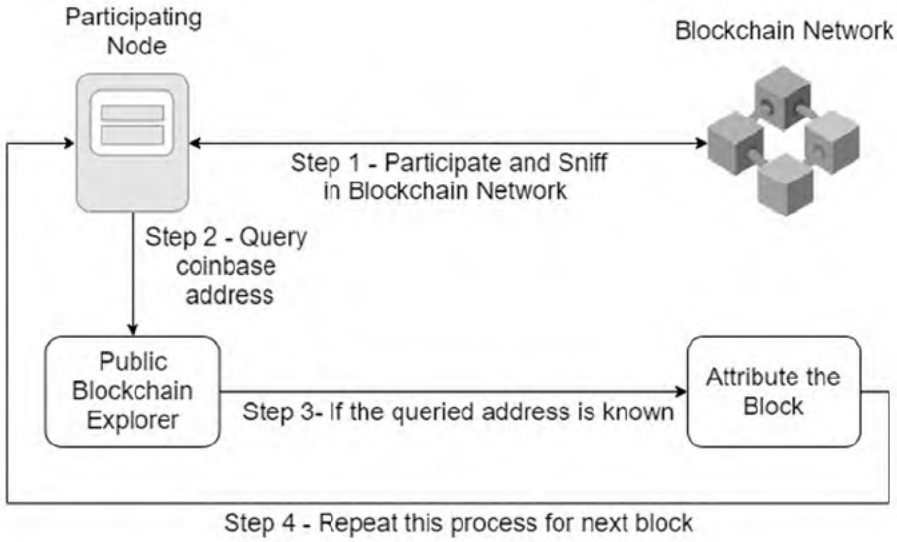

Fig. 9. Block attribution scheme.

Gini Value Measurement: These are economic measures of inequality (Dorfman, 1979; Gastwirth, 1971) for consensus power concentration (Bruschi et al., 2019; Caccioli et al., 2016).

The *Lorenz curve* is a graphical representation of the distribution of wealth. The curve illustrates the proportion of the income earned by any given percentage of the population. This curve has proven to be of significant importance in economic disparity measurement. To numerically describe this distribution, we can use the Gini Coefficient, which is based on the difference between the Lorenz curve and the line of equality.[13] We can calculate the Gini Coefficient as follows:

$$Gini = A/(A + B) \tag{4}$$

where A is the area between the line of equality and Lorenz curve, and B is the area under the line of equality. The value of Gini can range between 0 to 1, where 0 represents complete equality, and 1 represents complete inequality.

Implications of consensus power centralization: The impact of centralization in consensus power has been widely studied in security literature (Chen et al., 2017; Gervais et al., 2016; Karame, 2016; Sai et al., 2019a; Sapirshtein et al., 2016; Zhang et al., 2019). A concentration of 26% in proof of work-based blockchain can lead to successful selfish mining attacks. Whereas a consensus power concentration of over 51% can result in a 51% attack.

Smaller cryptocurrecies tend to be more prone to 51% attack as evident by successful attacks on Aurum Coin, Bitcoin Gold, Ethereum Classic, Flo Blockchain, Monacoin, Verge, Vertcoin and ZenCash (Sayeed & Marco-Gisbert, 2019). These 51% attacks have, on average, resulted in a loss of $2.5 million per cryptocurrency (Sayeed & Marco-Gisbert, 2019). The significance of these attacks is evident by the agreement of all our interviewees on the centralization implications of a 51% attack caused by consensus power concentration.

4.3.2. State-of-the-art for centralization on the consensus layer

Recent successful double-spending attacks due to the consensus power centralization have highlighted the need for a more encompassing understanding of the incentives behind the honest behavior for participation. Based on the results of Sai et al. (2019a), it seems that a better understanding of the economics behind the consensus system is needed to ensure secure operation. This is especially important for smaller cryptocurrencies as the barrier to conduct a 51% attack is lower when compared to major cryptocurrencies. We suggest a further in-depth investigation of the economic incentives behind conducting a double-spending attack against smaller cryptocurrencies be conducted.

4.4. Incentive layer

Bitcoin and similar decentralized cryptocurrencies are inherently dependent on the economics associated with rewards (Sai et al., 2019a). Sai et al. (2019a) reports that the exchange rate of Bitcoin is related to the overall consensus power of the network. If the exchange rate falls below a given threshold of profitability, the participants of the network may withdraw from active mining, which may result in a fall in overall hashing power of the network. A low value of hashing power of the network makes it easier for attackers to attain a higher consensus proportion; thus it may increase the threat of selfish mining and 51% attack. This interplay between the monetary aspect of public cryptocurrencies and security makes it essential to inspect centralization on the economy

[13] Line of equality refers to the equal distribution of hashing power among miners.

Table 7
Categories of centralization in incentive layer.

Ref	Wealth concentration	
	Identification	Measurement
R16	✓	×
R51	✓	Gini
R52	✓	Percentage value
R55	✓	×
R62	✓	Percentage value
R67	✓	Gini

driven incentive aspect of the network. A high concentration of wealth to a select few may be an aspect of centralization that can prove to be harmful to the network. Attacks such as the Whale Transaction Attack (Liao & Katz, 2017) have exploited wealth concentration. In a whale transaction attack, the attacker attempts to induce disagreement[14] between the participants by providing a high transaction fee in an already published block.

The issue of wealth concentration was raised by 60% of our interview participants unprompted. P_7, for example, noted how they focused on wealth concentration: *"In general, I follow the money. If the trail of funds leads to one natural person or group of natural persons (regardless of number of addresses), then the process is relatively centralized along the spectrum of centralized–decentralized blockchain"*.

Table 7 outlines the result of our review, identifying relevant articles and shortlisted techniques for measurement. In this subsection, we review the centralization based on Wealth Concentration in depth. This type of centralization may be of significance for a blockchain solution that employs a wealth-oriented consensus mechanism such as Proof-of-Stake (Kiayias, Russell, David, & Oliynykov, 2017).

4.4.1. Wealth concentration

High accumulation of native cryptocurrency may give a unique advantage to an adversary. The high wealth concentration can also be used to increase the overall cost of transactions (Liao & Katz, 2017), as demonstrated in the iFish attack on the Ethereum network (Cryptoslate, 2018). In the iFish attack, the attacker induced a large number of transactions with a high transaction fee in a short period. This influx of high transaction fees resulted in a considerable increase in the transaction fee. Another form of network abuse arising from high wealth concentration involves transaction fee manipulation by artificially increasing the overall fee required for a successful transaction.

Based on the results from our review, we point that this wealth concentration also has economic impacts on the network. As reported by Kondor, Pósfai, Csabai, and Vattay (2014), already wealthy nodes in the bitcoin's transaction graph tend to increase their wealth at a higher speed than smaller nodes. They call this phenomena the *"rich get richer"* scheme.

Measurement Technique : Wealth concentration measurement is at the center of disparity studies in economics (Gini, 1921). One of the most commonly used measures is the Gini Coefficient calculated from the Lorenz Curve. The wealth concentration is measured in the form of inequality based on the population and what proportion of population controls how much wealth. Translating this directly to the blockchain could mean calculating Gini over a cryptocurrency and all existing addresses on the blockchain. But we argue that this may not be the most efficient way as techniques such as Hierarchical Deterministic Wallets (Gutoski & Stebila, 2015) promote the generation of new addresses for every transaction. To overcome this limitation, Srinivasan (2017) proposes establishing a lower bound value on the cryptocurrency contained in the address for inclusion in the measurement, i.e., a wallet with 0 cryptocurrencies may be excluded from the study, as it most likely resembles an inactive address. Another reported measurement technique is to use a percentage measure. However, a simple percentage measure fails to capture the distribution. Machine learning-based transaction clustering approaches have also been employed to extract behavioral patterns (Hu et al., 2021); these heuristics may also be useful for the calculation of wealth distribution.

Implications of Wealth Concentration: Wealth concentration is linked with a number of potential attacks, such as the possibility of a 51% attack in the case of a wealthy attacker during a fall in exchange rate (Sai et al., 2019a). Whale attack, as discussed above, is another example of a wealth oriented security threat to the network. However, both of these potential attacks are without any real-world incident reports.

One example of wealth concentration in a real-world attack is the transaction fee price manipulation caused by the iFish attack (Cryptoslate, 2018). During the iFish attack, the attacker was able to artificially inflate the transaction fee of Ethereum by 35%. Another example of a wealth oriented attack is the bZx hack, where a smart contract designed for lending Ether was exploited by sending high-value transactions and manipulating the platform (Zmudzinski, 2020).

A public blockchain with high wealth concentration contradicts the foundational notion of a more even and open monetary system. This has a direct implication on the adoption of the technology.

[14] The disagreement is in the form of blockchain fork.

Table 8
Categories of centralization in operational layer.

Ref	Storage constraint		Specialized equipment concentration	
	Identification	Measurement	Identification	Measurement
R9	✓	×	×	×
R23	×	×	✓	×
R24	✓	Rate of growth	×	×
R38	✓	×	×	×
R39	✓	×	×	×
R51	×	×	✓	×
R52	×	×	✓	×
R53	×	×	✓	×
R55	×	×	✓	×
R62	×	×	✓	×
R63	✓	×	×	×
R67	×	×	✓	×
R80	✓	×	×	×

4.4.2. State-of-the-art for centralization on the incentive layer

High wealth concentration in public blockchain poses security threats, specifically in the form of manipulating economics associated with the blockchain system, such as transaction fees and exchange rates. However, measuring the current state of wealth concentration in the public blockchain is a nontrivial problem due to widespread adoption of technologies that aim to increase users' anonymity. Based on our review, we note that the use of clustering techniques and setting a lower bound on the amount of cryptocurrency stored at an address can help establish the state of wealth distribution. We suggest that further investigation into deanonymizing the blockchain can improve the accuracy of the techniques used to calculate the Gini value and that this is an important agenda for research going forward.

4.5. Operational layer

The uncertainty of reward imposes a constraint on participation for rational investors. This reasoning is primarily based on the cost of mining (Sai, Le Gear and Buckley, 2019). A miner can earn rewards in the form of mining incentives and accumulated transaction fees from the mined block but to profitably mine on a Proof-of-Work blockchain, the difference between rewards earned and the expenses of the mining operation should be positive. This is the 'operations' we are referring to in this 'operational' layer. The expenses of mining operations include capital costs such as the acquisition of adequate hardware and other recurrent costs such as the cost of electricity.

After conducting the systematic review, we report two types of centralization associated with operational aspects of the public blockchain. The first is the move from commercially available mining equipment to proprietary application-specific integrated circuit machines. This increased capital, operational cost has proven to be a significant barrier to entry for new miners in Bitcoin (Borge et al., 2017). We categorize this type of specialized hardware centralization as *Specialized Equipment Concentration*.

Another factor that contributes to the cost of mining is the storage requirements for operating on the network. As all full nodes in the network are required to store and process all the transactions, the data stored increases (Dai, Zhang, Wang, & Jin, 2018). This imposes a significant barrier as traditional computing devices may not be able to participate in the network given high storage requirements. This may limit the participation in consensus to only the participants who can afford greater computational resources imposing a constraint on participation. A significant storage requirement may deter users with conventional computing devices from participating in the consensus altogether, resulting in a more centralized network converged on participants with high computational capabilities. This high storage requirement has been discussed as a centralization causing factor (Guo, Gao, Mei, Zhao, & Yang, 2019; Reddit, 2019; StopAndDecrypt, 2018).

In this layer of centralization, interviewee I_{10} had an interesting perspective suggesting a restructuring of contract layer to widen our definition of the layer to include other operational concerns.

In this subsection, we report the centralization caused by the operational cost involved in participating in the consensus of the blockchain. We also manifest the result of our systematic literature review in Table 8.

4.5.1. Size of the blockchain

The traditional computing devices are often limited in-memory capabilities and can only hold a constrained amount of data. Attaining a higher storage capacity may prove to be costly if the growth rate of the storage requirement is significantly high (Guo et al., 2019). This growth in requirement may act as a deterring factor for non-organizational users as the requirement of the investment may be significant (Raman & Varshney, 2017), thus prompting centralization of mining effort.

The issue of storage requirement was articulated by 20% of our interview participants. I_{10} said: *"Nothing really stops blockchains from becoming so large that we will run out of capacity. Personally, I have just experienced the first challenge because my Linux partition ran out of capacity; however, if I bought additional hard-disks, I will still be able to run a full node, but it is getting more expensive to run full nodes".*

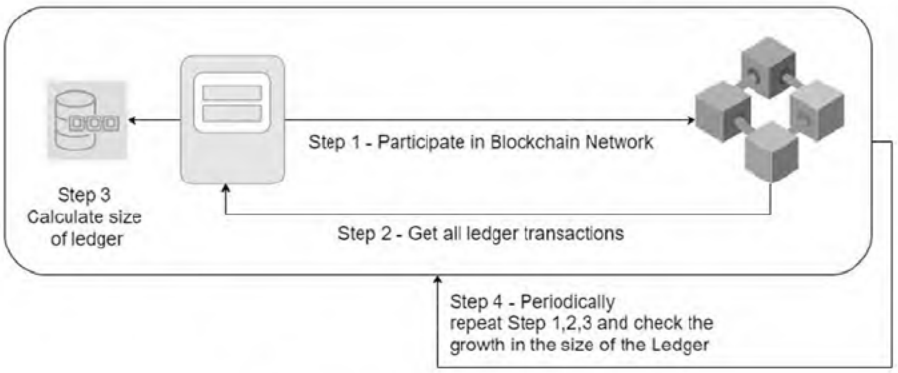

Fig. 10. Storage growth rate measurement technique.

Measurement Technique : To capture the storage-oriented centralization, Raman and Varshney (2017) suggests using the growth rate as a metric. This growth rate is determined based on historical data about the total size of the blockchain. The growth rate can be calculated periodically, ideally after every difficulty recalibration.[15] Fig. 10 illustrates the growth rate measurement technique graphically.

I_{10} stated their expectations for storage growth rate: *"considering that Moore's Law applies to hard drives, it will be interesting to measure the growth rate in comparison with Moore's law"*.

Implication of high storage requirements: Every blockchain instance may have a different storage requirements, based on its implementation. For example, Bitcoin does not pose significant storage issues as the overall requirement is still low. In contrast, Ethereum has an important storage requirement where the growth rate may limit participation. A growing storage requirement for Ethereum may result in fewer people being able to participate in the network as the participating nodes on Ethereum are expected to store code of smart contracts. A low number of participating nodes increases the likelihood of a successful DDoS attack as it reduces the attack surface.

4.5.2. Specialized equipment concentration

Proof-of-work based blockchains have seen a surge in the overall computational power of the network (Sai et al., 2019a). This surge has made it harder to get higher proportional control over consensus and, consequently, over the rewards associated with incentive. This higher computational requirement has induced an arms race in miners to acquire more efficient and specialized hardware (Ekblaw, Barabas, Harvey-Buschel, & Lippman, 2016). This type of specialized hardware is often not open source and gives the developers an advantage over others (Ekblaw et al., 2016).

60% of our interview participants acknowledged specialized equipment concentration as an issue for a public blockchain. I_7 suggested that this concentration may undermine the whole proposition of public blockchains: *".. but blockchain does not live in a vacuum, so really it was/is the externalities (ASICs and other special hardware for example) that threw the biggest spanner in the experiment"*.

Measurement Technique : Despite the significance of specialized equipment in Proof-of-work based mining operations, there is no existing metric to measure the centralization of hardware. Based on our literature review, we reason that this may be due to the non-public nature of this specialized hardware. As discussed earlier, most of these hardware implementations are not open source and often not available for public use.

Implication of Specialized Equipment Concentration: As reported by several studies listed in Table 8, the specialized equipment concentration may have given commercial entities an advantage over normal users. If this results in those commercial entries becoming focal, they may utilize the efficient computing equipment to attain higher consensus power and only release it to the public when it becomes less profitable to operate that computing equipment. This approach to hoarding efficient computing equipment is illustrated as the superhashing power dilemma by Bruschi et al. (2019). As a result of our review, we suggest that further investigation is warranted into the measurement of specialized equipment and its impact on centralization.

Apart from the above reported DDoS attack due to the low number of nodes, the specialized equipment requirement severely contains the participation. This higher barrier of entry and lack of profitability with old hardware makes it impractical to contribute to the network without significant investment. This lack of involvement has been shown to increase the likelihood of a successful selfish mining and double-spending attack (Sai et al., 2019a).

4.5.3. State-of-the-art for centralization on the operational layer

Our survey reports on two operational aspects of blockchain: the size of the blockchain and the specialized equipment concentration. The append-only nature of blockchain leads to an ever-increasing size of the ledger that needs to be stored in full

[15] In Proof-of-Work based blockchains, the difficulty of finding the solution of the computational puzzle is updated after a predefined number of blocks to maintain a static block creation time.

Table 9
Categories of centralization in application layer.

Ref	Wallet concentration		Exchange concentration		Reference client	
	Identification	Measurement	Identification	Measurement	Identification	Measurement
R4	×	×	×	×	✓	×
R6	×	×	×	×	✓	Satoshi index
R8	×	×	×	×	×	×
R11	✓	×	✓	×	×	×
R13	✓	×	✓	×	×	×
R26	×	×	×	×	✓	×
R27	×	×	✓	×	×	×
R34	×	×	✓	×	×	×
R36	✓	×	×	×	✓	×
R37	×	×	✓	Centrality	×	×
R40	✓	×	✓	×	×	×
R50	×	×	×	×	✓	×
R57	×	×	✓	×	×	×
R64	×	×	✓	×	×	×
R67	×	×	×	×	✓	×
R73	×	×	✓	Centrality	×	×
R76	✓	×	×	×	×	×
R78	✓	×	✓	×	×	×
R83	×	×	×	×	✓	×
R84	✓	×	✓	×	×	×
R86	✓	×	✓	×	×	×
R88	✓	×	×	×	×	×
R89	×	×	✓	×	×	×

nodes. This growth in the storage requirement is considered an adoption barrier. The current research only suggests metrics to measure it; however, there is a lack of techniques to counter the issue of growth in the storage requirement. We recommend the development of strategies to reduce the storage requirement in order to decrease this type of centralization. In terms of specialized equipment concentration, we note that this is yet to be quantifiably measured and we identify this as a high-potential avenue of future work.

4.6. Application layer

Users often rely on third-party applications to facilitate user interaction with the blockchain (Sai et al., 2019b). These third-party applications include reference implementations, wallets, and exchanges (Gervais et al., 2014). As a result of our review, we report on centralization on these three application layer entities. We also suggest that a monopoly in the user end applications for a blockchain instance is a contributor to the centralization of the blockchain. This issue of centralization on third-party applications was also pointed out by I_8: *"If you remember the catastrophe that centralized implementations such as Mt. Gox, Bitfinex have brought to the blockchain world, you can clearly see the desperate need for decentralization in user-facing applications"*.

Results from our literature review are outlined in Table 9.

This subsection is a manifestation of the identified centralization prone application layer entities.

4.6.1. Reference client development concentration

As described in Section 2.2, the data layer definition is implemented by a reference client, which acts as the gateway to the blockchain system. As any client that implements the protocol can become a part of the network, it is desirable from the decentralization point of view to have as many developers working on the reference implementation. Each client is expected to fulfill the protocol specification suggested by the core protocol. The development of the core protocol is decentralized by developing an open-source reference implementation. If a select few developers primarily drive the development of the core client, it contributes to centralization (Azouvi et al., 2018; Gervais et al., 2014). The decentralized protocol development factor captures this type of centralization. We note that this centralization is different from the improvement protocol centralization as the focus here on the development of a reference client and not improvements to the protocol.

Despite the reported adverse impact of this type of centralization on the blockchain, in Section 6, we present an argument in favor of some centralization in client development as the developer concentration may be a result of highly skilled developers making useful contributions.

Measurement Technique : (Gervais et al., 2014) suggests examining the number of unique developers contributing to the open-source project with the number of commits on the main core client codebase. This approach is then extended by Azouvi et al. (2018), where they propose using the Satoshi index, which represents the minimum percentage of all contributors required to reach 51% of data contribution.

Implication of reference client development concentration: If only a select few developers work on the reference implementation, they may gain unfair influence over the network. This concentration of power in the hand of select few feeds into the

governance issues discussed earlier. As discussed by Azouvi et al. (2018) and Gervais et al. (2014), this type of concentration is harmful to the decentralization of the network as a few developers may influence the implementation of change to the codebase. One of the major implications of influential actors in the public blockchain ecosystem is the defiance of open and equal monetary system assurance provided by the blockchain. As this open and equal system is one of the primary contributions of the public blockchain, the existence of influential entities severely limits systems capabilities to perform in an open and equal manner.

4.6.2. Exchange

Incentives for honest behavior are at the core of the decentralized, trustless transaction ledger. These incentives are often offered in the native cryptocurrency such as BTC and Ether. The real-world value of these cryptocurrencies has been debated (Sai et al., 2019a) with the recommendation that they be determined by the exchange rate to traditional fiat currencies. The exchange of cryptocurrency to traditional fiat currency is aided by application layer entities known as exchanges. These exchanges act as the means of consensus formation around the exchange value. This process is also known as *Price Discovery*. Due to the vital importance of the exchanges, the exchange applications must not be monopolized.

Measurement Technique : To measure the state of centralization in exchanges, Marvin (2017) propose measuring the centrality of exchanges by examining the flow of cryptocurrencies between addresses on the blockchain (Marvin, 2017). Addresses with high centrality in transactions may point to exchanges. This is observed by graphing the transaction flow and identifying nodes with a high degree of centrality. This was followed by the calculation of a Gini Coefficient that reports on the trend of centralization due to exchanges.

Other studies, such as Hileman and Rauchs (2017), have employed a percentage-based value measure, where they measure the proportion of all bitcoin transactions processed by exchanges.

Implication of centralized exchanges: A large number of successful attacks on Bitcoin and Ethereum have focused on exploiting vulnerabilities in exchanges (Chia et al., 2018). These centralized systems act as a single point of failure in case they also serve as a central repository of keys. A prominent example of this is the closure of Mt. Gox due to numerous security flaws leading to loss of Bitcoins owned by its users (Abrams, Goldstein, & Tabuchi, 2014).

Attacks on centralized exchanges not only impact the users of the exchange but the broader cryptocurrency community as it can instill doubts over the security of the ecosystem. These security attacks contribute to the barring trust and adoption by the wider community.

The use of these centralized exchanges may also be reflected through the uneven wealth distribution in the blockchain. Using the block attribution scheme discussed in Section 4.3.1, we report that major centralized exchanges such as Binance can store over $ 1 Billion in a single wallet location.[16]

4.6.3. Wallet concentration

Wallet applications are another form of centralized service on the application layer, as these applications are often developed and maintained by centralized organizations (Sai et al., 2019b).

Measurement Technique: Based on the review of the relevant literature, we report that there are no suggestions regarding the measurement of wallet concentration. We reason that this may be due to the nature of how wallets operate in a closed commercial environment. However, as most of these wallets use an exchange service to transmit funds such as Coinbase (Hileman & Rauchs, 2017), it may be reasoned that exchange centralization may provide a rough proxy for wallet based centralization as well.

Implication of centralized wallet: Applications such as wallets have been identified as a single point of failure and are considered a security threat (Sai et al., 2019b). A high concentration of wealth in centrally managed wallets may give the host an advantage feeding into the issue of wealth concentration. This concentration may also result in a dependence on centralized organization, consequently reducing the decentralization.

Similar to exchanges, a centralized wallet poses a potential barrier of entry in the ecosystem. Due to the technical ability required to host their wallets, most end users tend to prefer hosted wallets, which provides attackers with a small attack surface. This can aid attackers in conducting more targeted yet profitable attacks on the centralized wallet hosting service.

4.6.4. State-of-the-art for centralization on the application layer

Based on the literature review, we report that the dependence on centralized third party services for user-end applications such as exchanges, wallets, and reference client leads is a centralization causing factor. To measure the reference client centralization, Azouvi et al. (2018) suggest using the Satoshi index. However, the empirical data present in the surveyed reports is primarily focused on Bitcoin and Ethereum. We suggest that further investigation into smaller cryptocurrencies may assist in better understanding the current state of centralization across cryptocurrencies. Another form of user-end application is wallet applications. Based on our survey, we report that there are no measurement techniques to quantify this type of centralization. The presence of centralized wallets can also lead to more wealth concentration and this is specifically true for blockchain addresses used by centralized exchanges, as reported earlier. As such we propose this as an important area for future research.

[16] The identified Binance wallet address on Bitcoin leader is 1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s.

Table 10
State of centralization in Bitcoin and Ethereum.

Centralization factor	Bitcoin	Ethereum
Wallet concentration	No measurement	No measurement
Exchange concentration	7 exchanges served more than 97.24% of all trades	7 exchanges served more than 28.27% of all trades
Reference client concentration	Single developer authored 25.11% of all files	Single developer authored 40% of all files
Storage growth rate	0.5 GB per week	0.68 GB per week
Specialized equipment concentration	No measurement	No measurement
Wealth concentration	Gini = 0.56	Gini = 0.64
Consensus power distribution	Top 4 mining pools with 50.36% consensus power	Top 4 mining pools with 63.04% consensus power
Node discovery protocol control	No measurement	No measurement
Geographic distribution	Network latency 26.7% less than Ethereum (Gencer et al., 2018)	26.7% higher than Bitcoin (Gencer et al., 2018)
Bandwidth concentration	1.9 to 2.7 times greater than Ethereum (Gencer et al., 2018)	1.9 to 2.7 times less than Bitcoin (Gencer et al., 2018)
Routing centralization	30% network with 10 ASes (Apostolaki et al., 2017)	28% network with 1 AS (Gencer et al., 2018)
Owner control	$C_{OwnerControl} = 0.033$	$C_{OwnerControl} = 0.11$
Improvement protocol	Mean = 11.41, Median = 2.0, IQR = 6.5, IQMean = 2.95	Mean = 9.16, Median = 2.0, IQR = 5.0, IQMean = 2.56

5. State of centralization in Bitcoin and Ethereum

The following subsection provides an overview of empirical evidence specific to the two most prominently used blockchain-based cryptocurrencies: Bitcoin and Ethereum. We present the view of the literature on the centralization of these two cryptocurrencies. Where feasible, we also report the present state of centralization by conducting measurements following the taxonomy guidelines. To structure this investigation, we use the initial taxonomy. The results from this investigation are manifested in Table 10.

5.1. Governance layer

Owner Control: Satoshi Nakamoto is largely credited for the authorship and development of bitcoin (Nakamoto, 2008). This as-yet unknown individual or organization is said to have performed active mining in the early days of Bitcoin, accumulating a considerable amount of BTC (Bohr & Bashir, 2014). As Bitcoin implements provisions for anonymity, the amount of BTC held by Satoshi is not publicly known. Based on the data obtained from the Bitcoin blocks mined in 2009 (Blockchain luxembourg s.a, 2019; Sergio, 2013) performed clustering of similar wallets to identify large entities gathering BTCs. They also report that the largest gain of around 700,000 BTC belonged to a single entity performing mining in 2009. This gives us a value of $C_{OwnerControl} = 0.033$ for Bitcoin.

This value is significantly less than that of Ethereum, where the value of $C_{OwnerControl}$ is 0.11. Bai, Zhang, Xu, Chen, and Wang (2020) argues that Ethereum is very unfair since "the rich are already very rich". This high wealth disparity may allow select participants to conduct attacks based on economic manipulation of the Ethereum ecosystem, such as the Whale transaction attack and transaction fee manipulation. A reason for the high value has been presented in Section 4.

Improvement Protocol: According to Bitcoin (2019), all changes must be approved by all the developers of the core client. However, Gervais et al. (2014) reported on one violation of this rule. In this violation, a subset of developers unilaterally decided to lower the minimum transaction fee to 0.0001 BTC. This illustration strengthens (Gervais et al., 2014) questioning of the transparency, in the process of handling improvement proposals.

Azouvi et al. (2018) measures the centralization by calculating the centrality metrics for both Bitcoin and Ethereum reported in Table 10. They report that the collected commits and comments data set contained many outliers pointing out that the top 25% of developers contribute significantly more than others. They also point out that in their data set for Ethereum, a vast majority of EIP contributions are from a single user, Vitalik Buterin, the founder of Ethereum. They report a similar trend in Bitcoin, where only a handful of people are contributing to the improvement protocol allowing these select groups of people to dictate the changes that are implemented in the protocol. In the past, the block size debate surrounding the scalability has often been cited as a prime example of this type of governance control. Based on our current analysis, we report that the state of centralization in improvement protocol remains largely unchanged since the calculation of Azouvi et al. (2018) with Vitalik Buterin still dominating the EIP contributions in Ethereum and Gavin Andresen remaining responsible for the vast majority of accepted BIP in Bitcoin[17]

5.2. Network layer

Node Discovery Protocol Control: As reported in Section 4, our literature review suggest that no study has focused on the measurement of centralization of seed DNS nodes. We postulate that further investigation is required to measure this type of centralization adequately.

[17] The results from our improvement protocol analysis on 02/01/2020 are available at www.github.com/ashishrsai/centralization.

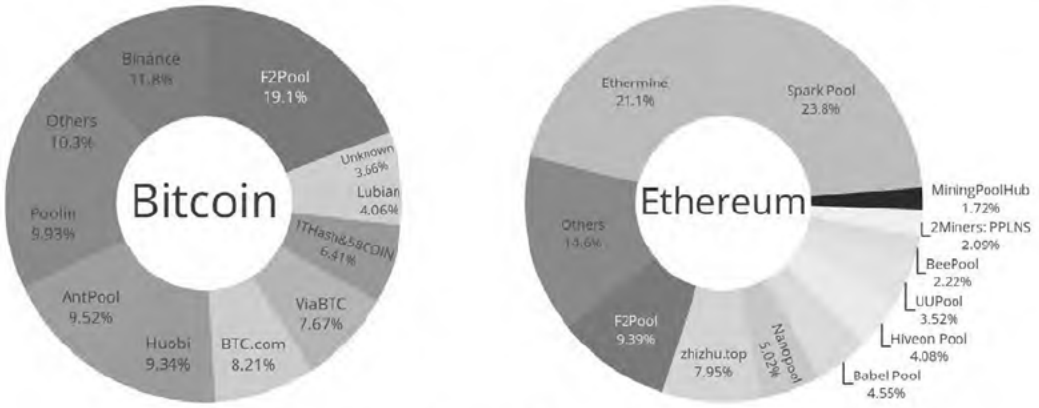

Fig. 11. Consensus centralization in Bitcoin and Ethereum.

Geographic Distribution: Gencer et al. (2018) conducted an extensive review of both Bitcoin and Etherum to measure centralization in the network layer. They reported that the Bitcoin network is more geographically centralized than Ethereum. The average peer-to-peer network latency of Ethereum is 26.7% higher than Bitcoin, suggesting that Ethereum nodes are located at a greater geographic distance. They reason that this is due to the data center focused approach to mining for Bitcoin, whereas Ethereum can be mined by using consumer hardware. This association between geographical distribution and operational centralization neatly illustrates the interdependency between different aspects of centralization, even those based in different layers.

Bandwidth Concentration: Gencer et al. (2018) states that nodes in Bitcoin tend to have about 1.9 to 2.7 times more network bandwidth than Ethereum nodes. They also report that based on the bandwidth, it can be assumed that Bitcoin nodes are located in data center clusters, whereas Ethereum exhibits a more spread out distribution of bandwidth.

Routing Centralization: Feld et al. (2014) reports that 30% of the bitcoin network was only made up of 10 ASes, which presents a level of security threat. This work was expanded by Apostolaki et al. (2017), where they report that 13 ASes covered about 30% of the network but only consisted of 36 IP prefixes. These 36 IP prefixes cover about 50% of mining power. However, the only investigation that has reported on AS-Level centralization in Ethereum, Gencer et al. (2018) reports that 28% of Ethereum nodes belonged to a single AS.

Replicating the experimentation of Gencer et al. (2018) to get the current measurement for network layer centralization requires access to an extensive geographically spread beacon network. Due to resource constraints, we leave such experimentation for other, larger research groups.

5.3. Consensus layer

Centralization of consensus power of bitcoin has been studied thoroughly in the literature (Beikverdi & Song, 2015; Gervais et al., 2014, 2016; Karame & Androulaki, 2016; Sai et al., 2019a). Beikverdi and Song (2015) uses a percentage based centralization value to derive a new metric called Centralization Factor. They report that at the beginning of 2011, 30% of all hashing power was controlled by eight mining pools. This concentration sees a significant increase in 2014 when, according to Gervais et al. (2014), the top mining pool alone controls close to 40% of all hashing power of the network.

Gencer et al. (2018) expands these analyses by also examining Ethereum's network. During the observation period, Gencer et al. (2018) reports that Bitcoin had a less centralized corsensus mechanism than Ethereum. On average, the top four mining pools in Bitcoin controlled 53% of the hashing power, whereas in Ethereum the top three mining pools controlled 61% hashing power.

We followed the block attribution scheme for both Bitcoin and Ethereum as discussed in Section 4.3.1 for a week's period.[18] Our results are in line with the observations of Gencer et al. (2018), with the top four mining pools in Ethereun constituting 63.04% of the hashing power, whereas, in Bitcoin, the top four mining pools controlled 50.36% of the hashing power. The results from our analysis are illustrated in the Fig. 11.

5.4. Incentive layer

According to Malik (2016), as of 2016, 11,000 unique Bitcoin addresses, out of a total of 12 million, contained 75.2% of all Bitcoin in circulation. This disparity shows a significant concentration of wealth to a select few. Chohan (2019) also supports the claim of significant inequality in the Bitcoin network. The author claims that the level of inequality reflects that of traditional economies and voids the proposed purpose of Bitcoin: decentralization. Gupta and Gupta (2017) conducted an in-depth investigation of the inequality of Bitcoin. They report that Bitcoin had a Gini value of 0.995 in the year 2013. This result is then refined by Srinivasan

[18] From DEC-25-2020 12:00:00 PM +UTC until JAN-01-2021 12:00:00 PM +UTC.

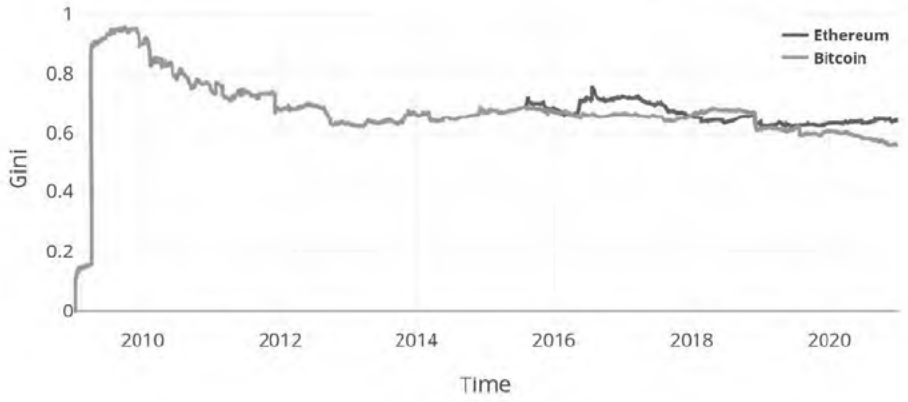

Fig. 12. Trend of Gini value in Bitcoin and Ethereum.

(2017), where they set a lower bound on the Bitcoin account to account for Hierarchical Deterministic wallets as described in Section 3. They report that in 2018, Bitcoin had a Gini value of 0.65, where they set the minimum threshold to 185 BTC per account. This Gini value suggests that wealth in bitcoin is highly centralized when compared to real economies where, according to the World Bank (World Bank, 0000), the highest reported Gini value is 0.63.

According to Srinivasan (2017), Ethereum demonstrates a similar trend of significant centralization with a Gini value of 0.76 with a minimum threshold of 2477 ETH per account. This suggested trend is in line with the report by Huobi Blockchain (0000), where they claim Ethereum to be more centralized in terms of wealth distribution.

In line with the investigation of Gupta and Gupta (2017) and Srinivasan (2017), we parsed daily transactions for both Bitcoin and Ethereum starting from the genesis block. Our results are in accordance with previous reports of significant wealth inequality in Bitcoin and Ethereum. However, we also report that there is a downwards trend in Bitcoin where the Gini has been steadily decreasing with the current Gini value of 0.56. The current Gini value for Ethereum is 0.64, which is lower than the Figure suggested by Srinivasan (2017), implying a downwards trend. We have visualized these results in Fig. 12.

5.5. Operational layer

In Pustišek, Umek, and Kos (2019), the authors report that the Bitcoin full node requires 204 GB storage space. This storage requirement is slightly lower than the 385 GB required by Ethereum for a full node (Afanasev, Krylova, Shorokhov, Fedosov, & Sidorenko, 2018). Pustišek et al. (2019) also reports that the storage growth rate is about 0.1–0.5 GB per day. Our review was unable to identify any longitudinal studies that observe the growth in storage requirements over a long time. To account for this, we collected the storage growth rate data for both Ethereum and Bitcoin for a period of a month by hosting full nodes. We report that Bitcoin's storage growth rate is on average 0.50 GB per week, whereas Ethereun tends to grow at a faster rate with the weekly growth rate of 0.68 GB.[19]

As reported in Section 4, numerous studies identify specialized equipment concentration as a cause of centralization. Despite the significant attention to this issue, our review suggests that there are no proposed measurement techniques.

5.6. Application layer

Reference Client Concentration: According to Azouvi et al. (2018), a single author wrote about 30% of all files in the bitcoin reference implementation.[20] This is significantly higher in Ethereum, where an individual author wrote 55% of all files. They also analyze the comments on the GitHub pages of Bitcoin and Ethereum reference clients. They report that only eight people contributed to half of all comments representing 0.3% of all commenters. This concentration in comments is also observable in Ethereum, where 0.6% commenters contributed to 50% of comments.

We analyzed the core clients for Bitcoin and Ethereum (Geth[21]) to observe the current state of centralization in the development. We report that Ethereum has a higher contribution of single developers than Bitcoin. In Ethereum, a single author has contributed to over 40% of all commits, whereas in Bitcoin, a single author wrote 25.11% of all commits. These observations are in line with Azouvi et al.'s (2018) results. We have reported the top 5 contributors to Ethereum and Github core clients in Fig. 13.

Exchange Concentration: Intermediary services such as Exchanges that also act as central key stores for Bitcoin have been suggested as a centralization causing factor by Böhme et al. (2015). A prominent example of the harm caused by exchange

[19] Both the full nodes were hosted from DEC-01-2020 to JAN-01-2021, the daily growth reports are available at www.github.com/ashishrsai/centralization.

[20] Azouvi et al. (2018) propose using the Satoshi Index to measure centralization in client development. However, the specific values of Satoshi Index for Bitcoin and Ethereum are not available.

[21] https://github.com/ethereum/go-ethereum.

Fig. 13. Github contributions of top 5 authors for Bitcoin and Ethereum.

concentration is the collapse of Mt. Gox in 2014 (Abrams et al., 2014). In 2014, Mt. Gox was the leading exchange for Bitcoin, and its closure resulted in a total loss of $450 Million. Böhme et al. (2015) reports that the concentration of excharges was still high in 2015 when the seven largest exchanges served more than 95% of all bitcoin trades.

An empirical analysis conducted by Böhme et al. (2015) reported that out of 40 Bitcoin exchanges examined, 18 had closed, wiping out customers' account balance as they stored the private keys of customers. They argue that these exchanges operate as the de facto centralized authorities in the Bitcoin network.

As for Ethereum, we report that there are no studies that explicitly report on the behavior of exchanges for Ethereum. However, as suggested by Kim and Lee (2018), most of the Bitcoin exchanges also exchange multiple other cryptocurrencies, including Ether.

To account for the lack of empirical data for Ethereum, we measure the centrality of exchanges by observing the flow of cryptocurrencies between addresses on the blockchain as discussed in Section 4.6.2. We measure centrality for both Bitcoin and Ethereum for a period of a week.[22] We report that the top 7 exchanges on Bitcoin processed over 97.24% of all trades. This ratio was significantly lower for Ethereum, with the top 7 exchanges contributing to 28.27% of all transactions.

As discussed earlier, based on our systematic review, we conclude that there is no suggestion regarding a measurement technique to capture wallet based centralization.

So, in terms of Bitcoin, the main centralization threats are at the Network, Consensus, and Application layers. Specifically, the centralization aspects of the Network layer: geographic distribution, bandwidth, and routing are vulnerabilities for bitcoin in that they allow the specific threats of geopolitical manipulation of the network, high resource requirement for participation, and possibility of network attacks. These threats for bitcoin are augmented by the high concentration of consensus power to centralized mining pools and application layer operations such as exchanges and wallets.

Ethereum also shares the issues of centralization on the application layer as they lead to reliance on centralized entities such as exchanges and wallets for participation in the network. Other significant centralization threats for Ethereum include the Governance, Consensus, and Incentive layers. Especially the centralization aspects of the Governance and Incentive layers may induce vulnerabilities for Ethereum in that they allow unilateral decision making on the governance layer and high wealth concentration on the incentive layer.

6. Discussion

In this first in-depth investigation of the centralization of public blockchain solutions, we conducted a systematic review of existing literature to produce an initial taxonomy of centralization. We then refined this initial taxonomy through expert interviews. We provide an overview of centralization in different aspects of the blockchain. We examine different means of measuring centralization, also pointing out the absence of measurement techniques in these research studies. This initial taxonomy provides a framework for a more systematic discussion around the centralization of major blockchain systems. The following section discusses the findings of our survey.

6.1. Non binary nature of centralization

We observe that decentralization in the public Blockchain literature is a loosely-defined term that can take many shapes and forms. We also observe that most of the non-decentralization-specific articles reviewed treat decentralization as a binary construct. That is: a blockchain instance is either centralized or decentralized. However, based on our taxonomy, we define centralization of

[22] From DEC-25-2020 12:00:00 PM +UTC until JAN-01-2021 12:00:00 PM +UTC.

public Blockchains as *the process by which one or more architectural dimensions (aspects) of the Blockchain are restrictive to the majority of participants by direct or indirect economic, social, or technical constraints* and so argue that centralization is not suited to binary classification.

This latter observation aligns with expert interviews, where 60% of participants preferred a spectrum of values for centralization rather than the conventional binary notion. However, the interviewees also acknowledged that the complexity of a more granular definition might dilute the meaning to non-experts in the blockchain domain. For example, I_5 said: *"I am an engineer, so I prefer precision and a multidimensional model, but I know when you are presenting to business people, a single score might be what they are looking for"*.

This survey presents a novel, initial taxonomy to address this dilution concern and allows for structured discussion on centralization. The following text discusses the key findings of the taxonomy.

Consensus power concentration was the most recognized form of blockchain centralization by both the literature and experts interviewed. We reason that this wide recognition is due to the dependence of significant security threats such as the Double Spending (Karame et al., 2012) and Selfish mining (Sapirshtein et al., 2016) attacks on the consensus power concentration. The practical implication of this centralization is the heavy the impact of mining pools when operating a profitable mining operation. The dominance of mining pools is observable in both Ethereum and Bitcoin. In Bitcoin the top 4 mining pools control over 53% of the hashing power, whereas in Ethereum the top 3 mining pools control over 61% of the hashing power (See Table 10).

A high concentration of consensus power can induce an arm's race to attain the most efficient hardware (Sai et al., 2019a). Our survey reports that this race often results in specialized proprietary hardware. The practical implication of this type of hardware concentration is an indirect limitation to participation as only efficient, and often proprietary hardware can result in a profitable operation. To remedy this situation, studies such as Cho et al. Hyungmin (2018) (Cho, 2018), have proposed using a consensus algorithm that is memory heavy, for which specialized hardware design is inefficient.

Surprisingly on a similar operational constraint, the Storage growth rate was less widely recognized to contribute to centralization. However, I_{10} raised an interesting issue on the ever-increasing append-only nature of Blockchain that may result in consistent growth in storage requirements. As reported in Table 10, the current growth rate for Bitcoin is around 0.1 to 0.5 GB per day. The practical implication of this increased storage requirement is the inability of conventional computing devices to serve as nodes in the blockchain (Guo et al., 2019). Guo et al. (2019) propose a storage optimization scheme based on the redundant residual number system that can reduce the storage requirement. We suggest that a further investigation into storage optimization in public Blockchain is warranted.

Another unexpected finding of our survey was that 50% of the interviewees accepted node discovery protocol control as a threat to decentralization, despite only one research article reporting on the issue. We reason that this may be due to the practical implications of setting up a new node such as the potential delay in network connection for new nodes due to high traffic through DNS nodes. This type of delay is often not accounted for in network simulation tools such as NS3, employed by studies such as (Gervais et al., 2016; Sai et al., 2019a).

Contrary to the previous example, routing and bandwidth centralization in the network was not widely recognized by the interviewees. One potential explanation could be the experimental nature of the measurement associated with the routing and bandwidth centralization. Despite these being recognized as issues, both the bandwidth and routing do not cause operational issues to most participants at present.

Another network-oriented centralization concern widely recognized by both the literature and interviewees is the geographic distribution of the nodes. Our findings suggest that the Ethereum network is more geographically spread out than Bitcoin. We reason that this is due to the possibility of using conventional hardware such as GPUs to participate in Ethereum. Despite the recognition, our literature review did not identify potential strategies to address this centralization. We suggest that strategies to limit geographic concentration should be investigated.

The lack of mitigation techniques is also persistent in the application layer aspects. The wallet and exchange centralization have been reported on by the literature and also recognized as centralization issues by expert interviews. As reasoned earlier, the centralized store of cryptocurrencies may give an advantage to the exchange or wallet operator. This advantage is often in the form of wealth concentration and can be observed in the centralization of Bitcoin exchange platforms, where only seven exchanges were reported to serve more than 95% of all trades.

Interviewees and literature also agree on the implication of wealth concentration on the decentralization. Surprisingly, despite the apparent issue of a *"Rich getting Richer"* effect in Proof-of-Stake cryptocurrencies (Fanti et al., 2019), most of the reported literature focused on the wealth concentration in Proof-of-Work. We suggest that the issue of wealth concentration be investigated in the context of Proof-of-Stake cryptocurrencies.

Another factor that may result in a *"Rich getting Richer"* effect is the distribution of wealth at the very start of the Blockchain captured by owner control in our taxonomy. The issue of owner control is also associated with how the Blockchain is governed. Governance centralization in Blockchain is widely recognized by both the literature and interviewees. Interestingly, Wang et al. (2017) argue for some centralization in the governance to facilitate quick response to security threats. We expand on this line of reasoning in the following subsection.

6.2. Aspect based measurement of implications of centralization

As pointed out earlier, not all aspects of our taxonomy are an equal contributor to the overall centralization of the blockchain. This was also substantiated by six interviewees agreeing that a combined value of centralization for the overall blockchain would

not be meaningful. For example, storage constraint oriented centralization may be an issue in Ethereum due to the requirement to store smart contracts. In contrast, this may not be a significant issue for Bitcoin as only transactions drive the storage requirements. We expand on this category-based significance reasoning that not all centralization is necessarily equally bad for the network:

The governance layer based centralization argument presented by Gervais et al. (2014) assumes that concentrating decision making power to a select few is bad for blockchain. However, we question this argument, as true decentralization is an impossibility in real world scenarios (Kwon et al., 2019; Szabo, 1970). The concentration in decision making had also proven to be useful in instances of network attacks when a prompt response was mandated (Wang et al., 2017). Delegation of controlling power during the cases of security bugs or attacks may have proven to be detrimental to the network. Despite the lack of decentralization in governance, it may be to the overall benefit of the network. We present this as a potential future research avenue to explore the most suitable governance structure for decentralized systems.

We also argue that the results obtained by Azouvi et al. (2018) regarding the centralization in source code development for core client implementation may not necessarily be bad. It may just be the case that only a handful of developers have an in-depth understanding of the source code to make useful contributions to the system. This reasoning of limited expertise feeds into the argument against the decentralization of the improvement protocol. As pointed out by Azouvi et al. (2018), the vast majority of the Ethereum Improvement Protocol recommendations originated from a single developer, Vitalik Buterin. We reason that this may be due to the quality of suggestions proposed by Vitalik.

These arguments in favor of some centralization are an example of the complex nature of decentralization in distributed systems. We propose that the significance of each aspect of centralization be determined based on the empirical evidence specific to each blockchain instance.

7. Conclusion

In this paper, we conduct a systematic literature review to provide a summary of the research done on the centralization aspect of blockchain. We structure our findings in a novel initial taxonomy of centralization. This taxonomy is then refined and validated through expert interviews.

Given the significant growth in the application of blockchain technology in information systems (Chen et al., 2020; Khalid et al., 2021; Li et al., 2020; Putz et al., 2021), it becomes imperative to understand centralization's socio-technical nature in the blockchain. This refined understanding of centralization helps us better understand the security and performance implications associated with adopting a blockchain-based decentralization approach in existing information systems. Another important aspect associated with the adoption of blockchain-based decentralization is the management of the decentralized system. This taxonomy report on the issues associated with the governance of a decentralized system and its potential implications.

7.1. Contribution

Decentralized blockchain solutions provide a means of monetary asset transfer without a trusted third party; this is attained through the delegation of the validation power to all participants of the system rather than the administrator. This delegation of control is often referred to as the original contribution of blockchain systems (Bonneau et al., 2015). Based on previous studies, Cong et al. (2019), Gencer et al. (2018), Gervais et al. (2014) and Sai et al. (2019a), we reason that the preconceived notion that blockchains are inherently decentralized may not hold in the present situation and that raises the potential of severe issues for blockchain instances. Due to the lack of an objective measure of centralization, it becomes impractical to discuss improvement in terms of centralization.

Centralization is a challenging variable to research, in part because of the multiple definitions and measures of centralization applicable in blockchain and, to date, the implicit nature of several of those aspects and the lack of an encompassing framework. We report on these myriads of definitions, conceptualizations, and dimensions used to describe this concept by segmenting them based on a generic architecture proposed by Zhang et al. (2019). Our study contributes to the existing body of knowledge by systematically surveying and synthesizing the blockchain literature, reporting on the adverse impact of centralization such as security threats, as well as identifying research gaps such as the lack of Ethereum specific research on centralization.

With this systematic review, we provide the reader with an overview of various forms of centralization in Blockchain resulting in an initial taxonomy. This taxonomy also contains numerous existing measurement techniques used to measure centralization. It may help researchers evaluate the centralization of a blockchain instance, but will also allow researchers add more aspects of centralization as they become known, providing them with a vocabulary of centralization that will allow them address the issues that arise.

We have also reported on the platform-specific findings for the two most prominently used blockchain-based cryptocurrencies: Bitcoin and Ethereum. We report that both Bitcoin and Ethereum have similar centralization issues with regards to reference client implementation, decentralized protocol development, and exchanges. However, in terms of wealth concentration, Ethereum is more centralized than Bitcoin, primarily due to high owner control. This trend continues with consensus power concentration, where Ethereum is reported to be more centralized than Bitcoin. Ethereum nodes, however, are geographically more spread out than Bitcoin, resulting in a low geographic concentration when compared to Bitcoin.

We also discuss that centralization on all aspects is not necessarily adverse for the blockchain by expanding the argument in favor of some centralization by Wang et al. (2017). We suggest that the unpropitious impact of centralization be measured on each aspect based on empirical evidence. This aspect-specific investigation may assist the move from the binary notion of decentralization to a multidimensional scale encompassing adequate measurement and control where necessary.

7.2. Threats to validity

As decentralization is fundamental to a public blockchain, the term is frequently used in the title and abstract of articles relating to public blockchains. To not omit any relevant articles, we kept the search queries generic by including any article that includes the term *"Blockchain"* and *"Decentralization"* along with suggested alternate words in Section 3. We acknowledge that despite the broad terms used, we may have missed relevant articles not present, or with different phrasing, on these leading search repositories. These missed articles may include *"gray literature"*, which is of significant importance in the blockchain research domain (Casino, Dasaklis, & Patsakis, 2019). To overcome this limitation, we included Google Scholar in our search process. However, as reported earlier, the Google Scholar search was limited to the top 1000 entries, even though the relevant articles dropped off significantly after the top four hundred returned articles.

The literature review may also be limited due to the strict inclusion and exclusion criteria for the title and abstract filtering. We reason that these strict criteria are warranted due to a large number of articles retrieved by the search queries (3574 non-duplicate entries). To overcome this limitation, we employed a two-step filtration by reviewing both the title and abstract. We also performed cross-validation of the filtration process by the independent review of the articles by two authors. This cross-validation process resulted in Cohen's Kappa value of 0.84, which is considered an almost perfect agreement. We repeated a similar cross-validation process for the full-text filtration.

The review process aimed to extract factors from all shortlisted articles despite their core focus. As the study of centralization in public blockchain is still in the early stage, we included articles where the core focus was not centralization. This inclusion may have limited the quality of shortlisted articles, as observed by the exclusion of 148 articles after full-text filtration. To overcome this limitation, we performed a quality review of all 212 shortlisted articles and shortlisted a final set of 89 articles.

To further evaluate the literature-review findings, we interviewed ten experts. The recruitment process was based on the prominence of authors in the bibliographic map generated by Ramona et al. (2019). As with any other qualitative research method, interviews have several limitations, as pointed out by Opdenakker (2006). In addressing them, we adhere to the validity dimensions put forth by Maxwell (1992) for qualitative studies. The first validity threat is the descriptive validity of the data obtained through interviews. To limit this, we transcribed the audio-captured interview in verbatim form. However, in the interviews that relied on contemporaneous notes, it is possible that the interviewer may have missed some observations. The second threat to validity is the interpretive validity of the interviews. To address this, we used open-ended questions and restricted the questions strictly to the research questions presented in Section 3.2. We also coded the interviews based on the terms used by the interviewees rather than an interpretation. The transcripts and notes were individually checked by researchers from the author list. The interviewees were also given back the interview transcripts and notes for validation.

7.3. Future work

Having provided a comprehensive overview of centralization in public blockchain, a case study focused on individual cryptocurrencies, and blockchain implementations would complement our study. This case study could include an in-depth centralization review of, for example, Bitcoin, Ethereum, and Libra (Pilkington, 2019).

The taxonomy developed by our study can also be expanded to provide an objective measure of centralization for blockchain instances, as a whole, to facilitate comparison. This objective measure may prove to be useful for the evaluation of centralization from a novice user, or governance perceptive. Four of our ten interviewees stated that they would prefer a single score to measure centralization objectively, and thought it would assist end-users and nonspecialist researchers.

We also hope to develop different flavors of this initial taxonomy that are specific to implementation details. For instance, the presented taxonomy is generic and does not consider consensus specific issues such as Stake bleeding (Gaži, Kiayias, & Russell, 2018). It also omits the consideration of source code dependencies in Smart Contracts. In future, we intend to statistically examine the source code of smart contracts to observe if a handful of libraries dominate the smart contracts in Ethereum.

The work presented here only examines the already identified factors that may lead to centralization and does not analyze the existence of other novel forms of centralization. As a part of future work, we will consider a thorough review of one of the reference blockchain implementations to identify factors that may also contribute to centralization directly or indirectly.

We also aim to review existing literature to identify potential solutions to the centralization avenues suggested by our review. These solutions may facilitate integrating centralization considerations during the development of public blockchains.

CRediT authorship contribution statement

Ashish Rajendra Sai: Conceptualization, Methodology, Data curation, Writing - original draft. **Jim Buckley:** Writing - review & editing, Supervision, Project administration, Validation. **Brian Fitzgerald:** Writing - review & editing, Supervision. **Andrew Le Gear:** Visualization, Investigation, Supervision, Validation, Writing - review & editing.

Acknowledgments

This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ipm.2021.102584.

References

Abrams, Rachel, Goldstein, Matthew, & Tabuchi, Hiroko (2014). Erosion of faith was death knell for mt. Gox. *New York Times*.

Afanasev, Maxim Ya, Krylova, Anastasiya A., Shorokhov, Sergey A., Fedosov, Yuri V., & Sidorenko, Anastasiia S. (2018). A design of cyber-physical production system prototype based on an ethereum private network. In *2018 22nd conference of open innovations association (FRUCT)* (pp. 3–11). IEEE.

Akram, Shaik V., Malik, Praveen K., Singh, Rajesh, Anita, Gehlot, & Tanwar, Sudeep (2020). Adoption of blockchain technology in various realms: Opportunities and challenges. *Security and Privacy*, *3*(5), Article e109.

Alzahrani, Naif, & Bulusu, Nirupama (2018). Towards true decentralization: A blockchain consensus protocol based on game theory and randomness. In *International conference on decision and game theory for security* (pp. 465–485). Springer.

Anceaume, Emmanuelle, Lajoie-Mazenc, Thibaut, Ludinard, Romaric, & Sericola, Bruno (2016). Safety analysis of bitcoin improvement proposals. In *2016 IEEE 15th international symposium on network computing and applications (NCA)* (pp. 318–325). IEEE.

Androulaki, Elli, Barger, Artem, Bortnikov, Vita, Cachin, Christian, Christidis, Konstantinos, De Caro, Angelo, et al. (2018). Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the thirteenth eurosys conference* (p. 30). ACM.

Antonopoulos, Andreas M. (2017). *Mastering bitcoin: Programming the open blockchain.* " O'Reilly Media, Inc.".

Antonopoulos, Andreas M., & Wood, Gavin (2018). *Mastering ethereum: building smart contracts and dapps*. O'reilly Media.

Apostolaki, Maria, Zohar, Aviv, & Vanbever, Laurent (2017). Hijacking bitcoin: Routing attacks on cryptocurrencies. In *2017 IEEE symposium on security and privacy (SP)* (pp. 375–392). IEEE.

Atzori, Marcella (2015). Blockchain technology and decentralized governance: Is the state still necessary? Available at SSRN 2709713.

Azouvi, Sarah, Maller, Mary, & Meiklejohn, Sarah (2018). Egalitarian society or benevolent dictatorship: The state of cryptocurrency governance. In *International conference on financial cryptography and data security* (pp. 127–143). Springer.

Bai, Qianlan, Zhang, Chao, Xu, Yuedong, Chen, Xiaowei, & Wang, Xin (2020). Evolution of ethereum: A temporal graph perspective. arXiv preprint arXiv: 2001.05251.

Bailey, Kenneth D. (1994). *Typologies and taxonomies: An introduction to classification techniques, Number 102*. Sage.

Baliga, Arati (2017). Understanding blockchain consensus models. *Persistent*, *2017*(4), 1–14.

Baniata, Hamza, Anaqreh, Ahmad, & Kertesz, Attila (2021). Pf-BTS: A privacy-aware fog-enhanced blockchain-assisted task scheduling. *Information Processing & Management*, *58*(1), Article 102393.

Beck, Roman, Avital, Michel, Rossi, Matti, & Thatcher, Jason Bennett (2017). *Blockchain technology in business and information systems research*. Springer.

Beck, Roman, Müller-Bloch, Christoph, & King, John Leslie (2018). Governance in the blockchain economy: A framework and research agenda. *Journal of the Association for Information Systems*, *19*(10), 1020–1034.

Beikverdi, Alireza, & Song, JooSeok (2015). Trend of centralization in bitcoin's distributed network. In *2015 IEEE/ACIS 16th international conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD)* (pp. 1–6). IEEE.

Berdik, David, Otoum, Safa, Schmidt, Nikolas, Porter, Dylan, & Jararweh, Yaser (2020). A survey on blockchain for information systems management and security. *Information Processing & Management*, *58*(1), Article 102397.

Bitcoin (2019). Bitcoin improvement proposals - github repository. URL=https://github.com/bitcoin/bips.

Blockchain luxembourg s. a (2019). Blocks mined. *Bitcoin Block Explorer and Currency Statistics*, URL=https://www.blockchain.com/btc/blocks.

Böhme, Rainer, Christin, Nicolas, Edelman, Benjamin, & Moore, Tyler (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, *29*(2), 213–238.

Bohr, Jeremiah, & Bashir, Masooda (2014). Who uses bitcoin? an exploration of the bitcoin community. In *2014 twelfth annual international conference on privacy, security and trust* (pp. 94–101). IEEE.

Bonneau, Joseph, Miller, Andrew, Clark, Jeremy, Narayanan, Arvind, Kroll, Joshua A., & Felten, Edward W. (2015). Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *2015 IEEE symposium on security and privacy* (pp. 104–121). IEEE.

Borge, Maria, Kokoris-Kogias, Eleftherios, Jovanovic, Philipp, Gasser, Linus, Gailly, Nicolas, & Ford, Bryan (2017). Proof-of-personhood: Redemocratizing permissionless cryptocurrencies. In *2017 IEEE European symposium on security and privacy workshops (EuroS&PW)* (pp. 23–26). IEEE.

Bradbury, Danny (2013). The problem with bitcoin. *Computer Fraud & Security*, *2013*(11), 5–8.

Briscoe, Neil (2000). Understanding the OSI 7-layer model. *PC Network Advisor*, *120*(2).

Bruschi, Francesco, Rana, Vincenzo, Gentile, Lorenzo, & Sciuto, Donatella (2019). Mine with it or sell it: the superhashing power dilemma. *ACM SIGMETRICS Performance Evaluation Review*, *46*(3), 127–130.

Buckley, Jim, & Exton, Christopher (2003). Bloom's taxonomy: A framework for assessing programmers' knowledge of software systems. In *11th IEEE international workshop on program comprehension, 2003.* (pp. 165–174). IEEE.

Buterin, Vitalik, et al. (2013). Ethereum white paper. *GitHub Repository*, 22–23.

Caccioli, Fabio, Livan, Giacomo, & Aste, Tomaso (2016). Scalability and egalitarianism in peer-to-peer networks. In *Banking beyond banks and money* (pp. 197–212). Springer.

Caffyn, Grace (2015). What is the bitcoin block size debate and why does it matter. URL: http://www.coindesk.com/what-is-the-bitcoin-block-size-debate-and-why-does-it-matter/**(visited on 27/11/2015).**

Casino, Fran, Dasaklis, Thomas K., & Patsakis, Constantinos (2019). A systematic literature review of blockchain-based applications: current status, classification and open issues. *Telematics and Informatics*, *36*, 55–81.

Chen, Qian, Srivastava, Gautam, Parizi, Reza M., Aloqaily, Moayad, & Ridhawi, Ismaeel Al (2020). An incentive-aware blockchain-based solution for internet of fake media things. *Information Processing & Management*, *57*(6), Article 102370.

Chen, Lin, Xu, Lei, Shah, Nolan, Gao, Zhimin, Lu, Yang, & Shi, Weidong (2017). On security analysis of proof-of-elapsed-time (poet). In *International symposium on stabilization, safety, and security of distributed systems* (pp. 282–297). Springer.

Chesterman, Xavier (2018). *The P2pool mining pool* (PhD thesis), Ghent University.

Chia, Vincent, Hartel, Pieter, Hum, Qingze, Ma, Sebastian, Piliouras, Georgios, Reijsbergen, Daniel, et al. (2018). Rethinking blockchain security: Position paper. arXiv preprint arXiv:1806.04358.

Cho, Hyungmin (2018). ASIC-Resistance of multi-hash proof-of-work mechanisms for blockchain consensus protocols. *IEEE Access*, *6*, 66210–66222.

Chohan, Usman W. (2019). Cryptocurrencies and inequality. Notes on the 21st Century (CBRI).

Chu, Dennis (2018). Broker-dealers for virtual currency: Regulating cryptocurrency wallets and exchanges. *Columbia Law Review*, *118*(8), 2323–2360.

Cong, Lin William, He, Zhiguo, & Li, Jiasun (2019). *Decentralized mining in centralized pools: Technical report*, National Bureau of Economic Research.

Conti, Mauro, Kumar, E. Sandeep, Lal, Chhagan, & Ruj, Sushmita (2018). A survey on security and privacy issues of bitcoin. *IEEE Communications Surveys & Tutorials*, *20*(4), 3416–3452.

Cryptoslate (2018). Ethereum network under assault: Gas price manipulation may indicate covert EOS attack [interview].

Dai, Mingjun, Zhang, Shengli, Wang, Hui, & Jin, Shi (2018). A low storage room requirement framework for distributed ledger in blockchain. *IEEE Access*, *6*, 22970–22975.

Davidson, Sinclair, De Filippi, Primavera, & Potts, Jason (2016). Economics of blockchain. Available at SSRN 2744751.

De Domenico, Manlio, & Baronchelli, Andrea (2019). The fragility of decentralised trustless socio-technical systems. *EPJ Data Science*, *8*(1), 2.

De Filippi, Primavera, & Loveluck, Benjamin (2016). The invisible politics of bitcoin: governance crisis of a decentralized infrastructure. *Internet Policy Review*, *5*(4).

Dietrich, Sven, Long, Neil, & Dittrich, David (2000). Analyzing distributed denial of service tools: The shaft case.. In *LISA* (pp. 329–339).

Dorfman, Robert (1979). A formula for the gini coefficient. *The Review of Economics and Statistics*, 146–149.

Dwivedi, Ashutosh Dhar, Srivastava, Gautam, Dhar, Shalini, & Singh, Rajani (2019). A decentralized privacy-preserving healthcare blockchain for IoT. *Sensors*, *19*(2), 326.

Ekblaw, Ariel, Barabas, Chelsea, Harvey-Buschel, Jonathan, & Lippman, Andrew (2016). Bitcoin and the myth of decentralization: Socio-technical proposals for restoring network integrity. In *2016 IEEE 1st international workshops on foundations and applications of self* systems (FAS* W)* (pp. 18–23). IEEE.

Esposito, Christian, Ficco, Massimo, & Gupta, Brij Bhooshan (2021). Blockchain-based authentication and authorization for smart city applications. *Information Processing & Management*, *58*(2), Article 102468.

Etherscan (2019a). Ethereum transaction information for the gensis block. URL=https://etherscan.io/txs?block=0.

Etherscan (2019b). Total ether supply and market capitalization. URL=https://etherscan.io/stat/supply.

Fanti, Giulia, Kogan, Leonid, Oh, Sewoong, Ruan, Kathleen, Viswanath, Pramod, & Wang, Gerui (2019). Compounding of wealth in proof-of-stake cryptocurrencies. In *International conference on financial cryptography and data security* (pp. 42–61). Springer.

Feld, Sebastian, Schönfeld, Mirco, & Werner, Martin (2014). Analyzing the deployment of bitcoin's P2p network under an AS-level perspective. *Procedia Computer Science*, *32*, 1121–1126.

Fleiss, Joseph L., & Cohen, Jacob (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619.

Galster, Matthias, Weyns, Danny, Tofan, Dan, Michalik, Bartosz, & Avgeriou, Paris (2013). Variability in software systems—a systematic literature review. *IEEE Transactions on Software Engineering*, *40*(3), 282–306.

Garay, Juan, Kiayias, Aggelos, & Leonardos, Nikos (2015). The bitcoin backbone protocol: Analysis and applications. In *Annual international conference on the theory and applications of cryptographic techniques* (pp. 281–310). Springer.

Gastwirth, Joseph L. (1971). A general definition of the lorenz curve. *Econometrica*, 1037–1039.

Gaži, Peter, Kiayias, Aggelos, & Russell, Alexander (2018). Stake-bleeding attacks on proof-of-stake blockchains. In *2018 crypto valley conference on blockchain technology (CVCBT)* (pp. 85–92). IEEE.

Gencer, Adem Efe, Basu, Soumya, Eyal, Ittay, Van Renesse, Robbert, & Sirer, Emin Gün (2018). Decentralization in bitcoin and ethereum networks. arXiv preprint arXiv:1801.03998.

Gervais, Arthur, Karame, Ghassan O., Capkun, Vedran, & Capkun, Srdjan (2014). Is bitcoin a decentralized currency? *IEEE Security & Privacy*, *12*(3), 54–60.

Gervais, Arthur, Karame, Ghassan O., Wüst, Karl, Glykantzis, Vasileios, Ritzdorf, Hubert, & Capkun, Srdjan (2016). On the security and performance of proof of work blockchains. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 3–16).

Gini, Corrado (1921). Measurement of inequality of incomes. *The Economic Journal*, *31*(121), 124–126.

Great Britain, Government Office for Science (2016). *Distributed ledger technology: Beyond block chain*. Government Office for Science.

Guegan, Dominique (2017). Public blockchain versus private blockhain. Documents de travail du Centre d'Economie de la Sorbonne 2017.20 - ISSN : 1955-611X.

Guerra García, José Manuel, Espinosa Torre, Free, & García Gómez, José Carlos (2008). Trends in taxonomy today: an overview about the main topics in taxonomy. *Zoológica Baetica*, *19*, 15–49.

Guo, Zhaohui, Gao, Zhen, Mei, Haojuan, Zhao, Ming, & Yang, Jinsheng (2019). Design and optimization for storage mechanism of the public blockchain based on redundant residual number system. *IEEE Access*, *7*, 98546–98554.

Guo, Ye, & Liang, Chen (2016). Blockchain application and outlook in the banking industry. *Financial Innovation*, *2*(1), 24.

Gupta, Manas, & Gupta, Parth (2017). Gini coefficient based wealth distribution in the bitcoin network: A case study. In *International conference on computing, analytics and networks* (pp. 192–202). Springer.

Gutoski, Gus, & Stebila, Douglas (2015). Hierarchical deterministic bitcoin wallets that tolerate key leakage. In *International conference on financial cryptography and data security* (pp. 497–504). Springer.

Halpin, Harry, & Piekarska, Marta (2017). Introduction to security and privacy on the blockchain. In *2017 IEEE European symposium on security and privacy workshops (EuroS&PW)* (pp. 1–3). IEEE.

Hardin, Taylor, & Kotz, David (2021). Amanuensis: Information provenance for health-data systems. *Information Processing & Management*, *58*(2), Article 102460.

He, Pu, Yu, Ge, Zhang, Y. F., & Bao, Y. B. (2017). Survey on blockchain technology and its application prospect. *Computer Science*, *44*(4), 1–7.

Heilman, Ethan, Kendler, Alison, Zohar, Aviv, & Goldberg, Sharon (2015). Eclipse attacks on bitcoin's peer-to-peer network. In *24th {USENIX} security symposium ({USENIX} security 15)* (pp. 129–144).

Hileman, Garrick, & Rauchs, Michel (2017). Global cryptocurrency benchmarking study. *Cambridge Centre for Alternative Finance*, *33*.

Hu, Teng, Liu, Xiaolei, Chen, Ting, Zhang, Xiaosong, Huang, Xiaoming, Niu, Weina, et al. (2021). Transaction-based classification and detection approach for ethereum smart contract. *Information Processing & Management*, *58*(2), Article 102462.

Huobi Blockchain Big Data Weekly Insights. Vol. 8.

Iansiti, Marco, & Lakhani, Karim R. (2017). The truth about blockchain. *Harvard Business Review*, *95*(1), 118–127.

Jin, Tong, Zhang, Xiang, Liu, Yirui, & Lei, Kai (2017). Blockndn: A bitcoin blockchain decentralized system over named data networking. In *2017 ninth international conference on ubiquitous and future networks (ICUFN)* (pp. 75–80). IEEE.

Jing, Nan, Liu, Qi, & Sugumaran, Vijayan (2021). A blockchain-based code copyright management system. *Information Processing & Management*, *58*(3), Article 102518.

Judmayer, Aljosha, Stifter, Nicholas, Krombholz, Katharina, & Weippl, Edgar (2017). Blocks and chains: Introduction to bitcoin, cryptocurrencies, and their consensus mechanisms. *Synthesis Lectures on Information Security, Privacy, & Trust*, *9*(1), 1–123.

Judmayer, Aljosha, Zamyatin, Alexei, Stifter, Nicholas, Voyiatzis, Artemios G., & Weippl, Edgar (2017). Merged mining: Curse or cure? In *Data privacy management, cryptocurrencies and blockchain technology* (pp. 316–333). Springer.

Karame, Ghassan (2016). On the security and scalability of bitcoin's blockchain. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 1861–1862). ACM.

Karame, Ghassan O., & Androulaki, Elli (2016). *Bitcoin and blockchain security*. Artech House.

Karame, Ghassan O., Androulaki, Elli, & Capkun, Srdjan (2012). Double-spending fast payments in bitcoin. In *Proceedings of the 2012 ACM conference on computer and communications security* (pp. 906–917).

Khairuddin, Irni Eliana, & Sas, Corina (2019). An Exploration of Bitcoin mining practices: Miners' trust challenges and motivations. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (1–13).

Khalid, Adia, Iftikhar, Muhammad Sohaib, Almogren, Ahmad, Khalid, Rabiya, Afzal, Muhammad Khalil, & Javaid, Nadeem (2021). A blockchain based incentive provisioning scheme for traffic event validation and information storage in VANETs. *Information Processing & Management*, *58*(2), Article 102464.

Kiayias, Aggelos, Russell, Alexander, David, Bernardo, & Oliynykov, Roman (2017). Ouroboros: A provably secure proof-of-stake blockchain protocol. In *Annual international cryptology conference* (pp. 357–388). Springer.

Kim, Chang Yeon, & Lee, Kyungho (2018). Risk management to cryptocurrency exchange and investors guidelines to prevent potential threats. In *2018 international conference on platform technology and service (PlatCon)* (pp. 1–6). IEEE.

Kim, Seoung Kyun, Ma, Zane, Murali, Siddharth, Mason, Joshua, Miller, Andrew, & Bailey, Michael (2018). Measuring ethereum network peers. In *Proceedings of the internet measurement conference 2018* (pp. 91–104).

Kim, Tae Wan, & Zetlin-Jones, Ariel (2019). The ethics of blockchain networks] the ethics of contentious hard forks in blockchain networks with fixed features. *Frontiers in Blockchain, 2*, 9.

Kitchenham, Barbara (2004). Procedures for performing systematic reviews.

Kondor, Dániel, Pósfai, Márton, Csabai, István, & Vattay, Gábor (2014). Do the rich get richer? An empirical analysis of the bitcoin transaction network. *PloS One, 9*(2).

Kwon, Yujin, Liu, Jian, Kim, Minjeong, Song, Dawn, & Kim, Yongdae (2019). Impossibility of full decentralization in permissionless blockchains. arXiv preprint arXiv:1905.05158.

Landis, J. Richard, & Koch, Gary G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

Lee, Wei-Meng (2019). Using the web3. js APIs. In *Beginning ethereum smart contracts programming* (pp. 169–198). Springer.

Lewenberg, Yoad, Bachrach, Yoram, Sompolinsky, Yonatan, Zohar, Aviv, & Rosenschein, Jeffrey S (2015). Bitcoin mining pools: A cooperative game theoretic analysis. In *Proceedings of the 2015 international conference on autonomous agents and multiagent systems* (pp. 919–927). International Foundation for Autonomous Agents and Multiagent Systems.

Li, Xiaoqi, Jiang, Peng, Chen, Ting, Luo, Xiapu, & Wen, Qiaoyan (2017). A survey on the security of blockchain systems. *Future Generation Computer Systems*.

Li, Jiaxing, Wu, Jigang, Jiang, Guiyuan, & Srikanthan, Thambipillai (2020). Blockchain-based public auditing for big data in cloud storage. *Information Processing & Management, 57*(6), Article 102382.

Liao, Kevin, & Katz, Jonathan (2017). Incentivizing blockchain forks via whale transactions. In *International conference on financial cryptography and data security* (pp. 264–279). Springer.

Lindley, John (1836). *A Natural System of Botany, or, A systematic view of the organization, natural affinities, and geographical distribution, of the whole vegetable kingdom: together with the uses of the most important species in medicine, the arts, and rural or domestic economy*. Longman, Rees, Orme, Brown, Green, and Longman.

Malík, Vladimír (2016). The history and the future of bitcoin. *Praha: Bankovní Institut Vysoká škola Praha*.

Marvin, Ian (2017). *Decentralised? A Study of Concentration in the Bitcoin Network* (PhD thesis), University of Cape Town.

Mattila, Juri (2016). *The blockchain phenomenon–the disruptive potential of distributed consensus architectures: Technical report*, ETLA working papers.

Maxwell, Joseph (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*(3), 279–301.

Meijer, David, & Ubacht, Jolien (2018). The governance of blockchain systems from an institutional perspective, a matter of trust or control? In *Proceedings of the 19th annual international conference on digital government research: Governance in the data age* (pp. 1–9).

Miller, Andrew, Litton, James, Pachulski, Andrew, Gupta, Neal, Levin, Dave, Spring, Neil, et al. (2015). Discovering bitcoin's public topology and influential nodes. et al.

Mingxiao, Du, Xiaofeng, Ma, Zhe, Zhang, Xiangwei, Wang, & Qijun, Chen (2017). A review on consensus algorithm of blockchain. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 2567–2572). IEEE.

Nakamoto, Satoshi (2008). Bitcoin: A peer-to-peer electronic cash system.

Neudecker, Till, & Hartenstein, Hannes (2018). Network layer aspects of permissionless blockchains. *IEEE Communications Surveys & Tutorials, 21*(1), 838–857.

Nguyen, Cong T, Hoang, Dinh Thai, Nguyen, Diep N, Niyato, Dusit, Nguyen, Huynh Tuong, & Dutkiewicz, Eryk (2019). Proof-of-stake consensus mechanisms for future blockchain networks: fundamentals, applications and opportunities. *IEEE Access, 7*, 85727–85745.

Nickerson, Robert C., Varshney, Upkar, & Muntermann, Jan (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems, 22*(3), 336–359.

Oberländer, Anna Maria, Lösser, Benedict, & Rau, Daniel (2019). Taxonomy research in information systems: A systematic assessment.

O'Dwyer, Karl J., & Malone, David (2014). *Bitcoin mining and its energy footprint*. IET.

Oham, Chuka, Michelin, Regio A., Jurdak, Raja, Kanhere, Salil S., & Jha, Sanjay (2021). B-FERL: Blockchain based framework for securing smart vehicles. *Information Processing & Management, 58*(1), Article 102426.

Opdenakker, Raymond (2006). Advantages and disadvantages of four interview techniques in qualitative research. In *Forum qualitative sozialforschung/forum: qualitative social research, Vol. 7*.

Panarello, Alfonso, Tapas, Nachiket, Merlino, Giovanni, Longo, Francesco, & Puliafito, Antonio (2018). Blockchain and iot integration: A systematic survey. *Sensors, 18*(8), 2575.

Peck, Morgen E. (2017). Blockchain world-do you need a blockchain? This chart will tell you if the technology can solve your problem. *IEEE Spectrum, 54*(10), 38–60.

Petersen, Kai, Feldt, Robert, Mujtaba, Shahid, & Mattsson, Michael (2008). Systematic mapping studies in software engineering. In *12th international conference on evaluation and assessment in software engineering (EASE) 12* (pp. 1–10).

Pilkington, Marc (2019). The libra project: A transnational monetary dystopia–analysis of the disruption generated by the facebook-led stable coin. Available at SSRN 3434079.

Pustišek, Matevž, Umek, Anton, & Kos, Andrej (2019). Approaching the communication constraints of ethereum-based decentralized applications. *Sensors, 19*(11), 2647.

Putz, Benedikt, Dietz, Marietheres, Empl, Philip, & Pernul, Günther (2021). Ethertwin: Blockchain-based secure digital twin information management. *Information Processing & Management, 58*(1), Article 102425.

Raman, Ravi Kiran, & Varshney, Lav R. (2017). Dynamic distributed storage for scaling blockchains. arXiv preprint arXiv:1711.07617.

Ramona, Orăștean, Cristina, Mărginean Silvia, Raluca, Sava, et al. (2019). Bitcoin in the scientific literature–a bibliometric study. *Studies in Business and Economics, 14*(3), 160–174.

Razzaq, Abdul, Wasala, Asanka, Exton, Chris, & Buckley, Jim (2018). The state of empirical evaluation in static feature location. *ACM Transactions on Software Engineering and Methodology (TOSEM), 28*(1), 1–58.

Reddit (2019). R/monero - blockchain size issue in future? reddit, URL=https://www.reddit.com/r/Monero/comments/9gymaf/blockchain_size_issue_in_future/.

Roubini, Nouriel (2018a). The big blockchain lie. *Project Syndicate. Blog Post, 15*.

Roubini, Nouriel (2018b). Blockchain isn't about democracy and decentralisation – it's about greed | nouriel roubini.

Sai, Ashish Rajendra, Buckley, Jim, & Le Gear, Andrew (2019a). Assessing the security implication of bitcoin exchange rates. *Computers and Security*.

Sai, Ashish Rajendra, Buckley, Jim, & Le Gear, Andrew (2019b). Privacy and security analysis of cryptocurrency mobile applications. In *2019 fifth conference on mobile and secure services (mobisecserv)* (pp. 1–6). IEEE.

Sai, Ashish Rajendra, Le Gear, Andrew, & Buckley, Jim (2019). Centralization threat metric.

Sapirshtein, Ayelet, Sompolinsky, Yonatan, & Zohar, Aviv (2016). Optimal selfish mining strategies in bitcoin. In *International conference on financial cryptography and data security* (pp. 515–532). Springer.

Saroiu, Stefan, Gummadi, P. Krishna, & Gribble, Steven D. (2001). Measurement study of peer-to-peer file sharing systems. In *Multimedia computing and networking 2002, Vol. 4673* (pp. 156–170). International Society for Optics and Photonics.

Sayeed, Sarwar, & Marco-Gisbert, Hector (2019). Assessing blockchain consensus and security mechanisms against the 51% attack. *Applied Sciences*, *9*(9), 1788.

Sergio (2013). The well deserved fortune of satoshi nakamoto, bitcoin creator, visionary and genius.

Sim, Julius, & Wright, Chris C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257–268.

Srinivasan, Balaji S. (2017). Quantifying decentralization. *Medium*.

StopAndDecrypt (2018). The ethereum-blockchain size has exceeded 1tb, and yes, it's an issue.

Szabo, Nick (1970). The dawn of trustworthy computing.

Tapsell, James, Akram, Raja Naeem, & Markantonakis, Konstantinos (2018). An evaluation of the security of the bitcoin peer-to-peer network. 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)* (pp. 1057–1062). IEEE.

Walport, Mark (2016). *Distributed ledger technology: beyond blockchain. UK Government Office for Science: Technical report*, Tech. Rep.

Wang, Wenbo, Hoang, Dinh Thai, Xiong, Zehui, Niyato, Dusit, Wang, Ping, Hu, Peizhao, et al. (2018). A survey on consensus mechanisms and mining management in blockchain networks. (pp. 1–33). arXiv preprint arXiv:1805.02707.

Wang, Sha, Vergne, Jean-Philippe J. P., & Hsieh, Ying-Ying (2017). The internal and external governance of blockchain-based organizations: Evidence from cryptocurrencies. In *Bitcoin and beyond* (pp. 48–68). Routledge.

Wirdum, Aaron van (2016). Rejecting today's hard fork, the ethereum classic project continues on the original chain: Here's why. *Bitcoin Magazine*, *20*.

Wolfson, Shael N. (2015). Bitcoin: the early market. *Journal of Business & Economics Research (JBER)*, *13*(4), 201–214.

Wood, Gavin, et al. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper*, *151*(2014), 1–32.

World Bank. GINI index (World Bank estimate). URL=https://data.worldbank.org/indicator/si.pov.gini.

Wüst, Karl, & Gervais, Arthur (2018). Do you need a blockchain? In *2018 crypto valley conference on blockchain technology (CVCBT)* (pp. 45–54). IEEE.

Xie, Junfeng, Tang, Helen, Huang, Tao, Yu, F Richard, Xie, Renchao, Liu, Jiang, et al. (2019). A survey of blockchain technology applied to smart cities: Research issues and challenges. *IEEE Communications Surveys & Tutorials*, *21*(3), 2794–2830.

Xu, Xiaoqiong, Sun, Gang, Luo, Long, Cao, Huilong, Yu, Hongfang, & Vasilakos, Athanasios V. (2021). Latency performance modeling and analysis for hyperledger fabric blockchain network. *Information Processing & Management*, *58*(1), Article 102436.

Xu, Xiwei, Weber, Ingo, Staples, Mark, Zhu, Liming, Bosch, Jan, Bass, Len, et al. (2017). A taxonomy of blockchain-based systems for architecture design. In *2017 IEEE international conference on software architecture (ICSA)* (pp. 243–252). IEEE.

Yli-Huumo, Jesse, Ko, Deokyoon, Choi, Sujin, Park, Sooyong, & Smolander, Kari (2016). Where is current research on blockchain technology?—A systematic review. *PLoS One*, *11*, Article e0163477.

Zhang, Rui, Xue, Rui, & Liu, Ling (2019). Security and privacy on blockchain. *ACM Computing Surveys*, *52*(3), 1–34.

Zhao, Quanyu, Chen, Siyi, Liu, Zheli, Baker, Thar, & Zhang, Yuan (2020). Blockchain-based privacy-preserving remote data integrity checking scheme for IoT information systems. *Information Processing & Management*, *57*(6), Article 102355.

Zheng, Zibin, Xie, Shaoan, Dai, Hongning, Chen, Xiangping, & Wang, Huaimin (2017). An overview of blockchain technology: Architecture, consensus, and future trends. In *2017 IEEE international congress on big data (BigData Congress)* (pp. 557–564). IEEE.

Zheng, Zibin, Xie, Shaoan, Dai, Hong-Ning, Chen, Xiangping, & Wang, Huaimin (2018). Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, *14*(4), 352–375.

Zhu, Qingyi, Loke, Seng W, Trujillo-Rasua, Rolando, Jiang, Frank, & Xiang, Yong (2019). Applications of distributed ledger technologies to the internet of things: A survey. *ACM Computing Surveys*, *52*(6), 1–34.

Zmudzinski, Adrian (2020). Decentralized lending protocol bzx hacked twice in a matter of days.