2020

Measuring Decentrality in Blockchain Based Systems

Sarada Prasad Gochhayat
Old Dominion University, sgochhay@odu.edu

Sachin Shetty
Old Dominion University, sshetty@odu.edu

Ravi Mukkamala
Old Dominion University, mukka@cs.odu.edu

Peter Foytik
Old Dominion University, pfoytik@odu.edu

Georges A. Kamhoua
Old Dominion University, gkamhoua@odu.edu

See next page for additional authors

Original Publication Citation

Gochhayat, S. P., Shetty, S., Mukkamala, R., Foytik, P., Kamhoua, G. A., & Njilla, L. (2020). Measuring decentrality in blockchain based systems. *IEEE Access, 8*, 178372-178390. https://doi.org/10.1109/ACCESS.2020.3026577

Authors
Sarada Prasad Gochhayat, Sachin Shetty, Ravi Mukkamala, Peter Foytik, Georges A. Kamhoua, and Laurent Njilla

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 31, 2020, accepted September 15, 2020, date of publication September 24, 2020, date of current version October 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3026577

Measuring Decentrality in Blockchain Based Systems

SARADA PRASAD GOCHHAYAT[1], SACHIN SHETTY[1], (Senior Member, IEEE), RAVI MUKKAMALA[2], (Associate Member, IEEE), PETER FOYTIK[1], GEORGES A. KAMHOUA[1], AND LAURENT NJILLA[3], (Member, IEEE)

[1]Virginia Modeling Analysis and Simulation Center, Old Dominion University, Norfolk, VA 23529, USA
[2]Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA
[3]Cyber Assurance Branch, Air Force Research Laboratory, Rome, NY 13441, USA

Corresponding author: Sarada Prasad Gochhayat (sgochhay@odu.edu)

ABSTRACT Blockchain promises to provide a distributed and decentralized means of trust among untrusted users. However, in recent years, a shift from decentrality to centrality has been observed in the most accepted Blockchain system, i.e., Bitcoin. This shift has motivated researchers to identify the cause of decentrality, quantify decentrality and analyze the impact of decentrality. In this work, we take a holistic approach to identify and quantify decentrality in Blockchain based systems. First, we identify the emergence of centrality in three layers of Blockchain based systems, namely governance layer, network layer and storage layer. Then, we quantify decentrality in these layers using various metrics. At the governance layer, we measure decentrality in terms of fairness, entropy, Gini coefficient, Kullback–Leibler divergence, etc. Similarly, in the network layer, we measure decentrality by using degree centrality, betweenness centrality and closeness centrality. At the storage layer, we apply a distribution index to define centrality. Subsequently, we evaluate the decentrality in Bitcoin and Ethereum networks and discuss our observations. We noticed that, with time, both Bitcoin and Ethereum networks tend to behave like centralized systems where a few nodes govern the whole network.

INDEX TERMS Blockchain, centrality, measurement, analysis.

I. INTRODUCTION

In recent years, the world has recognized Blockchain as one of the technological advances to provide distributed and decentralized means of trust among untrusted peers [1]–[3] (See Fig. 1). Although there have been several academic and corporate efforts to design peer-to-peer network-based systems, Blockchain stands out among the rest because of its ability to withstand Sybil attacks [4], [5]. Normally, in a Sybil attack, the attackers gain a large influence in a peer-to-peer network based system by creating a large number of pseudonymous identities. Hence, this attack can impede peer-to-peer network-based systems which solely rely on distributed and decentrality properties. Baran, in his seminal work [6], first discussed centralized, decentralized and distributed networks. Later, Buterin [7] used Baran's work as a framework to discuss the meaning of decentrality in a Blockchain system.

Decentralization in distributed systems refers to the fact that there is no central point of control among distributed and connected peers or nodes [8]. It ensures that no single user or a group of users could control a system's assets or impose changes which other users don't consent to. Hence, decentralization increases the number of decision-makers and thereby removes the need to trust a central authority [9].

Although, in distributed systems, most nodes work autonomously to achieve a common goal, there are some nodes with more important roles than others, for example, miners in bitcoin [10], [11] or super-nodes in bittorrent [12]. These important nodes within a decentralized system could create a potential centrality.

Furthermore, decentralized system design does not ensure decentralized control. Even the most apparently decentralized systems have shown the ability to produce structurally centralized control. For example, the early decentralized technologies of the Internet and Web relied on key points of centralization, such as the Domain Name System and the World Wide Web Consortium.

The associate editor coordinating the review of this manuscript and approving it for publication was Chien-Ming Chen.

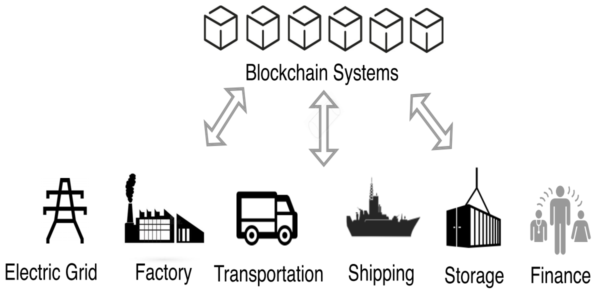


FIGURE 1. Application of Blockchain.

Recently, based on [7], a similar trend has also been observed in Blockchain based systems. For example, the location of nodes running a Blockchain system can gravitate towards centrality [13]. Similarly, the existence of Bitcoin consortium which maintains Bitcoin source code also shows central authority of the code [7].

Motivation: Recognizing the presence of centrality in a Blockchain system is very important since Blockchains offer distinct, potentially liberating opportunities for reinventing financial systems, organizations, governance, supply chains, and more [14]. When we recognize that the subsystems of a decentralized system can have centralizing effect, we can prevent such centralization of the subsystems to undermine a decentralized design [15]. In order to make these systems truly decentralized, we must ensure that each and every subsystem should minimize any central points of control.

A scientific quantification of decentralization is critical in assessing the level of decentralization in Blockchain-based systems. Most of the previous work on defining centrality in Blockchain considered only a particular aspect of Blockchain. For example, in [16] the authors focus on network topology using graph theory, and [7] considers the economic aspect to define decentrality.

In [17], Bach *et al.* analyzed the algorithmic steps taken by various consensus algorithms and carried out a comparative study on energy saving and tolerated power of adversary. Similarly, in [18], Porat, *et al.*, analyzed Blockchain consensus in terms of problem complexity. In [19], Ren modeled the consensus algorithm as a stochastic model and provided an insight into the impact of network and hashing on forking. Thus, the previous works captured only certain aspects of Blockchain.

Contribution: In this work, we take a holistic approach to quantify decentrality.

- First, we identify the emergence of centrality in different layers of Blockchain, namely governance layer, network layer and storage layer.
- Second, we present various metrics which capture the decentrality in respective layers. In particular, for measuring decentrality in the governance layer, we use fairness index, entropy, Gini coefficient, Euclidean distance, Minkowski distance, cosine similarity and Kullback-Leibler divergence metrics. Similarly, to measure decentrality in the network layer, we use degree centrality, betweenness centrality and closeness centrality metrics. For storage layer decentrality measurement, we use a distribution index, which captures the idea of how well the storage information is distributed among several storage systems.
- Third, we illustrate our methodology by evaluating the decentrality for Bitcoin and Ethereum networks. We first create baseline metric measurements for the three layers. Then, we gather data from Bitcoin and Ethereum systems, and calculate the decentrality metrics for the three layers. Finally, we compare the baseline measurements with the Bitcoin and Ethereum decentrality metrics.

The rest of the paper is organized as follows: Section II discusses background and some of the related work; Section III presents different metrics to quantify decentrality; Section IV discusses our results; and finally Section V presents the conclusions and future works.

II. BACKGROUND AND RELATED WORK

In this section, we discuss both permissioned and permissionless Blockchains, the concept of decentrality, benefits and needs of decentrality and distributed storage for Blockchains.

A. BLOCKCHAIN

A Blockchain is a special kind of distributed and decentralized system, which helps users or nodes, who cannot trust each other, to reach a consensus without relying on a single centralized controlling entity [5], [20]. It is a chain of blocks where each block contains a set of records. Some special nodes, *aka* validating nodes or miners, add blocks to the Blockchain through a procedure called "mining" [10]. In case more than one miner adds blocks to the chain simultaneously, a fork occurs in the Blockchain. In such a case, the "longest-chain" rule is applied, where the nodes follow the branch containing the most number of blocks [16]. By design, Blockchain is tamper-proof, and once a new block has been recorded, the data in that block cannot be altered retroactively [21], [22].

1) PERMISSIONED BLOCKCHAIN

In Permissioned Blockchains, nodes trust only a set of validating nodes and a governing authority [23]. Hence, they are more centralized. These systems sacrifice some of their decentrality for better scalability and performance [24], [25]. By having a governing authority that provides an inherent level of trust between participants, these Blockchain systems enable design decisions, such as sharding and channels, to be implemented without much complexity. Furthermore, the governing body can enforce data access controls to participants in the channel and only allow them to view sensitive transaction data [26].

2) PERMISSIONLESS BLOCKCHAIN

In permissionless Blockchains, nodes do not particularly trust any set of validating nodes. Instead, they try to reach a

TABLE 1. Summary of the related work for the Measurement of Blockchain based Systems.

Papers	Approaches
[27], [28]	Valdivia et al. analyzed decentrality of cryptocurrencies in terms of distributed transaction processing among different entities, in a theoretical fashion without quantification.
[29]	Chu and Wang defined centralization level, a quantitative measures of Blockchain decentralization, which reflects the distributions of transactions contributed by Blockchain providers.
[30]	Srinivasan defined decentrality using Gini coefficient and Lorenz curve; and observed that the systems unintentionally may move to centrality.
[8]	Wu et al. proposed an entropy method to quantify the decentralization for Bitcoin and Ethereum. Using the information entropy, they measure the discrete degrees of blocks mined and address balances to quantify the degrees of decentralization for Blockchain systems.
[27], [31]	Croman et al. first discussed layering of Blockchain systems, namely mining, transaction validation and storage. However, they didn't quantify the decentrality holistically.
[7]	Buterin considered the economic aspect to define decentrality.
[16]	Lischke and Fabian focused on network topology using graph theory, and used network metrics such as Degree centrality, betweenness centrality and closeness centrality to analyze the structure and the dynamics of the Bitcoin network.
[19]	Ren modeled the consensus algorithm as a stochastic model and gives insight into the impact of network and hashing on forking. The work discussed so far only tried to capture certain aspects of Blockchain.
[17]	Bach et al. analyzed the algorithmic steps taken by various consensus algorithms and did comparative study on energy saving and tolerated power of adversary.
[18]	Porat, et al., analyzed Blockchain consensus in terms of problem complexity.
[32]	Zhuang et al. mentioned how, in a decentralized cloud setting, a group of small data centers can cooperate with each others to improve the performance.
[33]	Raman and Varshney proposed a distributed storage system by combining private key encryption and Shamir's secret sharing scheme, which distributes transaction data without significant loss in data integrity.

consensus in a probabilistic or incentivised way, obviating the need to rely on centralized authorities [34]. The lack of a central authority enables Blockchain systems to provide stronger integrity and liveliness properties for distributed systems. Because there is no central authority in control, corrupting a Blockchain or hampering the propagation of its contents are only possible through collusion among powerful nodes. Consequently, permissionless Blockchains can achieve higher resistance against tampering and censorship, even in the presence of malicious nodes [23], [35].

B. DECENTRALITY

Decentralized systems are a subset of distributed systems where multiple authorities control different components and no authority is fully trusted by all [6], [7]. Decentrality is a property related to the control over the system. Better decentralization means higher resistance against censorship and tampering.

In Bitcoin and Ethereum systems, nodes generate blocks at a rate proportional to their computational power. Despite envisioned decentralization in Bitcoin, the high cost of mining has led to considerable centralization of consensus in practice. In order to share the risk of spending the resources but failing to win the competition, groups of miners form mining pools. This resulted in just a few mining pools validating most transactions. Better decentralization of miners means higher resistance against censorship of individual transactions and consequently a higher trust in the system [36]. In [27], [28] and [30], the authors discuss several ways to define decentrality and the ways in which systems unintentionally may move to centrality. In [29], the authors introduce a new metric, the centralization level, a quantitative measure of Blockchain decentralization, reflecting the distributions of transactions contributed by different Blockchain providers. A Blockchain is N_ϵ centralized if the top N nodes performed more than $1-\epsilon$ fraction of transactions. A Blockchain is more centralized if it has a smaller N and incurs a small centralization if N is large. Hence, a public chain's central trust is $T = N_{0.49}$; a consortium chain's central trust level is $T = N_{0.33}$; and a private chain's central trust level is $T = N_1$. Wu *et al.*, in [8], proposed an entropy method to quantify the decentralization for Bitcoin and Ethereum. Using the information entropy, they measure the discrete degrees of blocks mined and address balances to quantify the degrees of decentralization for Blockchain systems.

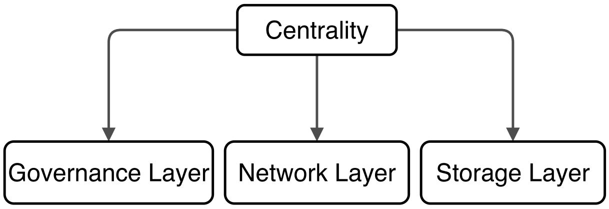


FIGURE 2. Three Layers of Decentrality in Blockchain systems.

C. BENEFITS OF DECENTRALITY

The foundation of centralized systems is the absence of mutual trust among nodes or users, so they need a trusted intermediary to cooperate with each other. The problem with

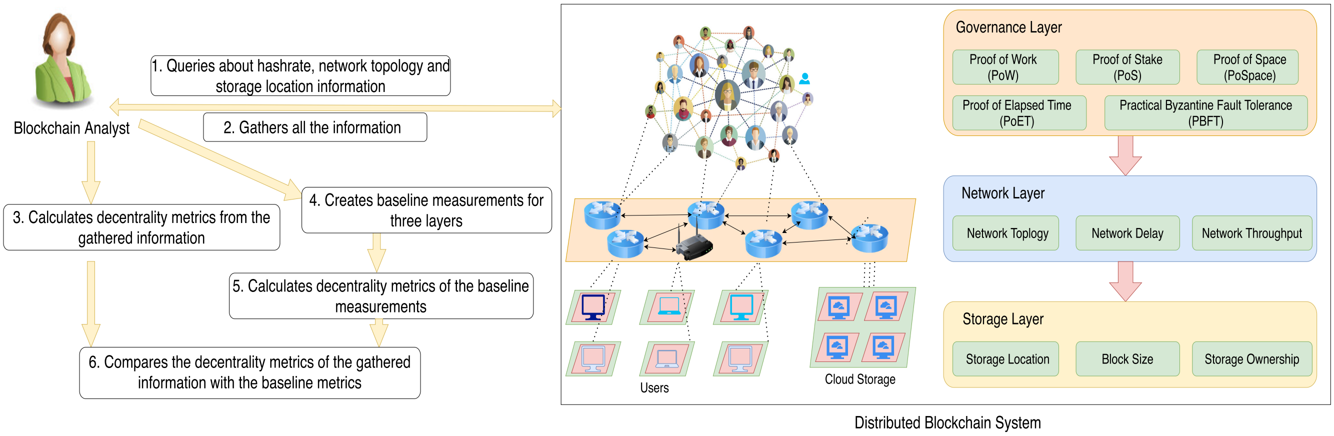

Blockchain Analyst
1. Queries about hashrate, network topology and storage location information
2. Gathers all the information
3. Calculates decentrality metrics from the gathered information
4. Creates baseline measurements for three layers
5. Calculates decentrality metrics of the baseline measurements
6. Compares the decentrality metrics of the gathered information with the baseline metrics
Users
Cloud Storage
Governance Layer
Proof of Work (PoW)
Proof of Stake (PoS)
Proof of Space (PoSpace)
Proof of Elapsed Time (PoET)
Practical Byzantine Fault Tolerance (PBFT)
Network Layer
Network Toplogy
Network Delay
Network Throughput
Storage Layer
Storage Location
Block Size
Storage Ownership
Distributed Blockchain System

FIGURE 3. Generic Architecture of decentrality measurement of Blockchain systems consisting Blockchain analyst and Distributed Blockchain system along with its submodules with layered view.

centralized systems is that they lack transparency, and therefore allow for single point of failure, censorship and abuse of power. Hence, there is a need for decentralized systems. The benefits of the decentralized systems are:

- Trust: Users do not put their trust in a central authority and instead put their trust on each other, hence any modification to data on the Blockchain by anyone can be observed by all others [29]. Here, users believe that no "trusted" group that exerts control could seize their assets or impose changes they did not consent to [12].
- Immutability: The data present in the Blockchain systems can not be changed, i.e., only read and append functionalities are applicable, not deletion [27], [37].
- Robustness: Failure of a node cannot take down the entire network as data is distributed across multiple Blockchain nodes. This results in high level of data availability. Even if a large number of nodes fail or are shut down by an attack, the data is still available for access at other nodes.
- Attack resistance: Decentralized systems are more expensive to attack, destroy or manipulate because they do not suffer from sensitive central points that can be attacked at much lower cost than the economic size of the system [38].
- Collusion resistance: It is much harder for nodes in a decentralized system to collude to act in ways that benefit them at the expense of other nodes [39].
- Central censorship free: There is no censorship. In a decentralized system, it is very difficult for a single party to censor communication traffic over the network [11].

D. REQUIREMENTS OF DECENTRALITY

For a system to claim to be decentralized, several requirements must be met:

- The system should not depend on a trusted third party.
- Any node can submit a transaction, i.e., every node has the right to submit to the Blockchain.
- Any node can validate a transaction. For example, in Bitcoin, it is envisioned that anyone can validate and add transactions.
- The distribution of effective power among the validating nodes should be even. It means some nodes should not have more control over how the chain should be extended.
- The incentive system for running the Blockchain should be fair. Otherwise, it may result in formation of a coalition of nodes and thereby reducing the number of independent nodes.

E. DISTRIBUTED STORAGE

An important point of vulnerability in Blockchain based systems is the storage of the Blockchain system [40], [41]. Running a Blockchain system on a particular cloud infrastructure makes Blockchain prone to a single point of failure. For example, Denial of Service attacks can disrupt cloud based systems [33], [42]. Similar trends have been observed in the most notable of distributed systems, Domain Name Systems. In order to overcome the single point failure, many organizations get the IP addresses from various DNS providers. Blockchain systems also have a similar vulnerability. To address rising storage costs and increasing transaction volumes in Blockchain systems, Raman and Varshney [33], proposed a distributed storage system, by combining private key encryption and Shamir's secret sharing scheme, which distributes transaction data without significant loss in data integrity. Furthermore, not only the distribution of the data, but also the impact of the workload in the data centers and arrival of requests to the data centers need to be considered to address the vulnerability in Blockchain systems [43], [44]. Zhuang *et al.* [32] show how, in a decentralized cloud setting, a group of small data centers can cooperate with each other to improve the performance.

III. QUANTIFYING DECENTRALITY

In this section, we first identify the emergence of centrality in three layers of Blockchain (See Fig. 2). Then, we present a generic process of decentrality measurement. Thereafter, we present various metrics which capture decentrality in respective layers.

Blockchain system may be conceptualized as consisting of three layers [27], [31], namely, governance layer, network layer and storage layer (see Fig. 3). In the governance layer, the nodes reach a consensus that is logically defined in the source code by selecting a leader or by probabilistically using a governance protocol. In Bitcoin, the node governance is achieved by creating a block and reaching consensus about the block by using proof of work (PoW) and the longest chain rule. Here, proof of work depends on the hash rate of individual nodes. The logic of the governance protocol changes as the Blockchain designers modify the code. Although miner nodes create blocks, a block is not accepted in the blockchain unless it reaches other nodes over a network. In the network layer, the network capacity in terms of topology and QoS play important roles to achieve a consensus. Another important parameter that delays the consensus process is the block size and block storage. If the block size is large, it delays the process of reaching consensus and the ability to validate the Blockchain. Similarly, if Blocks are stored or processed at a single cloud service provider, it might lead to single point failure. Hence, the parameter layer covers both block size and block storage. So clearly, centrality could arise because of node governance protocol, network structure, and cloud storage.

A. THE PROCESS OF DECENTRALITY MEASUREMENT

In this section, we discuss a generic process to measure the decentrality in a Blockchain-based system. First, we discuss the actors in the measurement. Then, we discuss the steps taken by the actors to calculate the decentrality of Blockchain. Subsequently, we also provide the pseudo-code of the Blockchain centrality measurement, in which we discuss the details of the involved sub-processes.

There are two actors in the process, namely, the Blockchain analyst and the Blockchain system (See Fig.3). Blockchain analyst wants to design and measure decentrality of Blockchain systems. There are six steps involved in the measurement. Blockchain analyst queries and gathers the information about the three layers of Blockchain (See Fig. 4). She calculates various metrics associated with the three layers and baseline measurements. Then compares different decentralized systems to see how centralized or decentralized the system being considered is. Algorithm 1 shows the process in great detail.

Decentrality is quantified in three dimensions:

B. GOVERNANCE LAYER

One way to quantify decentrality is based on nodes participating in the block creation or mining process. In particular, the number of nodes, or the number of organizations actually controlling the nodes, and their power measured in hash rate. Here, hash rate is the ability of a node to perform hash computations within a time interval.

Algorithm 1 Pseudocode for Decentality Measurement Process

```
1) Blockchain analyst queries and gathers Blockchain system about hashrate, network topology and storage location information from Blockchain system
2) Using the gathered information, the analyst calculates the following metrics for the three layers:
    a) For the governance layer, analyst calculates fairness index, entropy, Gini coefficient, distance, and similarity measures.
    b) For the network layer, analyst calculates degree, betweenness, and closeness centrality.
    c) For the storage layer, analyst calculates distribution index.
3) Baseline measurements are computed for the three layers based on the information received (See Table 2)
4) Analysts use the baseline measurements to calculate the metrics.
5) The comparison metrics derived from the gathered data of the Blockchain system and the baseline measurements, i.e., for governance layer, compare the result with Table 4. Similarly, for network and storage layer, comparisons of the values shown in Figure 16, 17, 18, 19, and 20
```

A simple metric of decentralization is simply the number of nodes. Here, the more nodes a Blockchain network has the more decentralized it is. The marginal utility of each additional node decreases with the network size. For example, a network with 1 million nodes is just as decentralized as a network with 1 billion nodes.

In order to consider the marginal utility of additional nodes, we could use logarithmic scale. Under a logarithmic scale a network with one node is maximally centralized and has minimum decentrality. The decentrality increases logarithmically with addition of more nodes. A network with 128 nodes would have a decentrality of 7, where a network of 1024 would have 10. Similarly, Bitcoin network with about 6000 to 7000 full nodes would have decentrality 12.5 if we only count the full nodes.

Moreover, it is the individual hash rate power of block miners which controls the fate of the network, in the short-term. They have the power to create forks in a blockchain. Hence, it is not enough to have 1000 participating nodes if 1 node has power to produce 51 percent of the blocks. All that matters is how many nodes can collaborate to generate 51 percent blocks and to some extent what is the largest percentage of block production held by a single individual.

Hence, by observing the probability of a node being selected as a leader or successfully creating a block, we can see whether the Blockchain is moving towards centrality. The Blockchain system assumes that every node has equal probability of being selected in the long-term. When some nodes have a higher probability of being selected than others,

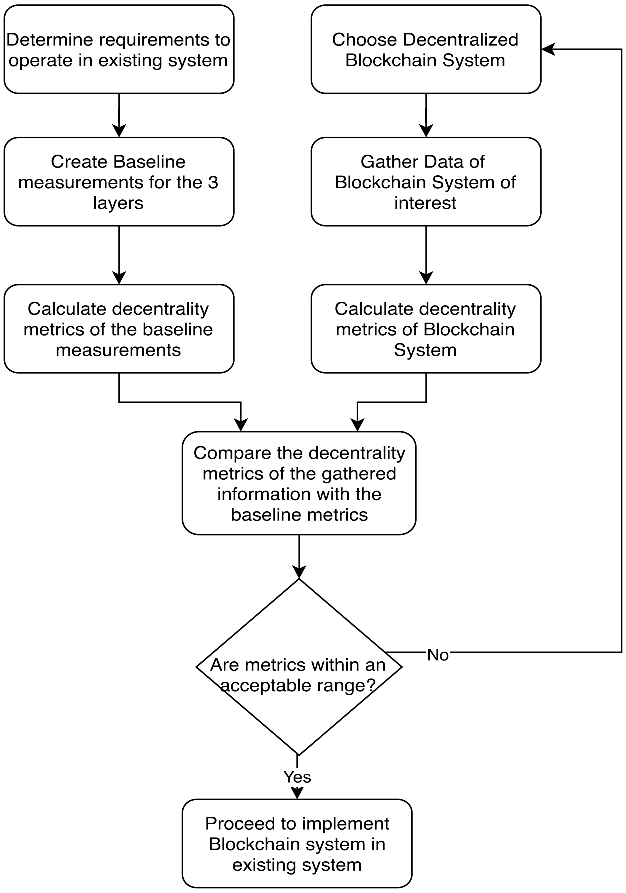

Determine requirements to operate in existing system
Create Baseline measurements for the 3 layers
Calculate decentrality metrics of the baseline measurements
Choose Decentralized Blockchain System
Gather Data of Blockchain System of interest
Calculate decentrality metrics of Blockchain System
Compare the decentrality metrics of the gathered information with the baseline metrics
Are metrics within an acceptable range?
No
Yes
Proceed to implement Blockchain system in existing system

FIGURE 4. Generic process of decentrality measurement of Blockchain systems by Blockchain analyst.

the Blockchain becomes more centralized. Once, we have the probability distribution of the nodes being successful at mining new blocks, we can calculate the decentrality using below mentioned metrics:

1) FAIRNESS

Fairness metrics have been used extensively in resource allocation in wireless networks. They measure the fairness level of resource decisions in allocations [45]. As the objective of a consensus protocol in Blockchain is to be fair among the miners, we can use the Fairness index to quantify decentrality, as

$$F(X) = \frac{(\sum_{i=1}^{i=N} p_i)^2}{N \sum_{i=1}^{i=N} p_i^2}, \tag{1}$$

where p_i is the fraction of total blocks mined by a node i and where N is the number of miners. When a system is completely distributed, when all p_is are the same, the fairness is 1. When it is completely central, the fairness will be $\frac{1}{N}$.

We can also define decentrality as a normalized fairness, i.e.,

$$NF(X) = \frac{F(x) - \frac{1}{N}}{1 - \frac{1}{N}} \tag{2}$$

When a system is completely distributed, the normalized fairness is 1. When it is completely central, the normalized fairness will be 0.

2) ENTROPY

Entropy has been employed in various fields to quantify uncertainty or randomness of an event or mechanism [46]. If we consider the Blockchain system as an information-source, we can model it as a random variable. Here, the amount of information emanating from a source is the amount of uncertainty that existed before the source released the information. In Blockchain systems, we can estimate the probability that a miner will create the next block based on its ability to add a block in the past. With respect to this model, we can use Shannon's entropy [47], $H(x)$, to quantify decentrality as,

$$H(X) = \sum_{i=1}^{i=N} -p_i log(p_i), \tag{3}$$

We can also define decentrality as a normalized entropy [48], [49], i.e.,

$$d(X) = \frac{H(X)}{log_2(N)}, \tag{4}$$

where $log_2(N)$ is the maximum entropy of the system.

We can define decentrality in terms of min-entropy [50], [51] as,

$$H_\infty = -log(max(p_i)). \tag{5}$$

Here, Shannon's entropy and Min-entropy are different instances of Renyi-entropy of order q [52], i.e.,

$$H = \frac{1}{1-q} ln(\sum_{i=1}^{N}(p_i^q)). \tag{6}$$

3) GINI COEFFICIENT

The Gini coefficient aims at measuring the degree of inequality in a distribution [53], [54]. It is most often used in economics to measure how far a country's wealth or income distribution deviates from a totally equal distribution.Gini coefficient is defined as,

$$G = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} |p_i - p_j|}{2 * N \sum_{j=1}^{N} p_j} \tag{7}$$

where p_i is the fraction total wealth or income of the i_{th} individual. In the most equal society, every person receives the same income ($G = 0$); and in the most unequal society, a single person receives 100 percent of the total income and the remaining $N - 1$ people receive none ($G = 1 - 1/N$).

4) DISTANCE MEASURES

We can use Euclidean distance to compare the resource distribution with the best case scenario [55], [56]. Here, Euclidean distance is defined as:

$$Euclidean\_dist = \sqrt{\sum_{i=1}^{N}(p_i - s_i)^2} \tag{8}$$

where $s_i = \frac{1}{N}$ (the best case scenario). Similarly, we can use Minkowski Distance, which is a generalization of Euclidean Distance. Minkowski Distance [57], [58] with parameter r is defined as:

$$Minkowski\_dist = (\sum_{i=1}^{N}(\mid p_i - s_i \mid)^r)^{(\frac{1}{r})} \tag{9}$$

When $r = 1$, Minkowski distance is considered as City block (Manhattan, taxicab, L_1 norm) distance. A common example of this is the Hamming distance, which is just the number of bits that are different between two binary vectors. When $r = 2$, Minkowski distance becomes Euclidean distance. Similarly, when r → ∞ "supremum" (L_{max} norm, L_∞ norm) distance. This is the maximum difference between any component of the vectors.

5) SIMILARITY MEASURE

We can also use Cosine Similarity to define decentrality [59], [60]. Cosine Similarity is defined as:

$$cos(P, S) = \frac{\sum_{i=1}^{N}(p_i * s_i)}{\sum_{i=1}^{N} p_i * \sum_{i=1}^{N} s_i} \tag{10}$$

We can also use Kullback–Leibler divergence (KL divergence) to quantify decentrality [61], [62]. KL divergence is used to compare two distributions. It is defined as

$$D_{KL}(P||S) = \sum_{i=1}^{N} p_i * log(\frac{p_i}{s_i}). \tag{11}$$

C. NETWORK LAYER

The main concern with centralization is trust. If there are only a few entities capable of running validating nodes or full nodes, then those entities could conspire to allow invalid transactions for their own gain, and there will be no way for other users to know without processing the block for themselves. Fairness among the nodes suffers when large miners have an advantage over small miners by controlling the network topology and bandwidth. This anomaly leads to centralization, where the mining power tends to be concentrated under a single controller, breaking the basic premise of the decentralized crypto currency vision [19], [63].

1) NETWORK ANALYSIS

We can model the block generation and dissemination among other nodes using stochastic processes. First, we model the time period $T(p_i)$ for a miner i with hardware capable of performing p_i operations per second to find a valid block is distributed exponentially with rate p_i in an interval 0 to t [64],

$$Pr[T(p_i) < t] = 1 - exp(-p_i * t) \tag{12}$$

Consider n miners with hash power rate $p_1, p_2, \ldots, p_n$. the time to find a block in time T is equal to the minimum value of random variable $T(p_i)$ assuming that the miner publishes a found block and it reaches other miners immediately. According to the properties of the exponential distribution, T is also distributed exponentially [64]:

$$Pr[T = min(T_1, T_2, \ldots, T_n) \leq t] = 1 - exp(-t\sum_{i=1}^{n} p_i) \tag{13}$$

Now let's consider another random variable X, which denotes the number of blocks mined within a time interval (t_1, t_2) and is independent of other non-overlapping intervals. The number of blocks mined within time interval t follows a Poisson distribution as (for complete analysis kindly see [65]):

$$Pr[X = k] = \frac{exp(-\sum_{i=1}^{n} p_i * t)(\sum_{i=1}^{n} p_i * t)^{-k}}{k!} \tag{14}$$

2) EFFECT OF NETWORK TOPOLOGY

From equations 12,13 and 14, we can see that the time taken by a node to generate a block and the delay induced due to the data transmission play an important role in Blockchain systems. Hence, the centrality of the nodes should be studied from a networking perspective also.

Network topology analysis will help us understand the ability of various nodes to propagate transactions and blocks.

Centrality in a network measures the importance, influence, or power of a node in the network and is widely applied in social network analysis. Important metrics to calculate centrality of a network are: Degree centrality, Betweenness centrality, and Closeness centrality [16], [66], [67]. Betweenness and closeness centrality count only geodesic paths assuming that messages or transactions in a network flow only along the shortest possible paths. The Eigenvector measure counts walks, which assumes that trajectories can also revisit nodes and edges multiple times (see [68] for more network analysis related metrics).

a: DEGREE CENTRALITY

The degree centrality is a measure of the number of direct links, or connections that one node i has with others, which captures a local property [16]. To compare the degree centrality among networks of different size, one has to normalize, dividing the measure by the maximum possible number of adjacent connections $N - 1$. Nodes with higher degree centrality tend to have more influence on others.

$$C_D(n_i) = \frac{\sum_{i=1}^{n} a_{ij}}{N - 1} \tag{15}$$

where a_{ij} is the direct link between node i and node j, i.e., $a_{ij} = 1$, if i and j are adjacent nodes, otherwise $a_{ij} = 0$

b: BETWEENNESS CENTRALITY

The betweenness centrality is based on the number of shortest paths passing through a node, which captures global property.

Nodes with high betweenness play a central role in connecting different groups in a network [16]. Nodes with the highest betweenness centrality measure result in the largest increase in a typical distance between others when they are removed.

$$C_B(n_i) = \frac{\sum_{j=1}^{N} \sum_{k>j}^{N} \frac{g_{jk}(i)}{g_{jk}}}{\frac{1}{2}N(N-1)}, \tag{16}$$

where $g_{jk}(i)$ is all geodesics linking node j and node k which pass through node i, and g_{jk} is the geodesic distance between nodes j and k.

c: CLOSENESS CENTRALITY

The closeness centrality emphasizes the distance of a node to all other nodes in the network by focusing on the geodesic distance from each node to all others [16]. Closeness centrality can be regarded as a measure of how long it will take information to spread from a given node to others in the network. The most central node in the network is that with the minimum costs or time for communicating with all others.

$$C_c(n_i) = \sum_{j=1, j\neq i}^{N} \frac{N-1}{d(n_i, n_j)}, \tag{17}$$

where $C_c(n_i)$ is the closeness centrality and calculated by the sum of inverse distances $d(n_i, n_j)$ between two nodes in the network.

D. STORAGE LAYER

Location of a node running a Blockchain system plays an important role while defining decentrality. A permissionless Blockchain enables any node to join the network, keeping it as decentralized as possible without knowledge of the hardware the nodes are running on [63]. This problem does not really exist in permissioned Blockchains since the endorsers that execute the contracts are known. They rely on the fact that everyone has a membership and can therefore be identified. If an application misuses the resources, appropriate action will be taken on the developers.

Three roles exist on Blockchain systems that interact with the storage layer. The Miner role or transactor role is one in which the software and hardware participate in the consensus process. This role proposes transactions and participates in voting, mining, staking, and any other tasks required by the protocol of a participating peer. The most commonly used role would be on a client or a user that interacts with the Blockchain system. This role often comes from a wallet application that stores credentials or cryptographic content required for the user to communicate and interact with the Blockchain system. The last role is not much talked about, but is potentially the most important. This is the role of the auditor, and involves software and hardware that downloads and verifies the ledger of the Blockchain system to ensure that all actors participating are following the rules. It is important that the auditor role be operational and run by independent parties to the system to ensure that expectations of the system are met.

All of these roles could be run on a variety of storage platforms including physical hardware, virtual hardware, and cloud based run systems. In Blockchain, many nodes run on cloud service provider, hence the cloud service provider becomes a crucial point of centrality. For example, if more than half of the miner nodes are being hosted on a single cloud service providers, and the cloud service provider starts to behave maliciously, or is hacked, or coerced by the government, then the Blockchain can be modified or disrupted [69]. This can also occur if the auditing roles are not able to independently audit the system due to the masking of a cloud based service provider. The best form of decentralization is when the users run software locally on their machines providing them full control. A more centralised approach is when all the users run their software on a single cloud service provider [12].

Based on the location of a node running, we can define three types of nodes:

- *Individual nodes*: These are the nodes who are running the instances of the Blockchain on different systems [70], [71].
- *Cloud based nodes*: These are the nodes which are running the Blockchain system at cloud service providers. Here, the nodes can use a particular cloud service provider to run the Blockchain instances or they can use various cloud service providers. Here, cloud based service providers essentially act as a single physical machine in a logical sense. Technically, it is distributed but it is run by a single owner.
- *Wallet-based nodes*: Wallets plays an important role for the centrality of the Blockchain. As users use wallets to store their private keys, any attack on the wallet can make Blockchain unavailable to the users. Hence, the number of wallets used by the users to store their private keys is crucial. Each user can have their own private wallets. However, users find it hard to maintain their wallets and do not want to download the entire Blockchain during client installation. They prefer to use the services of wallet providers, who store wallets online, regardless of how the mining is done. [27], [28]

The Blockchain system can be considered storage-wise decentralized if each and every node runs their code on individual machines, and centralized if all the nodes run their services at a single cloud service provider. The notion of a single machine or single cloud service provider should always be considered as the owners and controllers of the machine. If one owner can manipulate (turn off, re-configure, access data, upgrade) the machines then it doesn't matter if individual organizations or people control them.

In [72], Zhan *et al.*, used the Shannon-Wiener index, which is an evenness indicator, to measure if each disk drive accommodates the same number of file blocks in a distributed storage system. Distribution index, a part of Diversity index, has been used in ecological research [73], which is the quantification of equality of abundance in a community,

TABLE 2. Six cases of synthetic resource distribution from the decentralized (Best case, S1) to centralized (Worst case, S6) for governance layer.

Nodes	S1	S2	S3	S4	S5	S6
	(Best-Case)					(Worst-case)
Node 1	10 (0.1)	17 (0.17)	42 (0.42)	25 (0.25)	50 (0.5)	100 (1)
Node 2	10 (0.1)	17 (0.17)	20 (0.2)	25 (0.25)	50 (0.5)	0 (0)
Node 3	10 (0.1)	13 (0.13)	14 (0.14)	25 (0.25)	0 (0)	0 (0)
node 4	10 (0.1)	12 (0.12)	7 (0.07)	25 (0.25)	0 (0)	0 (0)
Node 5	10 (0.1)	11 (0.11)	4 (0.04)	0 (0)	0 (0)	0 (0)
Node 6	10 (0.1)	9 (0.09)	4 (0.04)	0 (0)	0 (0)	0 (0)
Node 7	10 (0.1)	7 (0.07)	3 (0.03)	0 (0)	0 (0)	0 (0)
Node 8	10 (0.1)	6 (0.06)	3 (0.03)	0 (0)	0 (0)	0 (0)
Node 9	10 (0.1)	5 (0.05)	2 (0.02)	0 (0)	0 (0)	0 (0)
Node 10	10 (0.1)	3 (0.03)	1 (0.01)	0 (0)	0 (0)	0 (0)

i.e., when there are similar proportions of all species within a community, then the distribution index tends to be one; but when the abundances are very dissimilar then the distribution index decreases. For example, suppose, we have n blocks which we want to distribute across c number of disk drives where each disk i stores x_i, then the distribution index (E) is calculated as follows

$$E = \frac{-\sum_{i=1}^{c}(\frac{x_i}{n} * log_2(\frac{x_i}{n}))}{log_2(n)} \tag{18}$$

When the cloud servers evenly distribute all n blocks on exactly c drives, i.e., $x_1 = x_2 = \ldots = x_c = n/c$, then the distribution index will be $\frac{log_2(c)}{log_2(n)}$.

Hence, we can calculate distribution index for individual nodes, cloud servers based nodes and wallet-based notes to quantify centrality. In the context of storage centrality, n represents the total number of blockchain nodes, c represents the total number of servers where data is stored, x_i is the number of nodes using server i. Here, when each node uses its own server, i.e., $c = n$, E becomes 1, and when all the nodes run at a single cloud server, i.e., $c = 1$, E becomes 0.

IV. RESULTS

In this section, we discuss the results for governance layer, network layer and storage layers of Blockchain systems. In each of these layers we compare Bitcoin and Ehtereum network with synthetic configuration.

A. GOVERNANCE LAYER DECENTRALITY

To achieve decentrality in the governance layer the computational power is considered. In a true decentralized system, computational power of all the nodes should be the same, i.e., in the Blockchain context, the hashrate should be the same at all miner nodes. We first consider synthetic networks to set the framework to measure different metrics, evaluate Bitcoin and Ethereum networks, and compare them.

1) SYNTHETIC DATA

Here, we consider six scenarios where 100 blocks are mined by 10 mining nodes. In the best-case scenario, each node mines 10 blocks, thus this scenario signifies a true decentralized system. In the worst-case scenario, only one node mines all the blocks, which represents a centralized system. We also consider four more cases where blocks are mined randomly. Table 2 shows six cases of resource distributions, from the best-case to the worst-case. The columns in the table represent from S1 (best-case) to S6 (worst-case). The rows represent the blocks mined by nodes in different scenarios. For example, in $S1$ column, each nodes has mined 10 blocks, i.e., Node 1 through Node 10 have each minded 10 blocks. In $S5$ column, only *Node* 1 and *Node* 2 have mined 50 blocks each, and others haven't mined any blocks.

Table 3 shows different metric measurements for the six scenarios discussed in Table 2. Here, the columns in the table represent from S1 (best-case) to S6 (worst-case). The rows represent different metrics. The observations are following:

- *Fairness:* The fairness value (see equation 1) decreases from 1 to 0.1 as a system goes from $S1$ to $S6$. That means fairness decreases with centrality.
- *Entropy and normalized entropy:* As we know, the entropy (shown in 3^{rd} row in Table 3) is high when the randomness in the system is high, i.e., in Blockchain system the probability of generating the block is equiprobable (see equation 3). Hence, we can observe the entropy in $S1$ is the highest, i.e., 3.32, among the given scenarios, and it decreases with centralization. We can observe that in the worst-case scenario, scenario $S6$, it is zero. The same can be observed in case of normalized entropy (see equation 4) in the 4^{th} row.
- *Gini Coefficient:* As discussed in equation 7, the Gini coefficient is zero when nodes generate the block equally. We can observe the same in $S1$ (5^{th} row), where the Gini coefficient is zero, and it increases with centrality.
- *Euclidean distance, and Minkowski Distance with $r = 1$ and $r = \infty$:* As we see, Euclidean and Minkowski distances (see equation 8 and 9) are zero for $S1$ (7^{th} and 8^{th} rows) and increase with centrality.
- *Cosine similarity and KL Divergence:* We can see both Cosine similarity and KL Divergence (equations 10 and 11) are zero (column $S1$ and 9^{th} and 10^{th} rows) when ndes generate blocks equally, and increase with centrality.

TABLE 3. Governance layer metric measurements for the six cases of resource distribution mentioned in Table 2.

Metric	S1 (Best-Case)	S2	S3	S4	S5	S6 (Worst-case)
Fairness index	1	0.82	0.40	0.40	0.2	0.10
Normalized Fairness Index	1	0.805	0.339	0.333	0.111	0
Entropy	3.32	3.16	2.51	2.0	1.0	0
Normalized Entropy	1	0.95	.75	.60	.30	0
Gini coefficient	0	0.262	0.56	0.6	0.8	0.9
Euclidean Distance	0	0.145	0.382	0.387	0.632	0.948
Minkowski Distance (r=1)	0	0.4	0.9	1.2	1.6	1.8
Minkowski Distance (r=∞)	0	0.07	0.31	0.15	0.4	0.9
Cosine Similarity	0	0.091	0.362	0.367	0.552	0.683
Kullback–Leibler divergence	0	0.16	0.81	1.32	2.32	3.32

2) BITCOIN DATASET

In Bitcoin, the hash rate of the node is a proxy for the number of blocks mined by a node. We collected Bitcoin hashrate for the following period: the entire chain, last 1 year, last 3 months and last 1 month (Dated April, 2, 2020, from btc.com). Fig. 5 shows the number of pools and blocks mined in Bitcoin. In Fig 5.a, x-axis shows the time and y-axis shows the number of pools. Similarly, in Fig 5.b, x-axis shows the time and y-axis shows the number of blocks mined. In both these figures, we can see that in the short-term only a few nodes are contributing blocks for the Blockchain.

Fig. 6, 7, 8, and 9 show Hashrate distribution of entire Bitcoin, for last 1 year, last 3 months and last 1 month, respectively, where x-axis shows the name of the mining pools and y-axis shows the hashrate. In Fig. 6, it may be observed that 36 percent of the hashrate belongs to unknown miners and the rest is distributed among the other pools. In the decentrality measurements, in this article, we have ignored the unknown pool of miners, as we do not know the exact number of miners, and only considered named pools. Here, in Fig. 7, 8 and 9, we can notice only a few pools, like, BTC.com, F2Pool, have more hashrate shares than others.

Table 4 shows different metric measurements of centrality of the Bitcoin network. Here, the columns in the table represent the four duration of data collection, i.e., the entire chain, the last 1 year, last 3 months and last 1 month, respectively. The rows represent different metrics. The following observations may be made from Tables 3 and 4.

- *Fairness:* The fairness index (see equation 1) of the entire chain is 0.159. When we compare that with Table 3, we can see that the bitcoin network is very centralized. However, when we observe the fairness index for the last 1 year, last 3 months and 1 month, we notice that during this period the Bitcoin network is more centralized among those pools.
- *Entropy and normalized entropy:* In both entropy (see equation 3) and normalized entropy (see equation 4) we see, the entropy decreases with period, which signifies that the network is more centralized among a few pools.
- *Gini Coefficient:* When we compare the Gini coefficient (see equation 7) of Bitcoin with Table 3, we see the entire Bitcoin network is very centralized. And the network is well distributed among a few pools, as we see it over 1 year, 3 months and 1 month.
- *Euclidean distance, and Minkowski Distance with $r = 1$ and $r = \infty$:* Similar to above observation, we can notice all these metrics (see equation 8 and 9) are away from zero, hence they all show the sign of centrality. However, as seen in previous cases, the data over 1 year, 3 months and 1 month show that the network is governed by only a few pools.
- *Cosine and KL Divergence:* We observe that both cosine similarity and KL divergence (equations 10 and 11) show that the Bitcoin network is highly centralized in case of the entire network and over other periods they are governed equally by only a few pools.

3) ETHEREUM DATASET

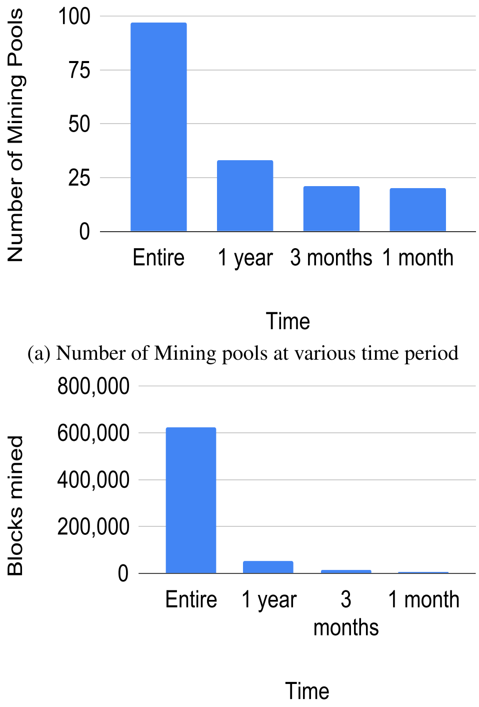
We collected Ethereum hashrate for the following period: the whole chain, the last 1 year, last 3 months and last 1 month (Dated April 07 2020 from eth.btc.com). Fig. 10 shows the number of pools and blocks mined in Ethereum. In Fig 10.a, x-axis shows the time and y-axis shows the number of pools. Similarly, in Fig 10.b, x-axis shows the time and y-axis shows the number of blocks mined. In both these figures, we can see that in the short-term only a few nodes are contributing blocks for the Ethereum.

Fig. 11, 12, 13 and 14 show Hashrate distribution of Complete Ethereum, for 1 year, 3 months and 1 month, respectively. Here, x-axis shows the name of the mining pools and y-axis shows the hashrate. In Fig. 11, it may be observed that most of the hashrate belongs to Ethermine miners, while, in Fig. 12, 13 and 14, we can notice only a few pools, like SparkPool, SparkPool_3, have more hashrate shares than others. Like Bitcoin, here also we can see that in short-term only a few nodes are contributing for the Blockchain.

Table 5 shows different metric measurements of Ethereum. Here, the columns in the table represent the four duration of data collection, i.e., the entire chain, the last 1 year, last 3 months and last 1 month. The rows represent different metrics. The observations are as follows:

TABLE 4. Governance layer metric measurements of Bitcoin for entire time period, last one year, last three months and last one month.

Metric	Entire	1 year	3 months	1 month
Fairness Index	0.159	0.104	0.097	0.095
Normalized Fairness Index	0.150	0.094	0.088	0.085
Entropy	4.514	3.661	3.567	3.537
Normalized Entropy	0.685	0.556	0.541	0.537
Gini Coefficient	0.821	0.901	0.906	0.908
Euclidean Distance	0.234	0.299	0.309	0.314
Minkowski Distance (r=1)	1.380	1.634	1.665	1.660
Minkowski Distance (r=∞)	0.124	0.147	0.172	0.176
Cosine Similarity	0.601	0.677	0.687	0.691
Kullback–Leibler Divergence	2.070	2.923	3.017	3.047



(a) Number of Mining pools at various time period

(b) Number of Block mined at various time period

FIGURE 5. Number of mining pools and blocks mined in Bitcoin for entire time period, 1 year, 3 months and 1 month (collected from https://btc.com/stats/pool?pool_mode=month dated April 2 2020).

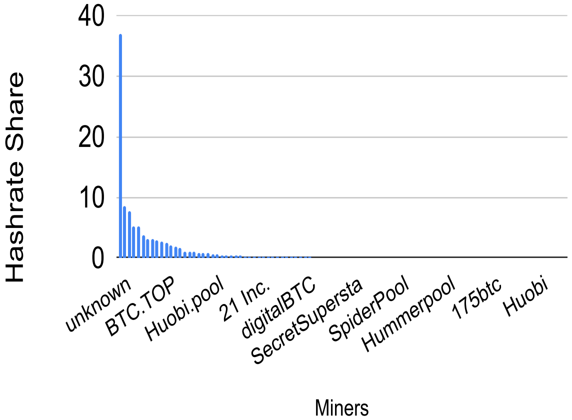


FIGURE 6. Hashrate distribution of entire Bitcoin collected from https://btc.com/stats/pool?pool_mode=all dated April 2 2020.

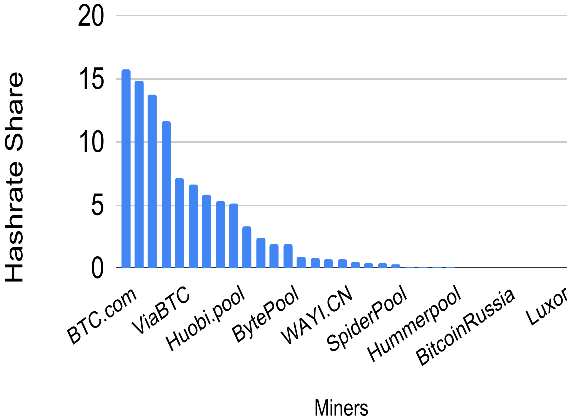


FIGURE 7. Hashrate distribution of Bitcoin for the last 1 year collected from https://btc.com/stats/pool?pool_mode=year dated April 2 2020.

- *Fairness:* The fairness index (see equation 1) of the entire chain is 0.159. When we compare that with Table 3, we can see the Ethereum network is very centralized. However, when we observe the fairness index for the last 1 year, last 3 months and 1 month, we notice that during this period the Bitcoin network is more centralized among those pools, hence 0.104 0.097, and 0.095. We can see this difference compared to Bitcoin, because of the number of pools involved here, i.e., in Bitcoin the number of pools is around 100 while in Ethereum it is around 6000.
- *Entropy and normalized entropy:* In both entropy and normalized entropy (see equation 3 and equation 4) we see, the entropy decreases with period, i.e., 4.514 for the entire network and 3.661, 3.567, 3.537 for 1 year, 3 months and 1 month. This signifies that the network is more centralized among a few pools over a short-period of time.
- *Gini Coefficient:* When we compare the Gini coefficient (see equation 7) of Ethereum with Table 3, we see the entire network is very centralized, i.e., Gini coefficient is 0.821. And a few pools control the network over 1 year, 3 months and 1 month.

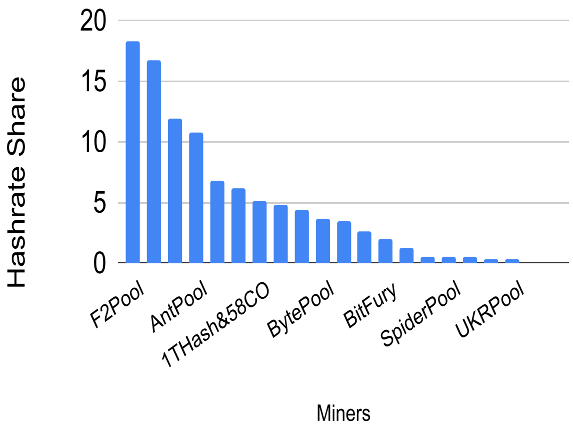


FIGURE 8. Hashrate of Bitcoin for the last 3 month collected from https://btc.com/stats/pool?pool_mode=month3 dated April 2 2020.

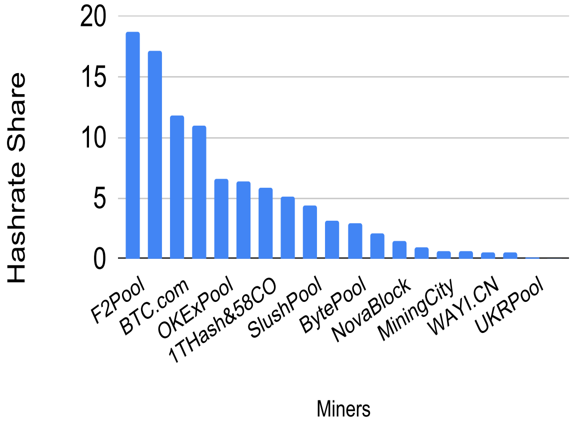
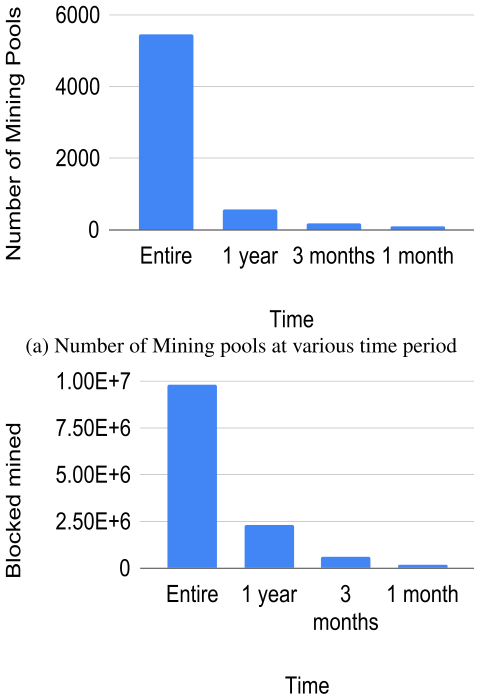


FIGURE 9. Hashrate of Bitcoin for the last 1 month collected from https://btc.com/stats/pool?pool_mode=month dated April 2 2020.

- *Euclidean distance, and Minkowski Distance with $r = 1$ and $r = \infty$:* Here, all these metrics (equations 10 and 11) are greater than zero, hence they all show the sign of centrality. However, as seen in previous cases, the data over 1 year, 3 months and 1 month show that the network is governed by only a few pools.
- *Cosine and KL Divergence:* Both cosine similarity and KL divergence show (equations 10 and 11) that the Ethereum network is highly centralized (i. e., cosine similarity and KL divergence are 0.601 and 2.070, respectively) in case of entire network and over other periods they are governed equally by only a few pools.

As we can see, both in Bitcoin and Ethereum networks, they do show some sign of centrality. As shown in Table 7, the number of pools combined with more than 51 percent of total hash rate is significantly low compared to the to the total number of pools (See Fig 5 and 10). Hence, only few polls are controlling both Bitcoin and Ethereum networks. In order to further find the extend of control, we can compare the measurements of various metrics over a particular time period, i.e., 1 month (see last column of Table 6 and 5) with the six scenarios mentioned in Table 3. Although, there are 10 metrics on the table, we can only consider normalized fairness index, normalized entropy, Gini coefficient, and cosine similarity, as these measurements are independent of the network size and their range is from zero to one. Here, we can notice that these metrics for both Bitcoin and Ethereum is similar to $S3$. This signifies that although there are many miners, only a few control the whole network.



(a) Number of Mining pools at various time period

(b) Number of Block mined at various time period

FIGURE 10. Number of mining pools and blocks mined in Ethereum for entire time period, 1 year, 3 months and 1 month (collected from https://eth.btc.com/miningstats dated April 07 2020).

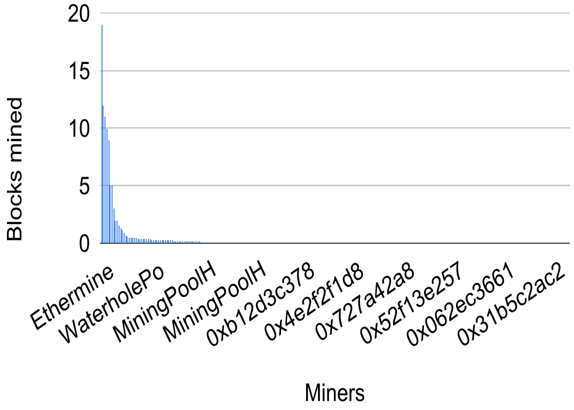


FIGURE 11. Hashrate distribution of entire Ethereum collected from https://eth.btc.com/miningstats dated April 07 2020.

B. NETWORK LAYER DECENTRALITY

In this case, the data are synthetic, as there is no data about the network topology of the Bitcoin or Ethereum network. For

TABLE 5. Governance layer metric measurements of Ethereum for entire time period, last one year, last three months and last one month.

Metric	Entire	1 year	3 months	1 month
Fairness Index	0.041	0.022	0.021	0.020
Normalized Fairness Index	0.038	0.019	0.017	0.017
Entropy	4.548	3.526	3.442	3.405
Normalized Entropy	0.562	0.435	0.425	0.421
Gini Coefficient	0.904	0.960	0.964	0.965
Euclidean Distance	0.289	0.397	0.413	0.416
Minkowski Distance (r=1)	1.569	1.730	1.744	1.753
Minkowski Distance (r=∞)	0.185	0.278	0.319	0.325
Cosine Similarity	0.795	0.849	0.854	0.856
Kullback–Leibler divergence	3.539	4.561	4.644	4.682

TABLE 6. Network layer metric measurements, i.e., Degree centrality (DC), Betweenness centrality (BC) and Closeness centrality (CC), of Complete graph, Path graph, Star topology and Random topology.

Nodes	Complete graph			Path graph			Star graph			Random graph		
	DC	BC	CC	DC	BC	CC	DC	BC	CC	DC	BC	CC
1	1.0	0	1.0	0.111	0.0	1.0	1.0	1.0	1.0	0.111	0.0	0.281
2	1.0	0	1.0	0.222	0.222	0.529	0.111	0	0.529	0.333	0.638	0.45
3	1.0	0	1.0	0.222	0.388	0.529	0.111	0	0.529	0.111	0.0	0.321
4	1.0	0	1.0	0.222	0.5	0.529	0.111	0	0.529	0.111	0.0	0.264
5	1.0	0	1.0	0.222	0.555	0.529	0.111	0	0.529	0.333	0.416	0.375
6	1.0	0	1.0	0.222	0.555	0.529	0.111	0	0.529	0.111	0.0	0.264
7	1.0	0	1.0	0.222	0.5	0.529	0.111	0	0.529	0.222	0.5	0.409
8	1.0	0	1.0	0.222	0.388	0.529	0.111	0	0.529	0.333	0.416	0.346
9	1.0	0	1.0	0.222	0.222	0.529	0.111	0	0.529	0.111	0.0	0.281
10	1.0	0	1.0	0.111	0	0.529	0.111	0	0.529	0.222	0.555	0.45

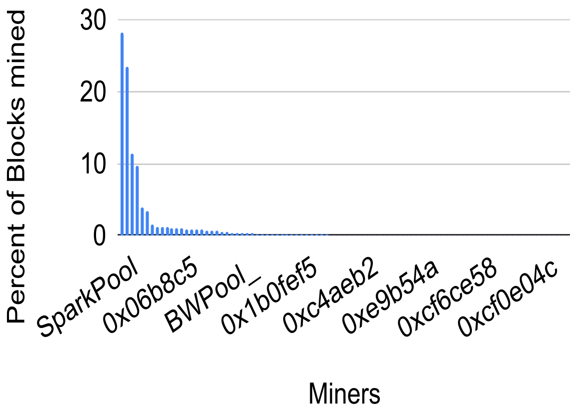


FIGURE 12. Hashrate distribution of Ethereum for the last year collected from https://eth.btc.com/miningstats dated April 07 2020.

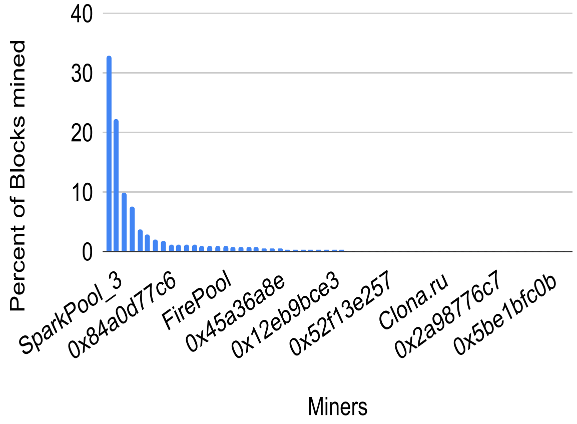


FIGURE 13. Hashrate distribution of Ethereum for the last 3 months collected from https://eth.btc.com/miningstats dated April 07 2020.

the purpose of pictorial illustration, we have only considered 10 nodes. However, the similar output can be observed with a higher number of nodes. Here, we discuss the results from the best case scenario to the worst case scenario. Fig. 15 shows topology of various cases. Table 6 shows degree (DC), betweenness (BC) and closeness centrality (CC) (see equations 15, 16 and 17) of complete graph, path graph, star topology and random topology. The columns represent DC, BC and CC of complete graph, path graph, star topology and random topology. The rows in the table show the DC, BC and CC nodes,i.e., from *node* 1 to *node* 10. Fig. 16, 17, 18, and 19 show DC, BC and CC measurements of the complete graph, path graph, star topology and random graph. Here, x-axis show the nodes, and y-axis represent DC, BC and CC.

1) COMPLETE GRAPH (BEST CASE)

In the best case scenario, all nodes are connected with others directly, forming a complete connected graph. In this case, starting as a new node is the most difficult. The node having

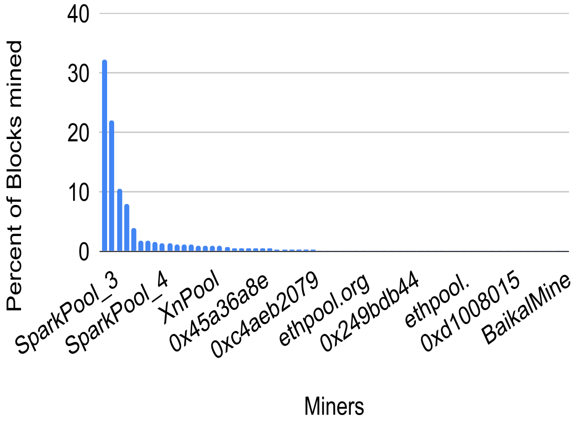


FIGURE 14. Hashrate distribution of Ethereum for the last month collected from https://eth.btc.com/miningstats dated April 07 2020.

TABLE 7. Number of pools combined having more than fifty-one percent of hash rate.

Duration of Mining	Bitcoin Network	Ethereum Network
Entire	15	4
Last one year	4	2
Last three months	4	2
Last one month	4	2

poor network connectivity will have issue becoming member. In this topology, each node has high degree and closeness centrality and low betweenness centrality. Fig. 16 shows measurements of the complete graph.

Both DC and CC of all nodes in the complete graph is 1 (see Table 6 as well as Fig. 16.a and 16.c). It signifies that when all the nodes are connected to each other directly, they relay the information quickly to others, hence, there is no centrality. Similarly, BC value *zero* signifies no centrality, which can be seen in Fig. 16.b (as the values are zero, the figure seems empty).

2) PATH GRAPH

In this case, all nodes, except the last nodes, are connected to only two nodes directly, forming a long chain. Here, starting as a new node is easier, i.e., it has to connect only two nodes. In this topology, every node, but the end nodes, has the highest degree, betweenness and closeness centrality. Fig. 17 shows measurements of the path graph.

The nodes at the center of the path graph play important roles for spreading the information. Hence, BC and CC of those nodes will be the highest among other nodes (see Table 6 as well as Fig. 17.b and 17.c). We see centrality arising in the path graph. Similarly, value *zero* signifies less centrality at the end of the path graph.

3) STAR TOPOLOGY (WORST CASE)

In this worst-case, all nodes will be connected to only a centralized directly forming a star topology. Here, if the central node is down the whole network is down. It is the most centralized system. Hence, the central node has the highest degree, betweenness and closeness centrality. Here, starting as a new node is very easy as a node can directly connect with the central node. Fig. 18 shows measurements of the star topology. Hence, DC, BC and CC of the central node is the highest among other nodes (see Table 6 as well as Fig. 18.a, 18.b and 18.c). Hence, we observe centrality in star topology.

4) RANDOM TOPOLOGY

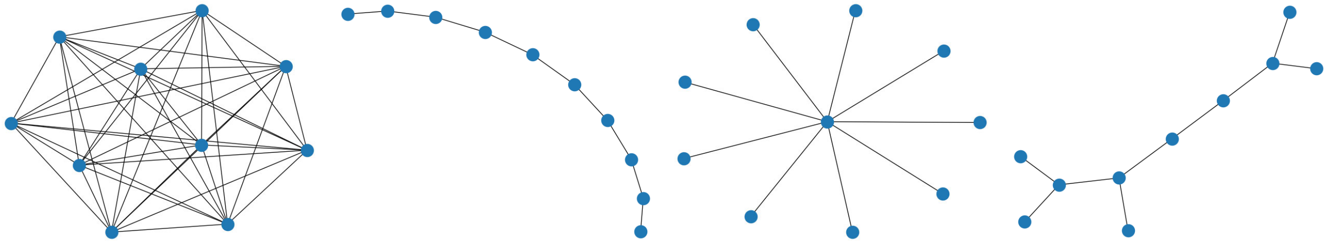
In random topology, different nodes will have different degree of connectivity with each other, which is normally found in real life. In this scenario, a few nodes emerge as contributing more, hence, the centrality emerges. Here, DC, BC and CC of the most connected nodes are the highest among other nodes (see Table 6 as well as Fig. 19.a, 19.b and 19.c). Bitcoin and Ethereum would have more resemblance to the random topology. Fig. 19 shows measurements of the Random graph. As we can see here, *node* 1, *node* 4 and *node* 7 show highest degree centrality; *node* 1 shows highest betweenness centrality; and *node* 1 and *node* 9 show highest closeness centrality. Hence, *node* 1, *node* 4 and *node* 7 are the important nodes where centrality arises.

C. STORAGE LAYER DECENTRALITY

To look at decentrality in the storage layer, we consider the storage used by nodes/users for local Blockchains and wallets. In a Blockchain system, a fully storage layer decentralization is achieved when all nodes run their instances of Blockchain on their own local physical machines. On the other hand, a Blockchain system is fully storage layer centralized when all the nodes run their Blockchain on a single server, which could be at a cloud service provider.

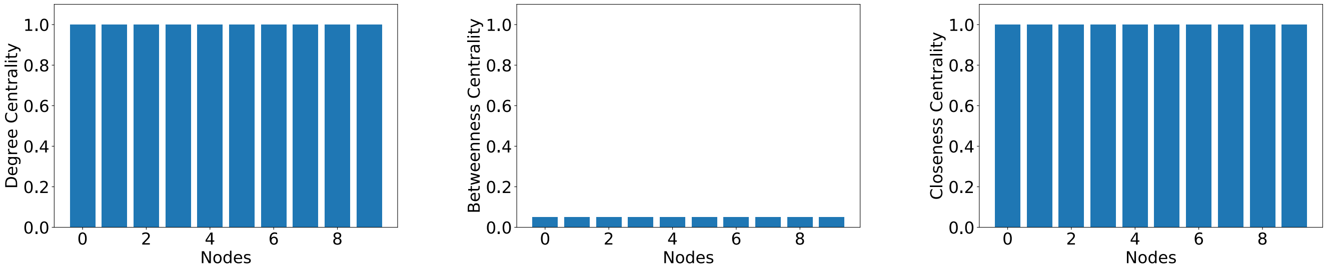
Similarly, storage layer decentrality could be defined for wallets. Generally, users use wallets to store their private keys. For example, if a node wants to offload their Blockchain to some cloud service provider, they can store the public and private keys locally or in a wallet. Hence, any attack on the wallet service provider can result in an attack on the Blockchain itself. In the context of storage layer decentrality, a system where each node stores its public and private keys locally (say in a wallet) is desirable to achieve full decentralization.On the other hand, when all nodes employ a single wallet provider, it contributes toward centrality at the storage layer.

Both decentrality of storage of the Blockchain and wallet can be quantified by using distribution index (See equation 18). Here, distribution index of 0 means complete centrality, and distribution index of 1 means complete decentrality. To illustrate the relevance of this equation in measuring storage decentrality, we consider a Blockchain system with 100 nodes. Depending on where each node stores its local Blockchain copy, we get a different distribution index. We chose a 100 node Blockchain system because we observed 97 minors in the Bitcoin network (see Fig. 5). We analyze the impact of the number of servers on a distri-



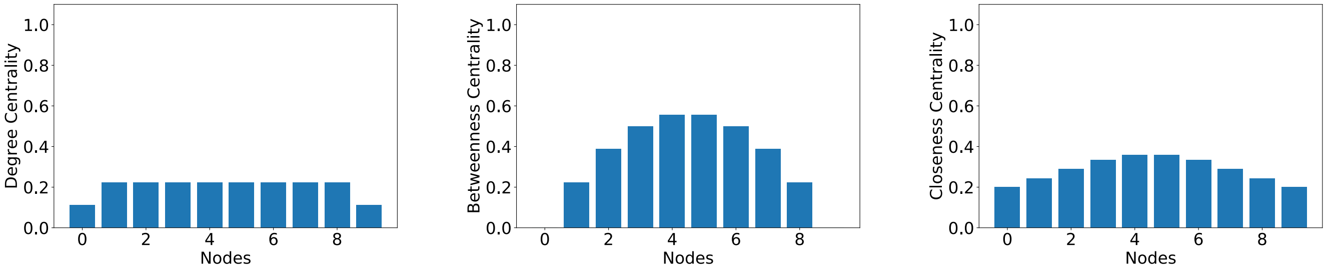
(a) Complete Graph (b) Path Graph (c) Star Topology (d) Random Topology

FIGURE 15. Topology of various cases.



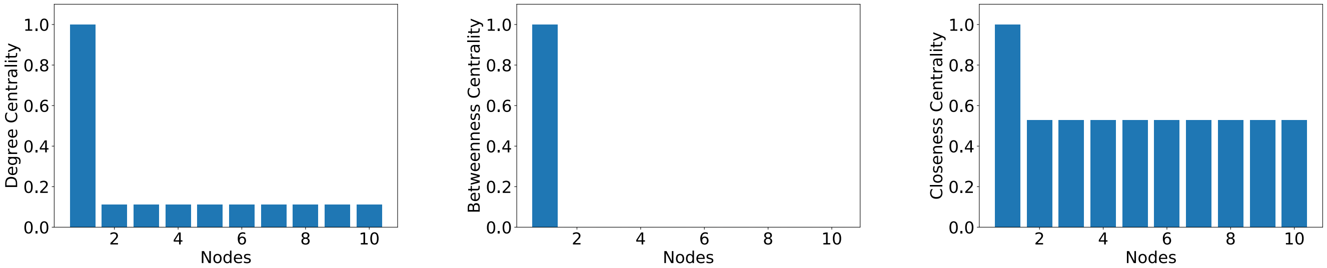
(a) Degree centrality (b) Betweenness Centrality (c) Closeness Centrality

FIGURE 16. Measurements of Complete Graph.



(a) Degree centrality (b) Betweenness Centrality (c) Closeness Centrality

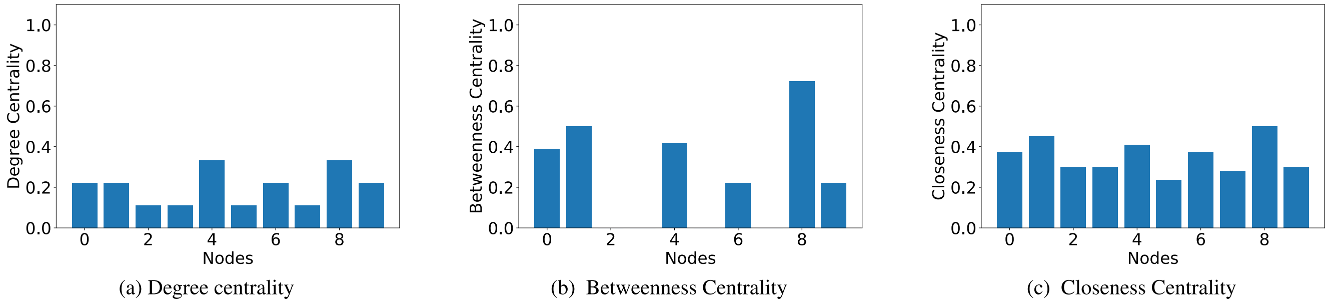
FIGURE 17. Measurements of Path Graph.



(a) Degree centrality (b) Betweenness Centrality (c) Closeness Centrality

FIGURE 18. Measurements of Star topology.

bution index. Fig 20 provides the result of distribution index versus number of servers on which the storage is distributed. Here, X-axis represents the number of storage servers and Y-axis shows the distribution index. For simplicity, here, we assume that the storage is uniformly distributed across the storage servers. With a single storage server($c = 1$) used by all 100 nodes, the distribution index is zero. The distribution index increases non-linearly with the number of storage servers (c). The distribution index is maximum when each of the 100 Blockchain nodes use a separate storage server ($c =$



(a) Degree centrality

(b) Betweenness Centrality

(c) Closeness Centrality

FIGURE 19. Measurements of Random Graph.

100). When the n node storage is not uniformly distributed across the c servers, the distribution will be different.

In summary, we observed that decentrality can be measured at the governance layer, network layer and storage layer using various metrics. For the governance layer, we considered synthetic data, Bitcoin network data and Ethereum network data; and showed decentrality in terms of fairness, entropy, Gini coefficient, KL divergence, etc. Similarly, in the network layer, we measured decentrality by using degree centrality, betweenness centrality and closeness centrality. At the storage layer, we applied a distribution index to define centrality. In summary, to achieve decentrality in a Blockchain system, one needs to achieve decentrality at each of the three layers. For example, a full distribution at governance layer, with a partial decentralization at network layer, and centralization at storage level will result in a system that is close to a centralized system.

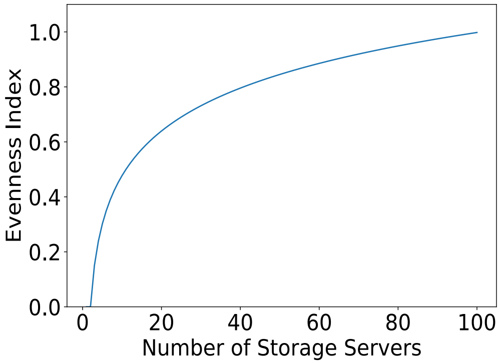


FIGURE 20. Change of Evenness Index with number of storage servers.

V. CONCLUSION AND FUTURE WORK

In this article, we discussed various metrics to quantify decentrality in Blockchain using information theoretic approach. In particular, we looked into the decentralization problem by focusing on three different layers, namely, the governance layer, network layer and storage layer. We discussed different metrics to evaluate decentrality in these layers. Subsequently, we evaluated the decentrality in Bitcoin and Ethereum networks and shared our observations. We noticed, with time, decentralised systems tend to be governed by a few nodes, hence they become more centralized.

Although we covered most of the metrics, there are a few more parameters which can introduce centrality which need worth attention. Those are:

- The team members involved in protocol design and upgrade.
- Company building Mining Hardware, i.e., the dependency of Blockchain system on Hardware. For example, PoW is hardware dependent, in which the success rate of becoming a leader or generation of the block is directly proportional to the hardware size.

Another interesting topic of interest would be to explore the factors that drive centrality. For example, the effect of market incentives, computing properties, demand for smoother user experience, which steer decentralized protocols into centralization need to be investigated. It appears the community discusses centrality as trade off in performance. To make these systems better performing aspects of centrality is added to them. Is the value of decentrality enough to justify the costs? This type of question can only be addressed once the measure of centrality/decentrality in the system can be compared. We need to find how to use the metrics proposed in the paper to quantify the centrality of the whole system.

REFERENCES

[1] K. Gai, Y. Wu, L. Zhu, L. Xu, and Y. Zhang, "Permissioned blockchain and edge computing empowered privacy-preserving smart grid networks," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7992–8004, Oct. 2019.

[2] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.

[3] A. S. Musleh, G. Yao, and S. M. Muyeen, "Blockchain applications in smart grid–review and frameworks," *IEEE Access*, vol. 7, pp. 86746–86757, 2019.

[4] S. Kim, Y. Kwon, and S. Cho, "A survey of scalability solutions on blockchain," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 1204–1207.

[5] W. Yang, E. Aghasian, S. Garg, D. Herbert, L. Disiuta, and B. Kang, "A survey on blockchain-based Internet service architecture: Requirements, challenges, trends, and future," *IEEE Access*, vol. 7, pp. 75845–75872, 2019.

[6] P. Baran, "On distributed communications networks," *IEEE Trans. Commun. Syst.*, vol. 12, no. 1, pp. 1–9, Mar. 1964.

[7] V. Buterin, "The meaning of decentralization," *Medium*, 2017.

[8] K. Wu, B. Peng, H. Xie, and Z. Huang, "An information entropy method to quantify the degrees of decentralization for blockchain systems," in *Proc. IEEE 9th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC)*, Jul. 2019, pp. 1–6.

[9] S. Talukdar, "The equitable distribution of the benefits from decentralization: A challenge for power system designers," in *Proc. IEEE Power Eng. Soc. Summer Meeting*, vol. 3, Jul. 2002, pp. 1691–1692.

[10] M. Conoscenti, A. Vetrò, and J. C. De Martin, "Blockchain for the Internet of Things: A systematic literature review," in *Proc. IEEE/ACS 13th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2016, pp. 1–6.

[11] M. Conoscenti, A. Vetrò, and J. C. De Martin, "Peer to peer for privacy and decentralization in the Internet of Things," in *Proc. IEEE/ACM 39th Int. Conf. Softw. Eng. Companion (ICSE-C)*, May 2017, pp. 288–290.

[12] C. Troncoso, M. Isaakidis, G. Danezis, and H. Halpin, "Systematizing decentralization and privacy: Lessons from 15 years of research and deployments," in *Proc. Privacy Enhancing Technol.*, Oct. 2017, vol. 2017, no. 4, pp. 404–426.

[13] T. Ali Syed, A. Alzahrani, S. Jan, M. S. Siddiqui, A. Nadeem, and T. Alghamdi, "A comparative analysis of blockchain architecture and its applications: Problems and recommendations," *IEEE Access*, vol. 7, pp. 176838–176869, 2019.

[14] W. Wang, D. T. Hoang, P. Hu, Z. Xiong, D. Niyato, P. Wang, Y. Wen, and D. I. Kim, "A survey on consensus mechanisms and mining strategy management in blockchain networks," *IEEE Access*, vol. 7, pp. 22328–22370, 2019.

[15] A. A. Mazlan, S. M. Daud, S. M. Sam, H. Abas, S. Z. A. Rasid, and M. F. Yusof, "Scalability challenges in healthcare blockchain system—A systematic review," *IEEE Access*, vol. 8, pp. 23663–23673, 2020.

[16] M. Lischke and B. Fabian, "Analyzing the bitcoin network: The first four years," *Future Internet*, vol. 8, no. 4, p. 7, Mar. 2016.

[17] L. M. Bach, B. Mihaljevic, and M. Zagar, "Comparative analysis of blockchain consensus algorithms," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 1545–1550.

[18] A. Porat, A. Pratap, P. Shah, and V. Adkar, "Blockchain consensus: An analysis of proof-of-work and its applications," Stanford Univ., Stanford, CA, USA, Tech. Rep.

[19] L. Ren, "Analysis of nakamoto consensus," Cryptology ePrint Arch., Tech. Rep. 2019/943, 2019. [Online]. Available: https://eprint.iacr.org …and https://arxiv.org/abs/1910.08510

[20] M. Wazid, A. K. Das, S. Shetty, and M. Jo, "A tutorial and future research for building a blockchain-based secure communication scheme for Internet of intelligent things," *IEEE Access*, vol. 8, pp. 88700–88716, 2020.

[21] J. Mern, "Structure and evolution of bitcoin transaction network," Dept. Aeronaut. Astronaut., Stanford Univ., Stanford, CA, USA. Tech. Rep.

[22] K. Singh, "Measuring node centrality in lightning network," *Medium*, 2017.

[23] K. Wüst and A. Gervais, "Do you need a blockchain?" in *Proc. Crypto Valley Conf. Blockchain Technol. (CVCBT)*, Jun. 2018, pp. 45–54.

[24] A. A. Monrat, O. Schelén, and K. Andersson, "A survey of blockchain from the perspectives of applications, challenges, and opportunities," *IEEE Access*, vol. 7, pp. 117134–117151, 2019.

[25] Q. Zhou, H. Huang, Z. Zheng, and J. Bian, "Solutions to scalability of blockchain: A survey," *IEEE Access*, vol. 8, pp. 16440–16455, 2020.

[26] T. Sato and Y. Himura, "Smart-contract based system operations for permissioned blockchain," in *Proc. 9th IFIP Int. Conf. New Technol., Mobility Secur. (NTMS)*, Feb. 2018, pp. 1–6.

[27] L. J. Valdivia, C. Del-Valle-Soto, J. Rodriguez, and M. Alcaraz, "Decentralization: The failed promise of cryptocurrencies," *IT Prof.*, vol. 21, no. 2, pp. 33–40, Mar. 2019.

[28] A. Gervais, G. O. Karame, V. Capkun, and S. Capkun, "Is bitcoin a decentralized currency?" *IEEE Secur. Privacy*, vol. 12, no. 3, pp. 54–60, May 2014.

[29] S. Chu and S. Wang, "The curses of blockchain decentralization," 2018, *arXiv:1810.02937*. [Online]. Available: http://arxiv.org/abs/1810.02937

[30] B. S. Srinivasan, "Quantifying decentralization," *Medium*, 2017. [Online]. Available: https://news.earn.com/quantifying-decentralization-e39db233c28e

[31] K. Croman, C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, E. Gün Sirer, D. Song, and R. Wattenhofer, "On scaling decentralized blockchains," in *Proc. Financial Cryptogr. Data Secur.*, 2016, pp. 106–125.

[32] H. Zhuang, R. Rahman, and K. Aberer, "Decentralizing the cloud: How can small data centers cooperate?" in *Proc. 14th IEEE Int. Conf. Peer-Peer Comput.*, Sep. 2014, pp. 1–10.

[33] R. K. Raman and L. R. Varshney, "Distributed storage meets secret sharing on the blockchain," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2018, pp. 1–6.

[34] T. Neudecker and H. Hartenstein, "Network layer aspects of permissionless blockchains," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 838–857, 1st Quart., 2019.

[35] C. Cachin and M. Vukolić, "Blockchain consensus protocols in the wild," 2017, *arXiv:1707.01873*. [Online]. Available: http://arxiv.org/abs/1707.01873

[36] N. Schneider, "Decentralization: An incomplete ambition," *J. Cultural Economy*, vol. 12, no. 4, pp. 265–285, 2019.

[37] R. Chatterjee and R. Chatterjee, "An overview of the emerging technology: Blockchain," in *Proc. 3rd Int. Conf. Comput. Intell. Netw. (CINE)*, Oct. 2017, pp. 126–127.

[38] A. Narayanan and M. Möser, "Obfuscation in bitcoin: Techniques and politics," 2017, *arXiv:1706.05432*. [Online]. Available: http://arxiv.org/abs/1706.05432

[39] M. A. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras, and H. Janicke, "Blockchain technologies for the Internet of Things: Research issues and challenges," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2188–2204, Apr. 2019.

[40] Y. Kim, R. K. Raman, Y.-S. Kim, L. R. Varshney, and N. R. Shanbhag, "Efficient local secret sharing for distributed blockchain systems," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 282–285, Feb. 2019.

[41] M. Liu, W. Dou, S. Yu, and Z. Zhang, "A decentralized cloud firewall framework with resources provisioning cost optimization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 3, pp. 621–631, Mar. 2015.

[42] P. Wendell, J. W. Jiang, M. J. Freedman, and J. Rexford, "DONAR: Decentralized server selection for cloud services," in *Proc. ACM SIGCOMM Conf. SIGCOMM*, 2010, pp. 231–242.

[43] K. Wang, M. Lin, F. Ciucu, A. Wierman, and C. Lin, "Characterizing the impact of the workload on the value of dynamic resizing in data centers," in *Proc. 12th ACM SIGMETRICS/PERFORMANCE Joint Int. Conf. Meas. Model. Comput. Syst.*, 2012, pp. 405–406.

[44] H. Li, M. Muskulus, and L. Wolters, "Modeling job arrivals in a data-intensive grid," in *Proc. Workshop Job Scheduling Strategies Parallel Process.*, 2006, pp. 210–231.

[45] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *ACM Trans. Comput. Syst.*, Sep. 1984. [Online]. Available: https://arxiv.org/abs/cs/9809099

[46] M. H. Jakubowski, R. Venkatesan, and Y. Yacobi. *Quantifying Trust*. Accessed: Nov. 30, 2020. [Online]. Available: https://eprint.iacr.org/2010/246.pdf

[47] G. Smith, "Quantifying information flow using min-entropy," in *Proc. 8th Int. Conf. Quant. Eval. Syst.*, Sep. 2011, pp. 159–167.

[48] S. J. Murdoch, "Quantifying and measuring anonymity," in *Proc. Data Privacy Manage. Auto. Spontaneous Secur.*, 2013, pp. 3–13.

[49] T. Lu, Z. Du, and Z. Jane Wang, "A survey on measuring anonymity in anonymous communication systems," *IEEE Access*, vol. 7, pp. 70584–70609, 2019.

[50] G. Smith, "On the foundations of quantitative information flow," in *Proc. Int. Conf. Found. Softw. Sci. Comput. Struct.*, 2009, pp. 288–302.

[51] G. Tóth, Z. Hornák, and F. Vajda, "Measuring anonymity revisited," in *Proc. NORDSEC*, 2004, pp. 85–90.

[52] S. Clauß and S. Schiffner, "Structuring anonymity metrics," in *Proc. 2nd ACM Workshop Digit. Identity Manage.*, 2006, pp. 55–62.

[53] M. Liu, F. R. Yu, Y. Teng, V. C. M. Leung, and M. Song, "Performance optimization for blockchain-enabled industrial Internet of Things (IIoT) systems: A deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3559–3570, Jun. 2019.

[54] M. Liu, Y. Teng, F. R. Yu, V. C. M. Leung, and M. Song, "Deep reinforcement learning based performance optimization in blockchain-enabled Internet of vehicle," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[55] B. Huang, Z. Liu, J. Chen, A. Liu, Q. Liu, and Q. He, "Behavior pattern clustering in blockchain networks," *Multimedia Tools Appl.*, vol. 76, no. 19, pp. 20099–20110, Oct. 2017.

[56] H. Tang, Y. Jiao, B. Huang, C. Lin, S. Goyal, and B. Wang, "Learning to classify blockchain peers according to their behavior sequences," *IEEE Access*, vol. 6, pp. 71208–71215, 2018.

[57] J.-W. Liao, T.-T. Tsai, C.-K. He, and C.-W. Tien, "SoliAudit: Smart contract vulnerability assessment based on machine learning and fuzz testing," in *Proc. 6th Int. Conf. Internet Things, Syst., Manage. Secur. (IOTSMS)*, Oct. 2019, pp. 458–465.

[58] M. Zhao-Hui, Z. Gan-Sen, L. Wei-Wen, M. Ze-Feng, W. Xin-Ming, C. Bing-Chuan, and L. Cheng-Chuang, "Research on DDoS attack detection in software defined network," in *Proc. Int. Conf. Cloud Comput., Big Data Blockchain (ICCBB)*, 2018, pp. 1–6.

[59] A. Goel, A. Agarwal, M. Vatsa, R. Singh, and N. Ratha, "Securing CNN model and biometric template using blockchain," 2020, *arXiv:2008.00054*. [Online]. Available: http://arxiv.org/abs/2008.00054

[60] R. Doku, D. B. Rawat, M. Garuba, and L. Njilla, "LightChain: On the lightweight blockchain for the Internet-of-Things," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, Jun. 2019, pp. 444–448.

[61] Y.-D. Yang, Y.-G. Li, Q. Yang, Z. Chen, and W.-L. Liu, "Blockchain application based smart power grid system," in *Proc. Int. Conf. Appl. Techn. Cyber Secur. Intell.*, 2020, pp. 383–390.

[62] Y. Zuo, S. Jin, and S. Zhang, "Computation offloading in the untrusted MEC-aided mobile blockchain IoT system," 2019, *arXiv:1911.08255*. [Online]. Available: http://arxiv.org/abs/1911.08255

[63] X. Liang, S. Shetty, and D. Tosh, "Exploring the attack surfaces in blockchain enabled smart cities," in *Proc. IEEE Int. Smart Cities Conf. (ISC)*, Sep. 2018, pp. 1–8.

[64] C. Laneve and A. Veschetti. (2020). *A Formal Analysis of Blockchain Consensus*. [Online]. Available: http://www.cs.unibo.it/~laneve/papers/AFABC.pdf

[65] J. Niu, C. Feng, H. Dau, Y.-C. Huang, and J. Zhu, "Analysis of nakamoto consensus, revisited," 2019, *arXiv:1910.08510*. [Online]. Available: http://arxiv.org/abs/1910.08510

[66] Wikipedia Contributors. (2020). *Centrality—Wikipedia, the Free Encyclopedia*. Accessed: May 22, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Centrality

[67] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Netw.*, vol. 1, no. 3, pp. 215–239, Jan. 1978.

[68] F. A. Rodrigues, P. R. Villas Boas, G. Travieso, and L. da F. Costa, "Seeking the best Internet model," 2007, *arXiv:0706.3225*. [Online]. Available: http://arxiv.org/abs/0706.3225

[69] E. Stefanov and E. Shi, "Multi-cloud oblivious storage," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2013, pp. 247–258.

[70] J. J. Wylie, M. W. Bigrigg, J. D. Strunk, G. R. Ganger, H. Kiliccote, and P. K. Khosla, "Survivable information storage systems," *Computer*, vol. 33, no. 8, pp. 61–68, 2000.

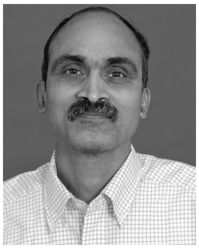
[71] A. Narayanan, V. Toubiana, S. Barocas, H. Nissenbaum, and D. Boneh, "A critical look at decentralized personal data architectures," 2012, *arXiv:1202.4503*. [Online]. Available: http://arxiv.org/abs/1202.4503

[72] Z. Wang, K. Sun, J. Jing, and S. Jajodia, "Verification of data redundancy in cloud storage," in *Proc. Int. Workshop Secur. Cloud Comput.-Cloud Comput.*, 2013, pp. 11–18.

[73] Wikipedia Contributors. (2020). *Diversity Index—Wikipedia, the Free Encyclopedia*. Accessed: May 22, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Diversity_index

SARADA PRASAD GOCHHAYAT received the M.Tech. degree in signal processing from IIT Guwahati, India, in 2010, and the Ph.D. degree in communication engineering from the Indian Institute of Science, India, in 2016. From 2016 to 2018, he was a Postdoctoral Fellow with the Department of Mathematics, University of Padua, Italy. He is currently a Postdoctoral Fellow with the Virginia Modeling, Analysis and Simulation Center, Suffolk, VA, and an Adjunct Faculty with Old Dominion University, Norfolk, USA. His research interests include security and privacy in distributed computing and networks, especially in the IoT, cloud computing, and blockchain.

SACHIN SHETTY (Senior Member, IEEE) received the Ph.D. degree in modeling and simulation from Old Dominion University, in 2007. He was an Associate Professor with the Electrical and Computer Engineering Department, Tennessee State University, USA. He is currently an Associate Professor with the Virginia Modeling, Analysis and Simulation Center, Old Dominion University. He holds a joint appointment with the Department of Modeling, Simulation, and Visualization Engineering and the Center for Cybersecurity Education and Research. He has authored or coauthored over 125 research articles in journals and conference proceedings and two books. His research interests include the intersection of computer networking, network security, and machine learning. He was a recipient of the DHS Scientific Leadership Award. He has served on the Technical Program Committee of ACM CCS, IEEE INFOCOM, IEEE ICDCN, and IEEE ICCCN.



RAVI MUKKAMALA (Associate Member, IEEE) received the Ph.D. degree in computer science from The University of Iowa, in 1987, and the M.B.A. degree from Old Dominion University, in 1993. He is currently a Professor and the Chair of the Department of Computer Science, Old Dominion University. He has published over 200 research articles in journals, books, and conference proceedings. His primary research interests include distributed systems, cybersecurity, blockchain and cryptocurrencies, and privacy-preserving data mining.

PETER FOYTIK received the bachelor's degree in computer science and the Master of Science degree in modeling and simulation. He is currently pursuing the Ph.D. degree in modeling and simulation with a research area in autonomous problem space exploration. He has been a modeling and simulation Professional for over 12 years at the Virginia Modeling Analysis and Simulation Center (VMASC). With a background in computer science, his initial expertise has been in software development of support tools for simulations and models. He has extensive experience with work on application development that integrates with various transportation modeling and simulation tools. His projects include analysis applications that support macroscopic and mesoscopic model development. His research interests include microscopic and agent-based simulations integrated with custom applications giving greater control and analysis capabilities to the user. His current research interests include utilized artificial intelligence methods to improve performance and calibration of models and simulations. His skill set has been applied even further to development with cyber-security applications for distributed systems and in the development of simulation tools to study these systems.

GEORGES A. KAMHOUA received the Ph.D. degree in computer science from Florida International University, in 2019. He is currently a Postdoctoral Research Associate with the Virginia Modeling, Analysis, and Simulation Center, Old Dominion University. His research interests include cybersecurity applied to online social networks, mobile wireless sensor networks, mobile computing, and crowdsourcing. His current research interests include federated learning platforms by evaluating various threats, developing new frameworks to show weaknesses, and enhancing existing defense techniques. His research contributions have been published in renowned conferences, such as ICDCS, IPCCC, and Trustcom. He has served on the Steering Committee, the Technical Program Committee, and a Reviewer for EAI EmergencyComm, IEEE ICCCN, and IEEE ICC, respectively.

LAURENT NJILLA (Member, IEEE) received the B.S. degree from the Department of Computer Science, University of Yaoundé, Yaoundé, Cameroon, the M.S. degree from the University of Central Florida, Orlando, FL, USA, and the Ph.D. degree from the Electrical and Computer Engineering Department, Florida International University, Miami, FL, USA. He is currently a Research Engineer with the Air Force Research Laboratory, Department of Defense, Rome, NY, USA. His current research interests include cyber security, game theory, hardware and network security, blockchain technology, cyber threat information, and advanced computer networking.