Unrolling the Performance of ZK-Rollups through Stochastic Modeling

Carlos Melo*, Johnnatan Messias†, José Miquéias‡, Glauber Gonçalves‡, Francisco A. Silva‡, and André Soares‡
*Universidade Federal de Sergipe, São Cristóvão - Sergipe, Brazil
†Max Planck Institute for Software Systems (MPI-SWS), Saarbrucken and Kaiserslautern, Germany
‡Universidade Federal do Piauí (UFPI), Picos - PI, Brazil
{carlos.alexandre, jmiqueias, ggoncalves, faps, andre.soares}@ufpi.edu.br, johnme@mpi-sws.org

***Abstract*—Sidechains offer partial solutions to Ethereum's scalability challenges; however, they introduce trade-offs related to security and implementation complexity. These limitations have been further addressed by Layer-2 solutions known as rollups, which combine off-chain computation with on-chain verification, preserving both security and decentralization on the Ethereum platform. This paper proposes a Stochastic Petri Net model to evaluate the feasibility of ZK-Rollups by analyzing their impact on throughput and latency. The results indicate that increased adoption of Layer-2 transactions can enhance system throughput by up to 20%. Conversely, latency may rise by more than 100% when larger batches are used, revealing a fundamental performance trade-off.**

***Index Terms*—Ethereum, Zero Knowledge Proof, Rollups, Blockchain, Performance, and Stochastic Models.**

I. Introduction

Blockchain technology ensures data integrity and security, by minimizing the risk of single points of failure in a distributed system. The introduction of smart contracts has broadened the scope of blockchain applications, enabling automated and trustless execution of agreements and transactions. This evolution has spurred further development and diversification of blockchain technology [1], leading to explorations of various consensus mechanisms [2] and interoperability solutions [3] to address scalability and performance challenges. These advancements have driven the exploration of blockchain's potential in diverse domains such as supply chain management [4], healthcare [5], and digital identity [6].

Despite continuous innovations, public blockchains face performance limitations related to the scalability trilemma. This trilemma highlights the challenge of simultaneously achieving high levels of security, decentralization, and scalability [7]. Usually, most public blockchains exhibit low transaction throughput and high mean response times (MRT), and during periods of high demand, this leads to network congestion and increased fees, hindering broader adoption, especially in computationally intensive domains.

The Ethereum platform addressed scalability limitations by implementing a Proof-of-Stake (PoS) consensus mechanism and sidechains, which operate parallel to the main blockchain (Layer-1), and offer higher transaction throughput. However, many sidechain implementations [8] raise security concerns due to the use of alternative consensus protocols, which may compromise blockchain's security.

In contrast, rollups emerged as an alternative to enhance Ethereum's scalability by processing transactions off-chain and recording only essential data or state changes on Layer-1 [9]. This approach significantly improves transaction throughput while leveraging the security of the underlying Layer-1 blockchain [10]. However, evaluating the performance of Layer-2 solutions presents unique challenges. While empirical methods are commonly used for blockchain performance evaluation, they often struggle to capture the dynamic complexities inherent in rollups. Traditional methods may not adequately address the interplay between on-chain and off-chain processing, the impact of different rollup designs, and the diverse range of potential applications. Existing studies have explored rollup performance using various methods [7], including test networks [11], benchmarking tools [12], [13], and simulations [9]. However, these approaches, while valuable, can be resource-intensive.

This paper leverages a Stochastic Petri Net (SPN) to capture the probabilistic nature of transaction processing and network dynamics. This approach allows for a more realistic representation of real-world blockchain systems compared to deterministic models [14]. Furthermore, our model incorporates empirical data derived from existing studies on rollups, such as the ZKsync dataset [15] and analyses of rollup benchmarks [12]. In summary, the contributions of this paper are as follows:

- We provide a formal analysis of how Layer-2 rollups mitigate the scalability trilemma.
- We introduce a novel application of Stochastic Petri Nets to model transaction dynamics and interactions between Layer-1 and Layer-2 components.
- We present an empirical performance evaluation showing that throughput can increase by up to 20%. However, we also observe that increasing the batch size can lead to over 100% growth in latency.

This paper is structured as follows. Section II reviews related work on blockchain performance modeling and evaluation. Section III provides an overview of Ethereum's transaction flow considering an approach of *Layer-2*. Section IV introduces the proposed model and its applications. Section V presents case studies and evaluation results. Finally, Section VI provides final considerations, limitations, and future work.

II. Related Work

This section positions our work within the context of existing literature, highlighting the distinctions between our approach and those used in other relevant studies.

Researchers have explored blockchain performance using various methods. Schaffer et al. [16] evaluated performance optimizations in private Ethereum infrastructures. While those results provide an overview of manageable blockchain micro-infrastructures, our work focuses on the public Ethereum network and the specific scalability challenges related to rollups. Unlike private blockchains, public networks require a broader assessment and consider factors such as decentralization and transaction-associated costs.

Spain et al. [17] investigated the impact of throughput and latency on Ethereum network transactions, particularly during high-demand periods such as Initial Coin Offerings. Although their work highlights the importance of costs in Ethereum, our research differs by focusing on the performance and cost implications of using ZK-Rollups as a scalability solution. By incorporating the probability of a transaction being processed on Layer-1 or Layer-2, our model provides a more comprehensive understanding of how ZK-Rollups can maintain a cost-benefit balance, considering latency and throughput in relation to infrastructure costs. Also, it is important to mention that the scenery covered by Spain et al. can also be analyzed by a stochastic framework, as proposed by this paper, once the number of requests is a parameter controlled by the arrival rate.

In previous work, Melo et al. [18] established a foundation for the performance evaluation of the Hyperledger Fabric platform, demonstrating the feasibility of using stochastic models to assess blockchains and pointing out the block formation and block size impact on the throughput and latency of permissioned environments. The present study expands this foundation by focusing on public blockchains, particularly Ethereum, incorporating the critical aspect of costs associated with Layer-2 scalability solutions, specifically rollups.

Ernstberger et al. [13] evaluated the performance of rollups through the zk-Bench tool, which helps evaluate the performance of different rollup architectures. However, our work focuses on developing a generalizable stochastic model to assess the performance of rollups, whereas zk-Bench aims at micro-benchmarking rollups. Our approach seeks to capture the performance dynamics of ZK-Rollups within the broader context of blockchain networks, considering transaction flow, costs, resource utilization, and network conditions.

Chaliasos et al. [12] analyzed the benchmarking and costs of ZK-Rollups, providing valuable data about throughput and transaction fees on specific implementations and specific Google Cloud infrastructures. Our paper takes a different approach by constructing stochastic models that enable a generalized performance evaluation for different types of ZK-Rollups. It considers the influence of costs on the choice between Layer-1 and Layer-2 on the Ethereum network, offering a broader and more flexible perspective on optimizing ZK-Rollup performance. The performance and infrastructure data extracted from Chaliasos et al. will be used to feed and evaluate the model proposed in our study.

Finally, Silva et al. [15] compiled a public dataset for the ZKsync Era rollup, providing real-world data about transaction throughput and latency that enables a better understanding of the performance characteristics of ZK-Rollups. While our work also focuses on evaluating Ethereum's performance with ZK-Rollups, we adopt a different approach by employing stochastic models that can be fed with data from Silva et al.. This allows us to explore a range of scenarios through a set of adjustable parameters based on system requirements.

III. Theoretical Designs

This section presents the foundational concepts required to understand Ethereum's transaction flow, and the interplay between Layer-1 and Layer-2 solutions. Ethereum enables smart contracts execution and the development of decentralized applications (dApps). Figure 1 provides Ethereum's transaction flow highlighting the pathways through Layer-1 and Layer-2.

Usually, a transaction is initiated by a client that may transfer assets between accounts, or a smart contract may be triggered based on some premises. These transactions are placed in the mempool, a temporary storage area, and soon they are propagated to processing on Layer-1 or Layer-2.

Layer-2 solutions improve scalability by aggregating and verifying transactions off-chain. For transactions processed on Layer-1, validators verify and group transactions into blocks, which are subsequently appended to the Ethereum blockchain, ensuring their permanence and immutability. The following describes the key components depicted in Figure 1:

- **Batch:** In Layer-2 solutions, multiple transactions are bundled off-chain into a batch. This process significantly reduces gas fees and bottlenecks on Layer-1 [9].
- **Block:** A collection of validated transactions. In Layer-2 systems, once a batch is validated, its result is submitted to Layer-1 as a single transaction and included in a block.
- **Batch Processing Time:** The average time required to append a batch to a block [8].
- **Block Generation Time:** The average time required for the network to generate a new block [8].
- **Processing Times:** The time required to process and validate transactions on both Layers [8].
- **Block Confirmation Time:** The time required for a block to be added to the blockchain and achieve finality, ensuring immutability.

This study focuses on performance metrics associated to a specific type of rollup named **ZK-Rollups** that employs Zero-Knowledge Proofs (ZKPs) for immediate validation. In ZK-Rollups, each batch is accompanied by a cryptographic proof that confirms its validity. This proof is generated off-chain

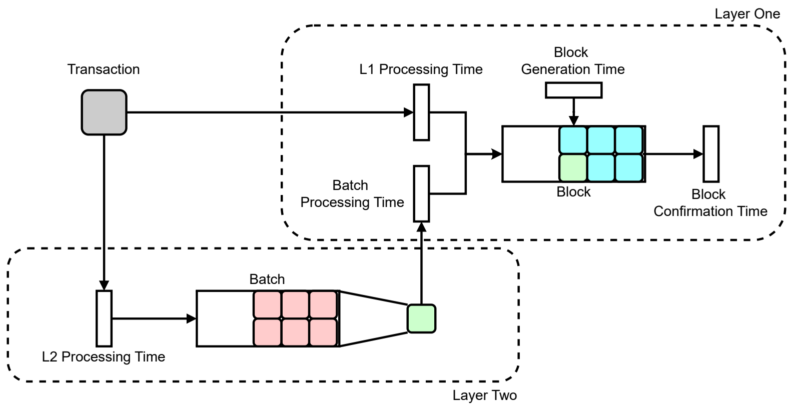


Fig. 1: Ethereum's Transaction Flow

and submitted to Layer-1 for verification. Since the proof guarantees batch validity, there is no need for a challenge period. This enhances security and transaction confirmation times.

IV. Proposed Model

The proposed model, depicted in Figure 2, is a Stochastic Petri Net (SPN) designed to formally represent the Ethereum platform. It focuses on the interactions between Layer-1 and Layer-2 operations by capturing the transaction flow on both layers. Table I details the server semantics and transition properties that define the model's behavior. The model representation and numerical analysis were performed using the Mercury tool [19].

TABLE I: Server Semantics in the Proposed Model

Transition	Type	Semantics	Distribution	Prob.
TE0	Timed	Single Server	Poisson	-
TE1	Timed	Infinite Server	Exponential	-
TE2	Timed	Single Server	Exponential	-
TE3	Timed	Infinite Server	Exponential	-
TE4	Timed	Infinite Server	Exponential	-
TE5	Timed	Single Server	Exponential	-
TE6	Timed	Single Server	Exponential	-
TI0	Immediate	Infinite Server	-	90%
TI1	Immediate	Infinite Server	-	10%
TI2	Immediate	Infinite Server	-	-

The proposed model can be interpreted sequentially, starting from the top-left corner and progressing to the right. Transactions enter the system through transition **TE0**, which represents transactions inter-arrival time.

A transaction reaches the intermediate state **P0**, from which it can be routed to either Layer-1 (blue dashed lines) or Layer-2 (red dashed lines). A key feature of our model is the probabilistic routing mechanism composed of the weighted immediate transitions **TI0** and **TI1**, which redirect transactions to Layer-1 and Layer-2, respectively. The weights associated with **TI0** and **TI1** define the probabilities for a transaction to be processed on each layer.

A. Layer-1 Path

Upon reaching state **P6**, a transaction undergoes a transaction validation process (transition **TE4**). The time associated with this transition reflects the average transaction processing time on Layer-1. After validation, the transaction moves to state **P7**, where it waits to be included in a block via transition **TI2**.

The Layer-1 block size is represented in the model as a queue composed of places **P8** and **P9**. Transactions arriving at the block through transition **TI2** remain in this queue until the block reaches its designated size. When place **P8** accumulates a number of tokens equal to the "Block Size", place **P9** is emptied, indicating that the block is complete.

At this point, all transactions in the block are consolidated into a single token and directed through the exponential transition **TE5**. The time of this transition corresponds to the block generation time. As a new block is generated, the previous block is moved to place **P10**, where it awaits persistence on the blockchain. This step is represented by the exponential transition **TE6**, whose time reflects the time required for block finality.

B. Layer-2 Path

When transactions are routed to Layer-2, a limited amount of resources is assumed. This limitation is represented in the model by a queue composed of places **P1** and **P2**, symbolizing the server capacity related to the number of vCPUs available. This configuration differs from Layer-1, where the size of the mempool or validation resources is abstracted as infinite.

Upon arrival at Layer-2, transactions undergo a quick evaluation before being directed to a queue representing a batch. This batch formation mechanism is analogous to block formation in Layer-1, with the transition to the batch queue triggered by transition **TE1**.

The batch size depends on the Layer-2 solution employed. Larger batches accommodate more transactions and require more time to reach their full capacity. Consequently, transactions in larger batches experience longer finalization times.

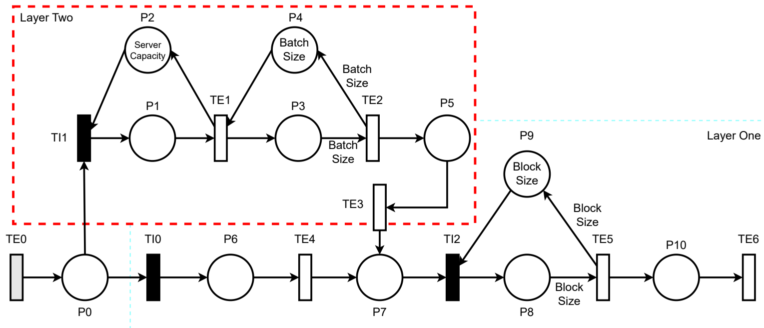


Fig. 2: Stochastic Petri Net for the Ethereum Platform

Conversely, smaller batches are submitted to Layer-1 more frequently, resulting in higher gas costs for users.

Once full, the batch is represented as a single token and transferred to place **P5** via transition **TE2**. The time of this transition reflects the time required to consolidate all transactions in the batch. Subsequently, the batch is submitted to Layer-1 as a single transaction, accompanied by the necessary proof for validation. This step is represented by transition **TE3**, whose time corresponds to proof verification or batch processing. After reaching Layer-1, the batch undergoes the same steps as transactions sent directly to Layer-1.

V. Case Studies

This section presents three case studies demonstrating the feasibility of the proposed model. The first case study provides a general evaluation of the system, considering both layers and analyzing the impact on throughput (transactions per second) and latency (seconds). The second case study shows the effect of Layer-2 components on general throughput and latency and the Layer-2 throughput through a $2k$ factorial design. The third case study investigates how processing time and batch size influence the throughput of the submodel representing *Layer-2*.

Table II presents the input data for evaluating the proposed model using the Mercury tool. The Layer-1 values reflect real-world Ethereum data extracted from etherscan.io, while the Layer-2 data were adjusted to represent ZK-Rollup characteristics and obtained from [12] and [15]. Additionally, most of these configurations are inherent to the Ethereum platform itself, such as block size and transaction arrival. Consequently, the minimum and maximum values for these components are defined by their base values.

A. Case Study I - Key Metrics and Layer-1

The first case study evaluates system latency and throughput, focusing on the relationship between both layers until transactions persist. Latency establishes a relationship between the average number of pending transactions in a system and the inter-arrival time of transactions. Figure 3a illustrates how system latency increases as the probability of transactions being processed by Layer-2 grows.

TABLE II: Input Parameters for the Proposed Model

Factor	Baseline	Variation {min,max}
Arrival (TE0)	≈13.7tps	{-}
L2 Probability	90%	{0%, 100%}
L1 Probability	10%	{0%, 100%}
Layer One Process. Time (TE4)	≈12.98s	{-}
Layer Two Process. Time (TE1)	100ms	{50ms, 200ms}
Block Generation Time (TE5)	13s	{-}
Block Confirm. Time (TE6)	60s	{-}
Batch Process. Time (TE2)	1075s	{600s, 1800s}
Batch Generation Time (TE3)	1s	{0.5s, 1.5s}
Batch Size	5000	{1000, 10000}
Server Capacity	32	{16, 64}
Block Size	167	{-}

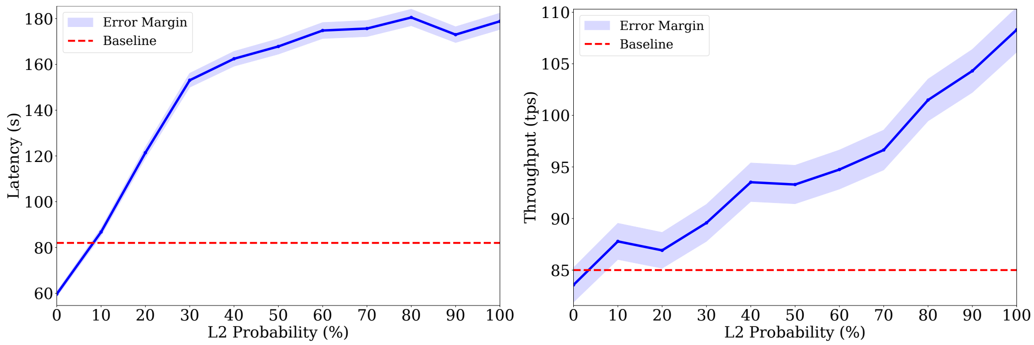
Throughput, defined as the number of transactions successfully processed per unit of time [14], is calculated separately for Layer-1 and Layer-2 in the proposed model. Figure 3b shows the increase in overall system throughput as the probability of transactions using Layer-2 grows. Higher throughput is desirable, while higher latency is undesirable. This relationship is critical from the user's perspective, as users expect their transactions to persist on the blockchain as quickly as possible. The parameters defined for our Layer-2 evaluation demonstrate a positive impact on system throughput. Throughput increases by approximately 20% in the evaluated scenario, reaching around 110 tps when 90% of transactions use Layer-2, compared to 85 tps when 90% of transactions use Layer-1.

B. Case Study II - Design of Experiments

This case study employs a 2^k factorial design to evaluate the systematic impact of key parameters on system performance. We focus on Layer-2 parameters, such as batch size, server capacity, and processing time, to understand their effects on system throughput and latency.

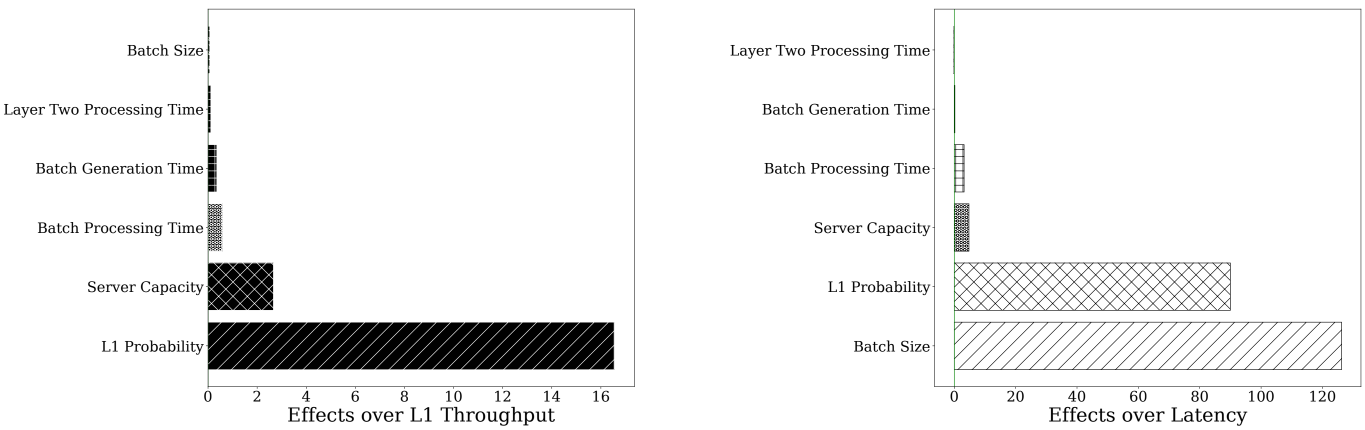
Figure 4a shows the impact of each Layer-2 component on the overall system throughput. The probability of a transaction using Layer-1 significantly impacts the overall system throughput, with variations of up to 16tps. Server capacity also influences this metric, albeit to a lesser extent, with an impact of just over 2tps. The remaining parameters show effects of less than 1tps over this metric.

Figure 4b highlights the substantial impact of batch size on latency, increasing it by more than 120 seconds. Additionally,



(a) Negative impact of Layer-2 on latency (b) Positive impact of Layer-2 on general throughput

Fig. 3: Latency and Throughput vs. Probability of Transactions Using Layer-2



(a) Impact of the probability of using Layer-1 on throughput (b) Impact of batch size and Layer-1 probability on latency

Fig. 4: Impact of Layer-2 parameters on Layer-1 performance

the probability of using Layer-1 also exerts a considerable influence on this metric. Finally, Figure 5 demonstrates that batch processing time has a significant negative effect on Layer-2 throughput, reducing it by more than 6tps.

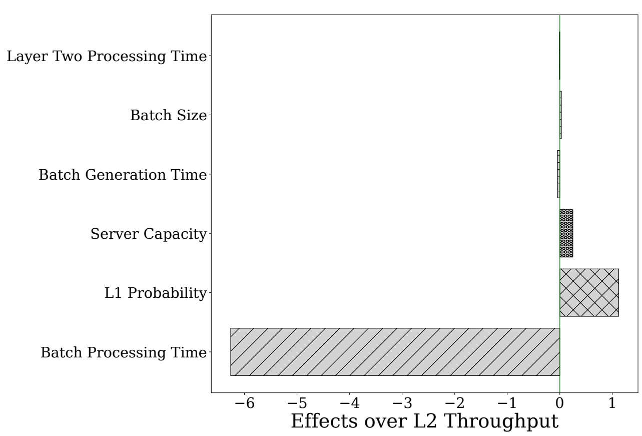


Fig. 5: High impact of batch processing time on Layer-2

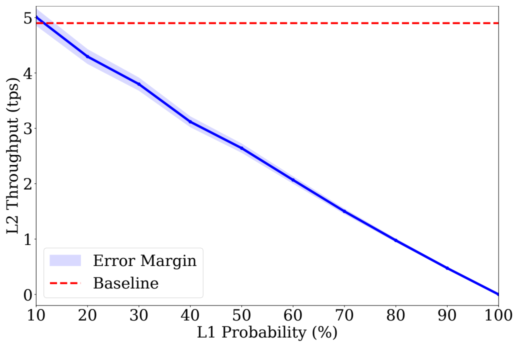
C. Case Study III - Exploring Layer-2

The throughput of the Layer-2 subsystem increases as the probability of transactions following this layer grows, as illustrated in Figure 6a. However, this growth of approximately 20% is modest compared to the negative impact caused by another factor directly related to batches: processing time. The larger the batch, the longer the processing time and, consequently, the lower the system throughput.

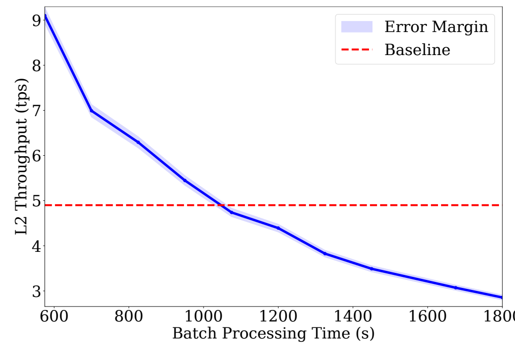
In the baseline scenario, with a batch processing time of 1075 seconds, the throughput was 4.9tps. However, it dropped to less than 3tps in scenarios with a batch processing time of 1800 seconds. Conversely, a batch processing time of 600 seconds nearly doubles Layer-2 throughput. In other words, smaller batches result in higher throughput, as evidenced in Figure 6b. However, this improvement in throughput comes at a higher cost for the user.

VI. Conclusion

This paper addressed scalability in blockchains, focusing on ZK-Rollups as a Layer-2 solution to overcome performance limitations in public blockchains like Ethereum. While Ethereum's transition to Proof-of-Stake and the emergence of sidechains have mitigated some performance bottlenecks,



(a) Lower transaction volume in Layer-2 vs. throughput



(b) Higher batch processing time reduces throughput

Fig. 6: Layer-2 Throughput vs. Probability of Layer-2 and Batch Processing Time

these approaches often involve trade-offs related to security or introduce implementation complexities. In contrast, ZK-Rollups offer a distinct advantage by leveraging off-chain computation and on-chain verification in compliance with Ethereum's requirements, increasing transaction throughput without compromising security. This directly demonstrate how rollups address the scalability trilemma by enhancing efficiency while preserving decentralization and security guarantees.

Additionally, this paper introduced a generalizable performance model based on Stochastic Petri Nets, enabling the simulation of network conditions and workloads through parameterized variables. This framework provides a comprehensive understanding of ZK-Rollup behavior and how it improves Ethereum performance in terms of throughput and latency. By identifying key performance factorssuch as batch sizes, processing times the results clarify how rollup design choices interact with Layer-1 and influence efficiency. For future work, we propose experimental validation of the model, focusing on ZK-Rollups and generalizing it to other Layer-2 technologies, such as Optimistic Rollups.

VII. Acknowledgments

This work had CAPES financial support under process No. 88887.173882/2025-00.

References

[1] X. Zhu, J. Shi, F. Xie, and R. Song, "Pricing strategy and system performance in a cloud-based manufacturing system built on blockchain technology," *Journal of Intelligent Manufacturing*, pp. 1–18, 2020.

[2] D. Huang, X. Ma, and S. Zhang, "Performance analysis of the raft consensus algorithm for private blockchains," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 1, pp. 172–181, 2019.

[3] R. Belchior, A. Vasconcelos, S. Guerreiro, and M. Correia, "A survey on blockchain interoperability: Past, present, and future trends," *ACM Computing Surveys (CSUR)*, vol. 54, no. 8, pp. 1–41, 2021.

[4] M. M. Queiroz, R. Telles, and S. H. Bonilla, "Blockchain and supply chain management integration: a systematic review of the literature," *Supply chain management: An international journal*, vol. 25, no. 2, pp. 241–254, 2020.

[5] A. Haleem, M. Javaid, R. P. Singh, R. Suman, and S. Rab, "Blockchain technology applications in healthcare: An overview," *International Journal of Intelligent Networks*, vol. 2, pp. 130–139, 2021.

[6] C. Sullivan and E. Burger, "Blockchain, digital identity, e-government," in *Business Transformation through Blockchain: Volume II*. Springer, 2018, pp. 233–258.

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

[8] D. Yang, C. Long, H. Xu, and S. Peng, "A review on scalability of blockchain," in *Proceedings of the 2020 2nd International Conference on Blockchain Technology*, ser. ICBCT '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 16. [Online]. Available: https://doi.org/10.1145/3390566.3391665

[9] L. T. Thibault, T. Sarry, and A. S. Hafid, "Blockchain scaling using rollups: A comprehensive survey," *IEEE Access*, vol. 10, pp. 93 039–93 054, 2022.

[10] L. D. Negka and G. P. Spathoulas, "Blockchain state channels: A state of the art," *IEEE Access*, vol. 9, pp. 160 277–160 298, 2021.

[11] A. Gluchowski, "Evaluating ethereum l2 scaling solutions: A comparison framework," *Accessed: Apr*, vol. 18, p. 2022, 2020.

[12] S. Chaliasos, I. Reif, A. Torralba-Agell, J. Ernstberger, A. Kattis, and B. Livshits, "Analyzing and benchmarking zk-rollups," *Cryptology ePrint Archive*, 2024.

[13] J. Ernstberger, S. Chaliasos, G. Kadianakis, S. Steinhorst, P. Jovanovic, A. Gervais, B. Livshits, and M. Orrù, "zk-bench: A toolset for comparative evaluation and performance benchmarking of snarks," in *International Conference on Security and Cryptography for Networks*. Springer, 2024, pp. 46–72.

[14] P. R. M. Maciel, *Performance, reliability, and availability evaluation of computational systems, volume I: performance and background*. CRC Press, 2023.

[15] M. I. Silva, J. Messias, and B. Livshits, "A public dataset for the zksync rollup," *arXiv preprint arXiv:2407.18699*, 2024.

[16] M. Schäffer, M. Di Angelo, and G. Salzer, "Performance and scalability of private ethereum blockchains," in *Business Process Management: Blockchain and Central and Eastern Europe Forum: BPM 2019 Blockchain and CEE Forum, Vienna, Austria, September 1–6, 2019, Proceedings 17*. Springer, 2019, pp. 103–118.

[17] M. Spain, S. Foley, and V. Gramoli, "The impact of ethereum throughput and fees on transaction latency during icos," in *International Conference on Blockchain Economics, Security and Protocols (Tokenomics 2019)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2020, pp. 1–15.

[18] C. Melo, G. Gonçalves, F. A. Silva, and A. Soares, "Performance modeling and evaluation of hyperledger fabric: An analysis based on transaction flow and endorsement policies," in *2024 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2024, pp. 1–6.

[19] P. Maciel, R. Matos, B. Silva, J. Figueiredo, D. Oliveira, I. Fé, R. Maciel, and J. Dantas, "Mercury: Performance and dependability evaluation of systems with exponential, expolynomial, and general distributions," in *Proc. of PRDC*. IEEE, 2017, pp. 50–57.