

# On the Trade-Off between Degree of Decentralization and Scalability in Proof-of-Stake–based Blockchain Systems

Frederic von Normann\*, Mansur Masama Aliyu†, Niclas Kannengießer†, Lan Bao Quang Le‡

Robert Heinrich‡, Ali Sunyaev†

\*SWIAT GmbH, Frankfurt am Main, Germany

frederic.vonnormann@swiat.io

†Technical University of Munich, Campus Heilbronn, Germany

{mansur.aliyu, niclas.kannengiesser, sunyaev}@tum.de

‡Ulm University, Ulm, Germany

{lan.le, robert.heinrich}@uni-ulm.de

**Abstract**—Proof-of-stake–based (PoS-based) blockchain systems exhibit a trade-off between degree of decentralization (DoD) and scalability: equitable participation of validating nodes can raise latency and degrade throughput, while scaling can reduce equitable participation. To meet application requirements, software architects must balance these dimensions using comparable metrics. We curated a compact metric set for DoD (block-proposal randomness [Shannon entropy], token concentration [HHI], wealth distribution [Gini]) and for scalability (transaction throughput, first-inclusion latency) and applied it to Tezos Hangzhou 2 (Octez v12.3). Across 1,573 controlled configurations with varying validating-node count  $n$ , block interval, and block size limit, we normalized DoD by  $n$  to enable cross-configuration comparison. Under this normalization, DoD–scalability associations were weaker and not uniformly negative. Scalability was primarily influenced by validating-node count. Our main contributions are a benchmarking approach with curated metrics, an empirical mapping of configuration levers to both dimensions, evidence for the dominant role of  $n$ , and guidance for configuring PoS-based blockchain systems.

**Index Terms**—Benchmarking, blockchain technology, proof-of-stake, software engineering, Tezos, trade-offs

## I. INTRODUCTION

Proof-of-stake–based (PoS-based) blockchain systems increasingly support decentralized applications (DApps) with non-functional requirements such as equitable validator participation, low latency, and high throughput. These requirements are closely tied to two fundamental blockchain system dimensions: degree of decentralization (DoD) and scalability. For example, equitable participation relates to a high DoD, while latency and throughput are associated with scalability. However, DoD and scalability are often in tension: enhancing one typically reduces the other. Maximizing both simultaneously is rarely feasible [1, 2]. As a result, software architects must identify Pareto-optimal blockchain system configurations that best align with the specific DApp requirements.

This work was supported by funding from the topic Engineering Secure Systems of the Helmholtz Association (HGF), by KASTEL Security Research Labs, and the project FeCoMASS by the German Research Foundation (DFG) under project number 499241390.

Identification of Pareto-optimal system configurations requires metrics that are consistent and comparable across configurations and a clear mapping from configuration levers to outcomes. While prior work proposes constructs and metrics for DoD and scalability individually [e.g., 3, 4], they are typically evaluated in isolation [e.g., 5, 6]. Their combined effectiveness in capturing the DoD–scalability trade-off remains unclear. For example, Shannon entropy [5, 7], used to assess block-proposal randomness, and availability [4], used to estimate scalability, may not exhibit the expected negative association that would signify a trade-off. This complicates metric selection and risks suboptimal configuration decisions. To support software architects in making better-informed and purposeful design decisions, we approach the following research question: *What combinations of metrics are suitable to quantify the DoD–scalability trade-off in PoS-based blockchain systems?*

We curated a set of constructs operationalized through metrics related to DoD and scalability. Such constructs include block-proposal randomness associated with DoD and transaction throughput associated with scalability [1]. We retained candidate metrics applicable to the Tezos system. Next, we collected data in a controlled private network across 1,573 configurations of the Tezos economic protocol Hangzhou 2 (PtHangz2, protocol 011). We then analyzed pairwise correlations between *normalized* DoD indices and scalability metrics. Consistent with our normalization by the validating-node count  $n$ , the resulting correlations are *weaker and not uniformly negative*. This isolates within-set equitability of participation (a component of DoD) from the dominant scalability effect of the validating-node count.

By presenting a benchmark approach, this work supports purposeful configuration of PoS-based blockchain systems under the DoD–scalability trade-off in four ways. First, it enables consistent quantification by curating a compact, comparable set of constructs operationalized through metrics, reducing cross-configuration inconsistencies. Second, this work empir-

ically maps how validating-node count, block interval, and block size limit shape both dimensions, identifying configurations that maximize either DoD or scalability. Third, it shows that correlations between *normalized* DoD and scalability values are weaker and not uniformly negative; this clarifies that validating-node count drives the dominant scalability effect and suggests that within- $n$  equitability plays a secondary role once isolated by normalization. Fourth, we distilled the results into concrete guidance: first set the validating-node count to meet DoD requirements, then tune block interval and block size limit for transaction throughput and first-inclusion latency—and provide a reproducible foundation for studies of the Tezos system and related PoS-based blockchain systems.

The remainder of this work is structured as follows. Section II discusses the theoretical background and related work. Section III introduces common metrics for measuring DoD and scalability. Section IV details the experimental setup and reports the experimental results. In Section V, we discuss principal findings, contributions, limitations, and future research directions. We conclude in Section VI.

## II. BACKGROUND AND RELATED WORK

This section outlines the notions of DoD and scalability in blockchain systems and summarizes related research on their empirical assessment.

### A. The Trade-Off between Degree of Decentralization and Scalability

The DoD–scalability trade-off is inherent in state-machine replication with consensus [1, 8, 9]. As illustrated in Figure 1, blockchain systems typically comprise *transaction originators* (clients) and *validating nodes*. Originators create transactions and submit them to validating nodes. Validating nodes participate in one or more consensus-critical tasks (e.g., transaction validation, block construction, state execution, and storage) [1]. We use *validating node* generically, independent of system-specific role names.

DoD refers to the extent to which validating nodes participate equitably and autonomously in consensus [1]. High DoD implies broadly distributed block-proposal opportunities and influence. Achieving such equitable participation generally increases communication and coordination overhead, which can reduce scalability [2, 10].

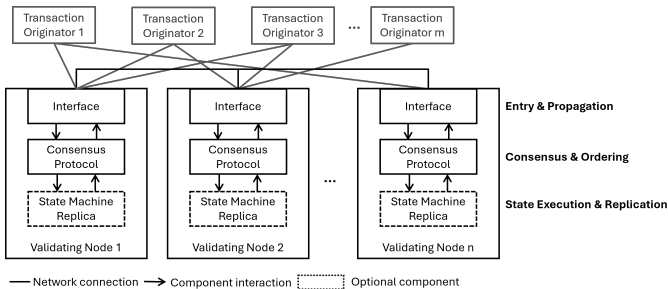


Fig. 1: Overview of blockchain system architectures (adapted from [1]).

Scalability denotes the ability to sustain higher workloads (e.g., transaction throughput) with acceptable transaction finalization latency. Typical scalability improvements entail reducing the validating-node count involved per decision or per time unit to lower communication costs [1, 10]. This can concentrate block-proposal frequency or influence, thereby reducing effective DoD. Consequently, maximizing DoD and scalability simultaneously is rarely feasible if at all; productive blockchain systems must balance the two.

### B. Related Research

Trade-offs between dimensions of blockchain system designs have been explored in both conceptual and empirical studies. Conceptual works [e.g., 2, 10, 11] provide valuable overviews of the DoD–scalability trade-off but lack empirical validation and guidance for identifying Pareto-optimal configurations that satisfy non-functional requirements.

Empirical works [e.g., 3, 6, 8, 12] complement conceptual research by presenting metrics to quantify constructs for measuring DoD and scalability [e.g., 5, 6, 13]. However, most empirical studies focus on the trade-off between scalability and security [e.g., 3, 4, 14, 15, 16], while the DoD–scalability trade-off remains underexplored. This limits software architects’ ability to select metric combinations that expose the relevant relationships—negative or otherwise—needed to identify Pareto-optimal blockchain system configurations for DApps.

In summary, while the DoD–scalability trade-off is well recognized, empirical quantification through combined metrics remains limited [1]. The diversity of constructs and their varied operationalizations further complicates metric selection for blockchain system design. This work addresses this gap by identifying and applying suitable metric combinations to empirically analyze the DoD–scalability trade-off in PoS-based blockchain systems, with a focus on the Tezos protocol.

## III. CONSTRUCTS AND METRICS FOR THE DoD–SCALABILITY TRADE-OFF

To investigate DoD and scalability in blockchain systems, prior work has proposed several benchmarking frameworks, including BBSF [14], BLOCKBENCH [3], and Diablo [4]. Building on Aliyu et al. [1], we summarize applicable constructs for PoS-based systems in Table I and formalize their operationalizations below.

### A. Metrics Related to Degree of Decentralization

We identified three operationalized constructs to quantify the DoD of PoS-based blockchain systems.

**Block-proposal randomness (H):** *Uncertainty of the next block proposer.* We measure block-proposal randomness as the Shannon entropy (bits) of the random variable ‘block-proposer identity’ computed from the empirical proposer distribution over the observation window:

$$H = - \sum_{i=1}^n p_i \log_2 p_i, \quad p_i = \frac{b_i}{\sum_{j=1}^n b_j}, \quad (1)$$

where  $n$  is the validating-node count and  $b_i$  the number of blocks proposed by validating node  $i$  (with  $0 \log 0 := 0$ ). Let

TABLE I: Overview of identified constructs and their applicability to PoS-based blockchain systems.

	Construct	Applicability	Justification for Applicability and Inclusion/Exclusion
Degree of Decentralization	Block-Proposal Randomness	✓	Validating nodes are (pseudo-)randomly selected to propose the next block.
	Geographical Diversity	✗	Geographic spread reflects dispersion sociotechnically, but controlled experiments are confounded by external factors beyond this study's scope.
	Hashing Power Distribution	✗	PoS protocols do not use mining for block generation or consensus.
	Token Concentration	✓	Block rewards accrue to proposers; higher concentration reflects repeated success of a few nodes.
	Wealth Distribution	✓	Reward dynamics affect token dispersion across validating nodes.
Scalability	Availability	✓	Validating-node availability underpins processing and consensus at scale.
	Transaction Finalization Latency	✓	Time to first on-chain inclusion captures user-perceived responsiveness.
	Transaction Throughput	✓	Blockchain systems process a bounded number of transactions per time unit.

*Availability is applicable but not evaluated in our experiments; we focus on throughput and first-inclusion latency.*

$k = |\{i : p_i > 0\}|$  be the validating-node count that proposed at least one block; then  $H \leq \log_2 k$ , with the maximum  $\log_2 n$  reached only if all validating nodes have equal nonzero proposal probability.

For comparability across configurations with different  $n$ , we use the  $n$ -based normalization

$$H_{\text{norm}} = \frac{H}{\log_2 n}, \quad H \in [0, \log_2 n], \quad H_{\text{norm}} \in [0, 1]. \quad (2)$$

$H_{\text{norm}} = 1$  when proposer probabilities are perfectly even across all  $n$  validating nodes.

**Token concentration:** *Distribution of stake across validating nodes.* We compute the Herfindahl–Hirschman Index (HHI) as

$$\text{HHI} = \sum_{i=1}^n \left( \frac{t_i}{t_{\text{total}}} \right)^2, \quad (3)$$

where  $t_i$  is the stake of validating node  $i$  and  $t_{\text{total}} = \sum_i t_i$ .

For comparability across systems with different  $n$ , we use the standard  $n$ -based normalization

$$\text{HHI}_{\text{norm}} = \frac{\text{HHI} - \frac{1}{n}}{1 - \frac{1}{n}} \in [0, 1], \quad (4)$$

which maps equal shares to 0 and monopoly to 1.

**Wealth distribution:** *Inequality of token holdings.* We use the Gini coefficient:

$$\text{Gini} = \frac{\sum_{i=1}^n \sum_{j=1}^n |t_i - t_j|}{2n \sum_{i=1}^n t_i}, \quad (5)$$

where  $n$  is the validating-node count and  $t_i$  is the token holdings of validating node  $i$ .

We normalize the Gini coefficient as

$$\text{Gini}_{\text{norm}} = \text{Gini} \cdot \frac{n}{n-1} \in [0, 1]. \quad (6)$$

### B. Metrics Related to Scalability

We identified two operationalized constructs to quantify the scalability of PoS-based blockchain systems.

**Transaction finalization latency (TFL):** *Time from transaction submission to first on-chain inclusion.*

$$\text{TFL} = t_{\text{incl}} - t_{\text{prop}}, \quad (7)$$

where  $t_{\text{prop}}$  is the submission timestamp and  $t_{\text{incl}}$  is the timestamp of the first block that includes the transaction. Following prior work, we use ‘finalization’ operationally to mean first inclusion; this is distinct from economic finality. We therefore refer to TFL as *first-inclusion latency*.

**Transaction throughput (TPS):** *Rate of first inclusions over an observation window.*

$$\text{TPS} = \frac{\text{NumOfConfTr}}{t_1 - t_0}, \quad (8)$$

where NumOfConfTr is the number of distinct transactions first included within  $[t_0, t_1]$ , and  $t_0, t_1$  denote the bounds of the observation window.

## IV. QUANTIFICATION OF THE TRADE-OFF BETWEEN DoD AND SCALABILITY

This section describes the data collection and analysis used to identify combinations of operationalized constructs that quantify the DoD–scalability trade-off.

### A. Experimental Design

We collected data on a private deployment of the Tezos protocol and analyzed the construct combinations in Section III. Below, we detail data collection and analysis.

1) *Data Collection:* We conducted a laboratory experiment [17] using a private deployment of the Tezos economic protocol (Hangzhou 2, PtHangz2, protocol 011) executed by Octez v12.3 nodes. The experimental environment consisted of 126 identical virtual machines, each hosting one validating node in a private cloud. A total of 1,703 experiments were executed, each with a unique blockchain system configuration. We changed blockchain system configurations in terms of the parameters *block interval*, *block size limit*, and *validating-node count* (see Table II).

Before each experiment, the Tezos system was initialized with equal stake allocation across all participating validating nodes. All validating nodes were fully synchronized and operational before workload injection to ensure experimental validity and eliminate initialization artifacts. The workload consisted only of standard value transactions that transferred tez, the native token of the Tezos protocol. To decouple transaction originator overheads from validating node/network behavior, all transactions were pre-built and signed.

Transactions were submitted evenly across the network by one transaction originator per validating node, using dedicated accounts that were not involved in staking. To avoid nonce management issues, each transaction originated from a separate account. This uniform distribution balanced the workload among all validating nodes.

For each configuration, the blockchain system was tested under the workload of 105% of the configuration-specific gas-bound theoretical throughput. We focused on experiments with a 105% workload to ensure that scalability metrics were not limited by insufficient transaction volume. The configuration-specific gas-bound theoretical throughput limit is

$$\text{TPS}_{\text{gas}} = \left\lfloor \frac{B_g}{g_{\text{tx}}} \right\rfloor / \Delta, \quad (9)$$

where  $B_g$  is the block size limit,  $\Delta$  is the block interval, and  $g_{\text{tx}}=1,420$  gas for a value-only transfer on Hangzhou 2. If a byte limit  $B_b$  binds, the limit becomes  $\min(\lfloor B_g/g_{\text{tx}} \rfloor, \lfloor B_b/s_{\text{tx}} \rfloor) / \Delta$ , with  $s_{\text{tx}}$  the average transaction size in bytes. Measured transaction throughput is expected to be below such ideals due to consensus overheads and network delays.

The workload level was executed as two consecutive workload cycles. Each cycle lasted exactly one block interval and was precisely synchronized with the block schedule. Transactions were submitted in a rate-controlled manner throughout each cycle to precisely match the target transaction throughput and avoid bursts.

All relevant data were logged in real-time. This included transaction submission timestamps recorded by the transaction originators as well as transaction processing and inclusion events captured in the validating node logs. The blockchain state, including blocks, endorsements, and transactions, was stored as part of normal protocol execution. Upon completion of each experiment, logs from all transaction originators and validating nodes were parsed and aggregated. Protocol-level data, such as block metadata and endorsements, were extracted

TABLE II: Configuration parameters and their domains used in the experiments

Input Parameter	Min Value	Max Value
Block Interval (s)	7	120
Block Size Limit (gas) <sup>1</sup>	28,400	7,500,000
Validating-Node Count	6	126

<sup>1</sup>Each standard value transfer consumes exactly 1,420 gas; the range corresponds to 20–5,281 transactions per block.

and stored centrally. These data served as inputs for the operationalized constructs (see Section III). Further details are available in our public repository [18].

To ensure data quality, 130 out of the 1,703 experiments were excluded due to permanent crashes of validating nodes, significant workload submission mismatches, and incomplete logs. The final data set comprised 1,573 configurations.

2) *Data Analysis*: We used the data collected to compute values for all applicable metrics in the scope of this work (see Section III). For token concentration (HHI) and wealth distribution (Gini), we used the stake balance per validating node at the *end* of each run (initial allocation plus block rewards). Because runs are short and initial balances are equal, rewards are small relative to initial balances. Block-proposal randomness (Shannon entropy) used block counts  $b_i$  per validating node. We normalized block-proposal randomness, token concentration, and wealth distribution using Eq. 2, Eq. 4, and Eq. 6, respectively, with respect to the validating-node count  $n$  to allow comparability across experiments.

Transaction throughput was computed as  $\text{TPS} = \text{NumOfConfTr}/(t_1-t_0)$  with  $(t_0, t_1)$  taken from the experiment schedule.

Transaction finalization latency (Eq. 7) used submission ( $t_{\text{prop}}$ ) and first-inclusion timestamps ( $t_{\text{incl}}$ ).

We computed Spearman’s rank correlation coefficient ( $\rho$ ) for all DoD–scalability pairs. For consistent interpretation (‘higher is better’), we inverted the values of *HHI*, *Gini*, and *TFL*. The original correlations between non-normalized values consistently showed strong negative correlations, primarily due to the significant influence of the validating-node count. For example, block-proposal randomness vs. transaction throughput  $\rho = -0.121$  ( $p < 0.001$ ), token concentration vs. transaction throughput  $\rho = -0.1663$  ( $p < 0.001$ ), block-proposal randomness vs. transaction finalization latency  $\rho = -0.343$  ( $p < 0.001$ ), and wealth distribution vs. transaction throughput  $\rho = -0.123$  ( $p < 0.001$ ).

We used the normalized values of DoD metrics to remove the dominant effect of validating-node count that couples DoD and scalability (i.e., larger  $n$  leads to lower transaction throughput and higher transaction finalization latency). To keep directions consistent (‘higher is better’), we used  $H_{\text{norm}}$ ,  $1 - \text{HHI}_{\text{norm}}$ ,  $1 - \text{Gini}_{\text{norm}}$ , and  $1/\text{TFL}$  (so higher values indicate faster first inclusion). Because Spearman is rank-based, these monotone transformations only flip signs where needed; the magnitudes of  $\rho$  and the  $p$ -values are unchanged. After normalization, correlations with scalability reflect within- $n$  equitability (our operational component of DoD), separated from validating-node count. We visualized relationships using rank-based smoothers.

### B. Trade-Off between DoD and Scalability in a Tezos System

The following first examines blockchain system configurations that maximize either DoD or scalability under heavy workload (105%), followed by an analysis of the correlations between these metrics that reveal the inherent trade-off.

1) *Optimization of Blockchain System Configuration to Maximize DoD and Scalability*: Maximization of DoD or scalability often requires conflicting system configurations.

a) *Maximization of DoD*: Figures 2a, 2b, and 2c illustrate influences of blockchain system configurations on DoD, using block-proposal randomness, token concentration, and wealth distribution. The configurations that lead to maximum DoD per construct are reported in the corresponding figure captions. Across all constructs, non-normalized DoD showed a strong dependency on network size. This is expected because the maximum block-proposal randomness is upper-bounded by the validating-node count:  $\max H = \log_2 n$ . Consequently, network size and stake dispersion set the ceiling for block-proposal randomness.

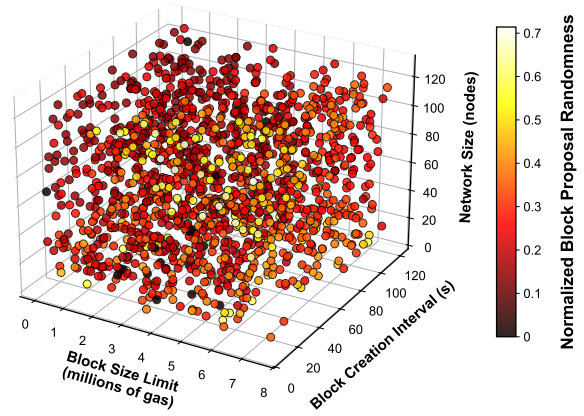
Block size limit and block interval exhibited only negligible effects on overall DoD. For block-proposal randomness specifically, we observe a mild association whereby smaller block size limits and longer block intervals correlate with slightly higher measured Shannon entropy. We attribute this to short-horizon dynamics under our offered-load pattern (e.g., fewer transient proposer concentration effects) rather than to a causal influence of these parameters on proposer selection. In Tezos (Hangzhou 2) and similar PoS-based systems, proposer identity is primarily determined by the protocol’s (stake-weighted) selection process and the stake distribution; block size and block interval can at most affect Shannon entropy indirectly (e.g., via empty/failed slots or timing artifacts). These observations should therefore be interpreted as *empirical and setup-specific*, not as general design levers for DoD.

For software architects, this implies that network size and stake dispersion are the primary levers for DoD; block size limit and block interval can be tuned mainly for scalability with little systematic impact on DoD when  $n$  is sufficiently large. Because runs start from equal stakes and are short, HHI and Gini remain near equal-share baselines; we therefore rely on block-proposal randomness as a more sensitive proxy for short-horizon DoD estimates.

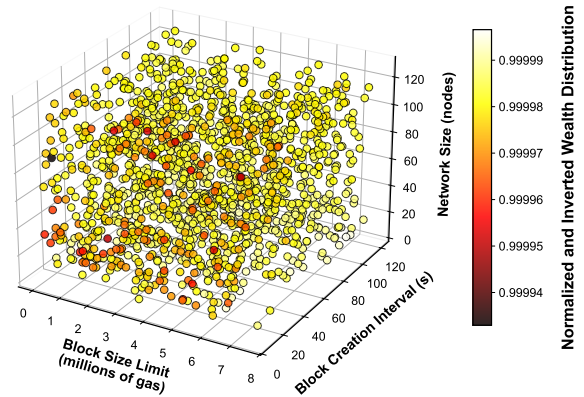
b) *Maximization of Scalability*: Figures 3a and 3b illustrate the identified influences of blockchain system configurations on scalability, measured in terms of transaction finalization latency and transaction throughput at a workload level of 105%. Configurations resulting in particularly high or low scalability per construct are reported in the figure captions.

Transaction finalization latency increases in configurations with large block intervals and large network sizes, which is plausible because transactions will be processed less frequently. Particularly high transaction throughput can be achieved through short block intervals, medium to large block size limits, and a small validating node count.

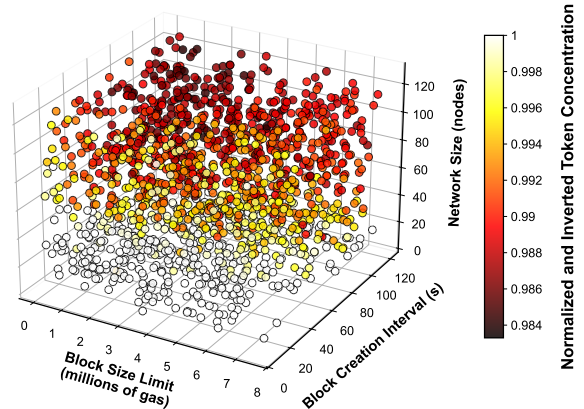
For the lower-bound configuration in Figure 3b ( $B_g \approx 6,000,000$ ,  $g_{tx} = 1,420$ ,  $\Delta \approx 10$  s),  $\text{TPS}_{\text{gas}} \approx [6,000,000/1,420]/10 \approx 422$  TPS; our measured  $\geq 350$  TPS is plausible given protocol and network overheads.



(a) Block size limit 5–7 million gas and block interval 50–100 s at a network size of 70–120 validating nodes yields high normalized block-proposal randomness ( $H_{\text{norm}}$  near 1).

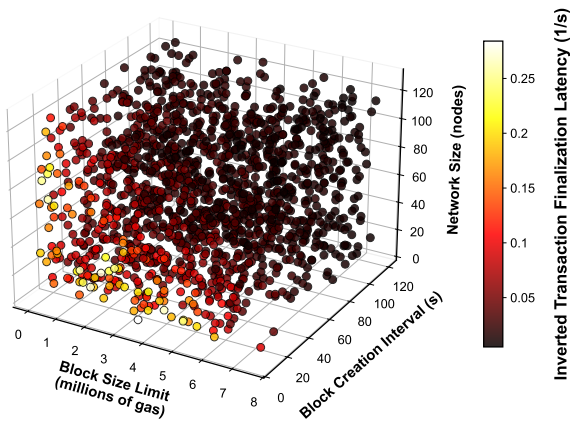


(b) Block size limit 4–8 million gas and block interval 20–40 s at a network size of  $\geq 40$  validating nodes yields low inequality (Gini $_{\text{norm}}$  near 0).

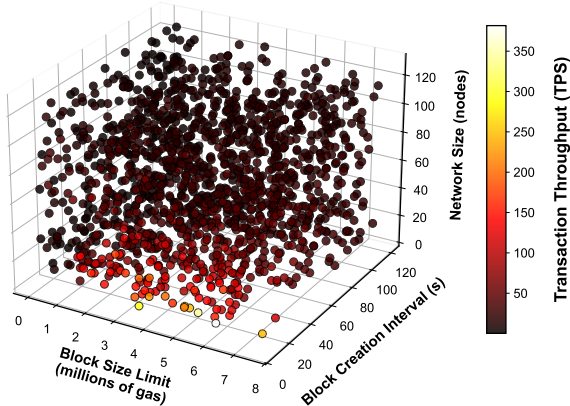


(c) Block size limit 28,400–8 million gas and block interval  $\geq 60$  s at a network size of 40–80 validating nodes yields low normalized token concentration (HHI $_{\text{norm}}$  close to 0, indicating near-equal stake distribution).

Fig. 2: Influences of blockchain system configurations on DoD, highlighting configurations with peak metric values. All DoD values are normalized. Wealth distribution and token concentration are inverted to align with the meaning ‘higher value indicates higher DoD.’



(a) Short block intervals (e.g., 20–40 s), medium–large block size limits (1–6 million gas), and small networks (e.g., 10–40 validating nodes) yielded low transaction finalization latency (e.g., < 50 s).



(b) Block size limit 6–8 million gas, block interval 10 to 25 s, and network size 10–30 validating nodes resulted in transaction throughput of  $\geq 350$  TPS.

Fig. 3: Influences of blockchain system configurations on scalability, highlighting configurations with peak metric values. The scalability values can be interpreted as ‘higher value indicates better scalability.’

Scalability was maximized through the use of short block intervals, large block size limits, and a small validating-node count.

The results confirm that scalability increases with shorter block intervals, larger block sizes, and smaller network sizes, complementing prior findings [2, 11] with new empirical evidence on their combined effects. We demonstrate that transaction finalization latency rises sharply with larger networks and longer block intervals, highlighting network size as a key latency constraint. In contrast, block size and block interval can be jointly adjusted to increase transaction throughput without significantly increasing the transaction finalization latency, if the network size is limited. This implies that software architects can achieve high scalability by balancing

block parameters while recognizing the inherent limits imposed by network size. These findings guide more informed configuration choices to optimize scalability.

2) *Correlations Between Metrics*: Across 1,573 configurations, correlations between *normalized DoD* and scalability values are weaker and sometimes non-negative (Figure 4). With the ‘higher is better’ convention ( $1/\text{TFL}$  for latency), we observe negative and slightly positive correlations for block-proposal randomness and scalability metrics, positive correlations for token concentration and scalability metrics, and negative correlations for wealth distribution with scalability metrics.

This pattern is expected: by normalizing with respect to  $n$ , we removed the direct size dependence of the DoD metrics that dominated raw correlations. What remains is equitability *within a fixed  $n$* . The stake-based token concentration measure ( $\text{HHI}_{\text{norm}}$ ) reflects *realized reward dispersion* over a short window, not long-run stake centralization. Because all runs start from equal stakes and rewards are small but sparse, configurations that increase scalability (smaller  $n$ , shorter block intervals, adequate block size limits) also increase the number of distinct validating nodes that produce at least one block in the observation window. This reduces observed token concentration (increases  $1 - \text{HHI}_{\text{norm}}$ ) and simultaneously increases TPS or  $1/\text{TFL}$ . The resulting positive correlations do not directly contradict a trade-off; they indicate that, over short horizons, stake-based concentration partly tracks the same drivers that improve scalability. As  $k$  (the number of proposers who actually propose) approaches  $n$  in longer horizons, this residual coupling should diminish.

The correlation analysis shows that *validating-node count is the dominant lever*: larger  $n$  reliably reduces scalability. Moreover, when  $n$  is comparable, more even proposer distributions do not uniformly reduce scalability in our Tezos setup; their effects are smaller than the size effect. This supports a configuration process where  $n$  is initially set to meet DoD requirements, and block parameters for scalability are subsequently set.

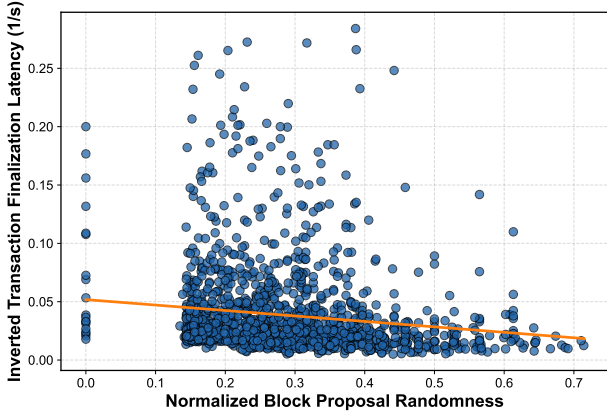
## V. DISCUSSION

The following highlights our principal findings and describes this work’s contributions to research and practice, limitations, and future research directions.

### A. Principal Findings

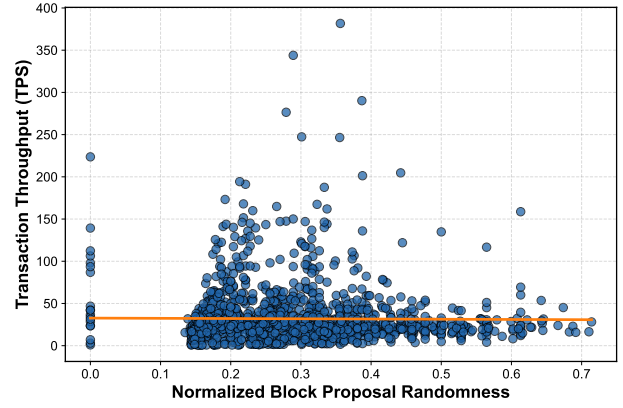
The analysis confirms that conflicting configurations are needed to maximize DoD or scalability. For example, while DoD can be enhanced by increasing the validating-node count, scalability benefits from the opposite. This highlights the inherent trade-off and the need for tailored configuration choices based on specific application goals, emphasizing that no single configuration fits all non-functional requirements of all DApps. The strong influence of the validating-node count on DoD also indicates that core aspects of DoD, such as equitability and autonomy of validating nodes, are not exclusively captured by the constructs used in this study. Confirming assumptions from

Workload Level: 105%  
Spearman: -0.238 (p = 9.6e-22)



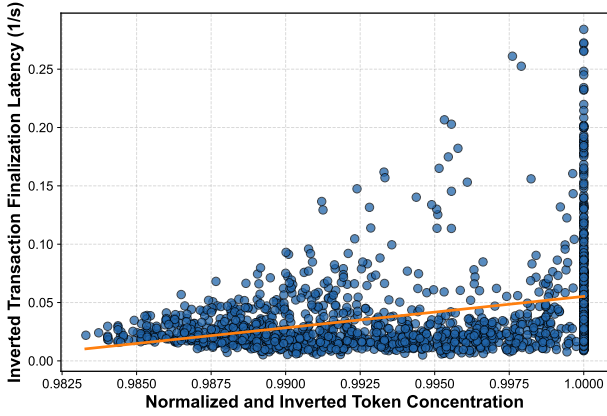
(a) Block-proposal randomness and inverted transaction finalization latency (1/TFL).

Workload Level: 105%  
Spearman: 0.025 (p = 0.32)



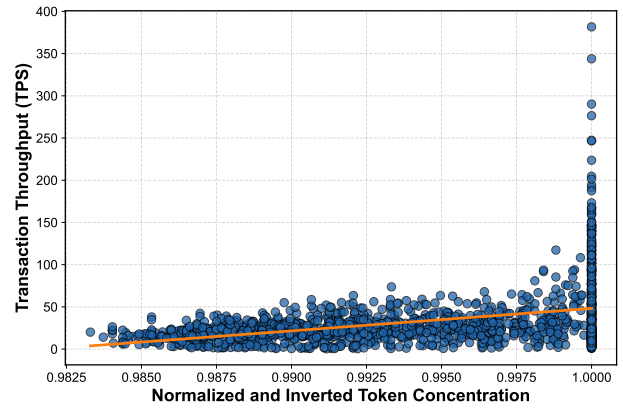
(b) Block-proposal randomness and transaction throughput.

Workload Level: 105%  
Spearman: 0.281 (p = 5e-30)



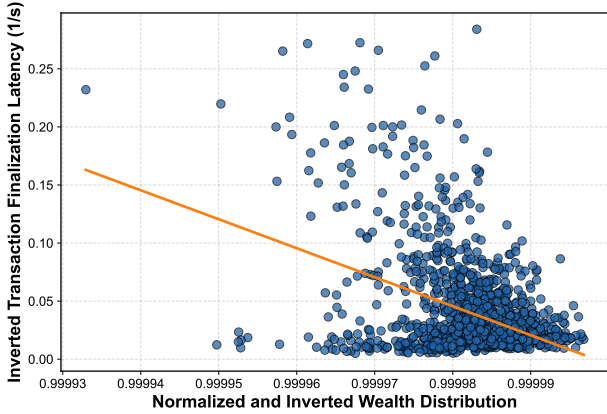
(c) Token concentration and inverted transaction finalization latency (1/TFL).

Workload Level: 105%  
Spearman: 0.359 (p = 3.6e-49)



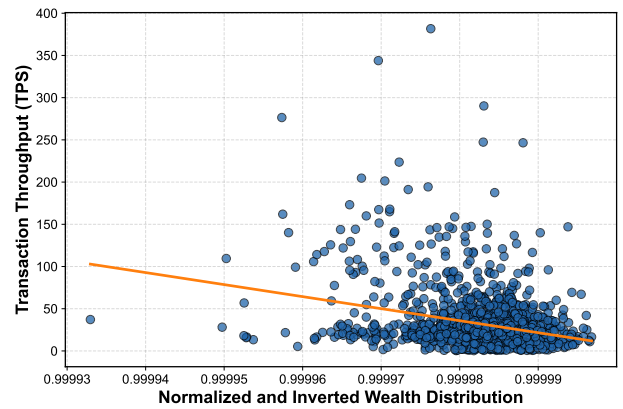
(d) Token concentration and transaction throughput.

Workload Level: 105%  
Spearman: -0.166 (p = 3.4e-11)



(e) Wealth distribution and inverted transaction finalization latency (1/TFL).

Workload Level: 105%  
Spearman: -0.148 (p = 3.9e-09)



(f) Wealth distribution and transaction throughput.

Fig. 4: Correlations between operationalized DoD and scalability constructs. All constructs are harmonized in their meaning: 'higher value is better.'

previous research [1], this necessitates the development of additional metrics that more effectively capture these aspects to measure DoD.

Because our observation windows are short and rewards are sparse, stake-based token concentration partly co-moves with scalability: in more scalable configurations, the Tezos system analyzed tends to distribute at least one reward to more validating nodes within the window, which lowers observed concentration. This explains the occasional positive correlations for  $1 - HHI_{\text{norm}}$ , indicating that  $HHI$  could be more useful in trade-off analyses with longer observation windows.

The varying correlation strengths among construct pairs underscore the importance of construct and metric selection in identifying Pareto-optimal configurations of blockchain systems. Employing multiple constructs is advisable to ensure an exhaustive evaluation. Moreover, to capture DoD as proposed in Section II-A—the degree to which validating nodes equitably and autonomously participate in consensus—DoD metrics should be normalized, as demonstrated in this work.

### B. Contributions

Quantifying the trade-off between DoD and scalability remains a significant challenge in PoS-based blockchain systems. Although various metrics exist to operationalize these concepts, it has been unclear which combinations are most effective for capturing the nuances of this trade-off. This work addresses that gap by presenting a curated set of metrics specifically tailored to the Tezos Hangzhou 2 protocol, enabling software architects to systematically identify Pareto-optimal configurations that balance DoD and scalability.

Using this curated set, the analysis demonstrates how each concept (i.e., DoD and scalability) can be maximized through specific system configurations. The uncovered configurations for the Tezos system not only corroborate existing research on scalability maximization but also extend the state of knowledge by showing, from a technical perspective, how DoD can be optimized. By contrasting configurations that are optimal for either DoD or scalability, we reveal the inherent limitations that prevent their simultaneous maximization in practice.

Furthermore, we provide empirical evidence of correlations between the proposed metrics, highlighting combinations that facilitate navigating the DoD–scalability trade-off. These findings lay a foundation for further investigation into this trade-off in Tezos (Hangzhou 2) and, by extension, in similar PoS-based blockchain systems. The observed correlations also offer practical value for formulating optimization problems aimed at identifying Pareto-optimal configurations within the Tezos protocol.

Overall, this work presents an analysis approach and actionable insights for practitioners and researchers seeking to balance DoD and scalability in PoS-based blockchain systems.

### C. Limitations and Future Research

We collected data on the Tezos protocol (Hangzhou 2) in a private network using a cloud service to maintain full control of parameter manipulations. The virtualization of compute

resources and internal cloud routing may have affected system behavior, potentially causing deviations from real-world conditions, particularly in terms of network delays in public blockchain systems. As a result, observed trade-offs between DoD and scalability metrics, both sensitive to latency, may be more pronounced in physically distributed settings. To overcome this limitation, future research should conduct a benchmark using geographically dispersed validating nodes across different data centers worldwide and from various cloud providers to examine diverse conditions and investigate additional metrics (e.g., geographical diversity).

The data collection is limited to various configurations of the Tezos Hangzhou 2 protocol (protocol 011), which uses the Liquid PoS consensus protocol. Because consensus protocols can differ strongly, even among PoS-based systems, our findings may not be generalizable to other PoS-based blockchain systems such as Solana or Ethereum. Future research should examine additional PoS-based blockchain systems to better understand the trade-off between DoD and scalability. Broader analysis could deepen the foundational knowledge of which system characteristics influence these metrics, aiding in the formulation of optimization problems to identify Pareto-optimal configurations. This could support software architects in designing blockchain systems that meet DApp requirements.

Given the large number of blockchain system configurations examined in this study, we intentionally constrained the observation period, collecting data only until the workload was fully processed. While this approach enabled broad coverage, it strongly limits insight into the long-term evolution of DoD and scalability. Future research could address this limitation by conducting longitudinal analyses to explore how system configurations influence the dynamics of DoD, scalability, and their trade-offs over time.

DoD and scalability are two of the core subconcepts of the blockchain trilemma [1]. Future research should focus on identifying construct combinations that are useful in investigating the full trilemma, involving DoD, scalability, and security. For example, future work could examine success rates or specific attacks in relation to the DoD and scalability trade-off. Such investigations should also include social aspects, such as off-chain governance influences, internet service providers, geographical diversity, and validator churn, that are outside the scope of this study.

## VI. CONCLUSION

Selecting appropriate metrics to identify Pareto-optimal configurations in PoS-based blockchain systems is challenging due to the varying correlations among operationalized constructs. This study emphasizes the importance of thoughtful metric selection and shows how different configurations optimize either DoD or scalability, but not both.

By empirically mapping correlations across a range of metric combinations, this work supports more informed construct selection and lays the groundwork for optimization strategies that explicitly address trade-offs in DApps.

## REFERENCES

- [1] M. M. Aliyu, N. Kannengießer, and S. Ali, “From concept to measurement: A survey of how the blockchain trilemma can be analyzed,” 2025. [Online]. Available: <https://arxiv.org/abs/2505.03768>
- [2] N. Kannengießer, S. Lins, T. Dehling, and A. Sunyaev, “Trade-offs between distributed ledger technology characteristics,” *ACM Computing Surveys*, vol. 53, no. 2, Apr. 2020.
- [3] T. T. A. Dinh, J. Wang, G. Chen, R. Liu, B. C. Ooi, and K.-L. Tan, “Blockbench: A framework for analyzing private blockchains,” in *Proceedings of the 2017 ACM International Conference on Management of Data*, New York, NY, USA, 2017, pp. 1085–1100.
- [4] V. Gramoli, R. Guerraoui, A. Lebedev, C. Natoli, and G. Voron, “Diablo: A benchmark suite for blockchains,” in *Proceedings of the Eighteenth European Conference on Computer Systems*, New York, NY, USA, 2023, pp. 540–556.
- [5] G. Quattrocchi, F. Scaramuzza, and D. A. Tamburri, “The blockchain trilemma: an evaluation framework,” *IEEE Software*, vol. 41, no. 6, pp. 101–110, 2024.
- [6] Y. Fu, M. Jing, J. Zhou, P. Wu, Y. Wang, L. Zhang, and C. Hu, “Quantifying the blockchain trilemma: A comparative analysis of algorand, ethereum 2.0, and beyond,” in *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, 2024, pp. 97–104.
- [7] K. Wu, B. Peng, H. Xie, and Z. Huang, “An information entropy method to quantify the degrees of decentralization for blockchain systems,” in *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication*, Beijing, China, 2019, pp. 1–6.
- [8] S. Mssassi and A. Abou El Kalam, “The blockchain trilemma: A formal proof of the inherent trade-offs among decentralization, security, and scalability,” *Applied Sciences*, vol. 15, no. 1, p. 19, 2024.
- [9] M. Leinweber, N. Kannengießer, H. Hartenstein, and A. Sunyaev, *Leveraging Distributed Ledger Technology for Decentralized Mobility-as-a-Service Ticket Systems*. Wiesbaden: Springer Fachmedien Wiesbaden, 2023, p. 547–567.
- [10] Y. Xiao, N. Zhang, W. Lou, and Y. T. Hou, “A survey of distributed consensus protocols for blockchain networks,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1432–1465, 2020.
- [11] O. O’Donoghue, A. A. Vazirani, D. Brindley, and E. Meinert, “Design choices and trade-offs in health care blockchain implementations: systematic review,” *Journal of medical Internet research*, vol. 21, no. 5, p. e12426, 2019.
- [12] T. Nakai, A. Sakurai, S. Hironaka, and K. Shudo, “A formulation of the trilemma in proof of work blockchain,” *IEEE Access*, vol. 12, pp. 80 559–80 578, 2024.
- [13] A. Gervais, G. O. Karame, K. Wüst, V. Glykantzis, H. Ritzdorf, and S. Capkun, “On the security and performance of proof of work blockchains,” in *Proceedings of the 2016 ACM SIGSAC conference on Computer and Communications Security*, Vienna, Austria, 2016, pp. 3–16.
- [14] K. Ren, J. F. Van Buskirk, Z. Y. Ang, S. Hou, N. R. Cable, M. Monares, H. F. Korth, and D. Loghin, “Bbsf: blockchain benchmarking standardized framework,” in *Proceedings of the 1st Workshop on Verifiable Database Systems*, New York, NY, USA, 2023, pp. 10–18.
- [15] M. Alharby and A. van Moorsel, “Blocksim: An extensible simulation tool for blockchain systems,” *Frontiers in Blockchain*, vol. 3, p. 28, 2020.
- [16] F. Gräbe, N. Kannengießer, S. Lins, and A. Sunyaev, “Do not be fooled: Toward a holistic comparison of distributed ledger technology designs,” in *Proceedings of the 53rd Hawaii international conference on system sciences*, 2020, pp. 6297–6306.
- [17] K.-J. Stol and B. Fitzgerald, “The abc of software engineering research,” *ACM Trans. Softw. Eng. Methodol.*, vol. 27, no. 3, Sep. 2018.
- [18] F. v. Normann, “Tezbed - a benchmarking framework for private blockchain systems based on the tezos protocol.” [Online]. Available: [https://github.com/Fredericvn/TezBed\\_A-Benchmarking-Framework-for-Private-Blockchain-Systems-Based-on-the-Tezos-Protocol](https://github.com/Fredericvn/TezBed_A-Benchmarking-Framework-for-Private-Blockchain-Systems-Based-on-the-Tezos-Protocol)