Public vs Private Blockchains lineage storage

Bilel Zaghdoudi
LIP6, CNRS
Sorbonne University
Paris, France

Maria Potop Butucaru
LIP6, CNRS
Sorbonne University
Paris, France

***Abstract*—This paper reports the experimental results related to lineage event storage via smart-contracts deployed on private and public blockchain. In our experiments we measure the following three metrics: the cost to deploy the storage smart contract on the blockchain, which measures the initial expenditure, typically in gas units, required to deploy the smart contract that facilitates lineage event storage, then the time and gas costs needed to store a lineage event. We investigated both single and multi-clients scenarios. We considered the following public blockchains: Hedera, Fantom, Harmony Shard0, Polygon Amoy, Ethereum Sepolia, Optimism Sepolia, Klaytn Baobab and Arbitrum Sepolia. Furthermore, we investigate the performances of Hyperledger Besu with different consensus algorithms as private blockchains.**

***Index Terms*—secure lineage storage, public blockchains, private blockchains, gas and time cost**

I. Introduction

The rapid proliferation of data across various domains has led to increasingly complex data ecosystems, characterized by diverse data sources, intricate data transformations, and multifaceted data consumption patterns. This complexity poses significant challenges in ensuring data quality, traceability, and compliance. Organizations struggle to maintain a comprehensive understanding of their data assets, leading to difficulties in identifying data lineage, ensuring data integrity, and meeting regulatory requirements. Moreover, the absence of standardized metadata management and data governance practices exacerbates the issue, resulting in data silos, inconsistent data definitions, and a lack of transparency in data operations. To address these challenges, tools and platforms such as OpenLineage, Apache Atlas, DataHub, and others have emerged, offering robust solutions for data lineage, metadata management, and data governance. These tools enable organizations to track and visualize data flow across systems, ensure data quality and consistency, and comply with regulatory standards by providing a unified view of data assets and their interdependencies. By implementing these tools, organizations can enhance data observability, streamline data governance processes, and foster a culture of data-driven decision-making.

The intricate nature of modern data ecosystems, while enabling enhanced data analytics and insights, also introduces significant security vulnerabilities. As data traverses through various systems and processes, it becomes susceptible to unauthorized access, data breaches, and malicious alterations. The decentralized architecture of contemporary data platforms, combined with the extensive integration of third-party services and cloud-based infrastructures, amplifies the risk of security incidents. These vulnerabilities can lead to substantial financial losses, reputational damage, and legal repercussions. Furthermore, the lack of comprehensive visibility and control over data movement complicates the enforcement of security policies and compliance with data protection regulations such as GDPR, CCPA, and HIPAA. Tools and platforms designed for data lineage, metadata management, and data governance must, therefore, incorporate robust security measures to safeguard data integrity and confidentiality. Ensuring secure data practices is imperative not only to protect sensitive information but also to maintain trust and compliance in an increasingly data-driven world. The integration of advanced security features, such as encryption, access controls, and continuous monitoring, within these tools can mitigate risks and enhance the overall security posture of an organization's data ecosystem.

II. Experimental Study

In this experimentation, three key metrics are being utilized to evaluate the performance and efficiency of various blockchain platforms for lineage event storage. The first metric is the cost to deploy the storage smart contract on the blockchain, which measures the initial expenditure, typically in gas units, required to deploy the smart contract that facilitates lineage event storage. Gas is a standard measure of computational work in blockchain systems. This cost is critical as it represents the upfront investment necessary to set up the system on the blockchain. The second metric is the time needed to store a lineage event on the blockchain, which gauges the latency or delay involved in recording a lineage event once the storage contract is operational. This metric is essential for understanding the responsiveness and efficiency of the blockchain in handling real-time data entries. Finally, the cost to store a lineage event on the blockchain evaluates the ongoing expense incurred each time a lineage event is recorded. This cost is typically expressed in gas units and reflects the economic feasibility of using the blockchain for continuous data storage operations. Together, these metrics provide a comprehensive view of the deployment and operational costs, as well as the performance efficiency of different blockchain platforms in supporting lineage event storage. For the following experimentation scenarios, the deployed smart contract had a size of 6,032 bytes. Additionally, the average size of each storage event recorded was approximately 2,200 bytes.

A. *Single-Batch Execution Scenario*

In this scenario, we execute 100 transactions sequentially on each blockchain system. The blockchain platforms evaluated include Hedera Testnet [1], Fantom Testnet [2], Harmony Shard0 Testnet [3], Polygon Amoy Testnet [4], Ethereum Sepolia Testnet [5], Optimism Sepolia Testnet [9], Klaytn Baobab Testnet [6] and Arbitrum Sepolia Testnet [8] for public Blockchains and Hyperledger Besu [7] with different consensus algorithms for private networks. The hardware environment for the private networks consists of 3 virtual machines with 4vCPUs and 8 GB RAM each, running Ubuntu 22.04. Performance metrics are collected using custom monitoring scripts and analyzed using statistical methods to determine significance.

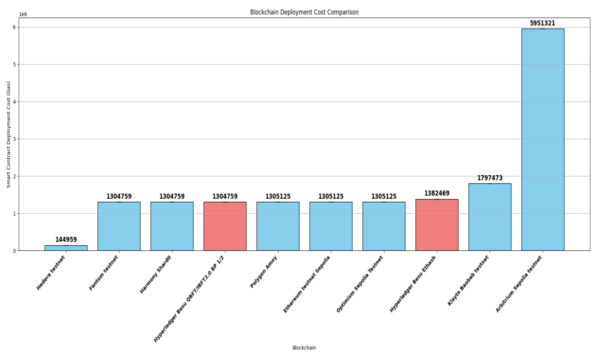

Fig. 1. Smart contract deployment cost

1) Data Presentation, analysis and observations: Smart contract deployment costs further underscore the cost efficiency of private blockchains. Hyperledger Besu configurations (IBFT2.0 and QBFT) maintain deployment costs at approximately 1,304,759 gas units, while the Hedera Testnet shows an exceptionally low cost of 144,959 gas units. In stark contrast, public blockchains like Arbitrum Sepolia Testnet and Ethereum Testnet Sepolia incur much higher costs, with Arbitrum Sepolia reaching 5,951,321 gas units. This significant variation underscores the economic considerations that organizations must evaluate when selecting a blockchain platform for deploying smart contracts.

The lineage event storage time comparison shows a similar trend, with private blockchains outperforming public ones. Arbitrum Sepolia Testnet stands out with the lowest storage time of approximately 814.65 ms. Hyperledger Besu configurations also demonstrate low storage times, with IBFT2.0 BP 1 achieving around 1,081.75 ms. Conversely, public blockchains like Ethereum Testnet Sepolia and Hedera Testnet exhibit higher storage times, around 13,761.73 ms and 7,240.37 ms, respectively. These results suggest that while public blockchains offer broader accessibility, they may face latency issues, impacting their suitability for applications

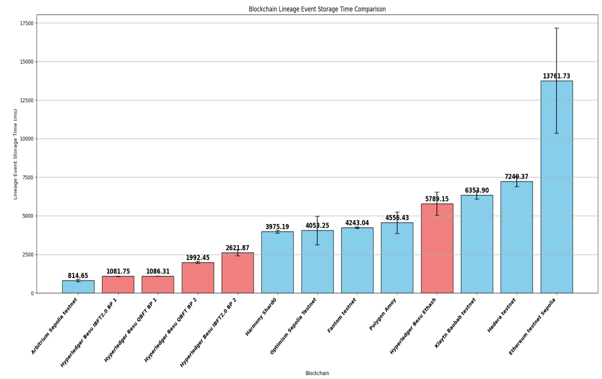

Fig. 2. Lineage Event Storage Time

requiring rapid data storage. The chart in 3 illustrating the

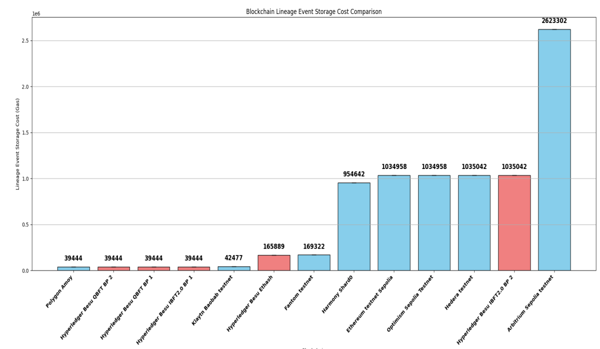

Fig. 3. Lineage Event Storage Cost

lineage event storage cost across different blockchains reveals significant disparities. The private blockchains (Hyperledger Besu) demonstrate remarkably low storage costs, around 39,444 gas units across various configurations. This indicates efficient performance and cost-effectiveness in a controlled environment. On the other hand, public blockchains, such as Arbitrum Sepolia, incur significantly higher costs, peaking at 2,623,302 gas units. Notably, Ethereum testnet Sepolia and Optimism Sepolia Testnet also exhibit high costs, around 1,035,042 gas units. This discrepancy highlights the cost implications of leveraging public blockchains for storage-intensive operations, primarily due to their broader network and higher operational complexities.

Private blockchains (Hyperledger Besu configurations) consistently demonstrate lower costs and faster storage times, highlighting their efficiency in controlled environments. This efficiency makes them preferable for enterprise applications where cost and performance are critical. Public blockchains, despite offering greater decentralization and security, exhibit higher costs and slower performance. These attributes might be a trade-off for their broader network accessibility and robustness against single points of failure.

The deployment and operational costs on public blockchains can be prohibitively high. For instance, the cost on the Arbitrum Sepolia Testnet is significantly higher than on private blockchains, which may deter cost-sensitive applications. Organizations need to balance the benefits of decentralization offered by public blockchains against the economic efficiencies of private blockchains. Performance Metrics:

The performance metrics indicate that for applications requiring high throughput and low latency, private blockchains provide a more suitable infrastructure. The higher storage times on public blockchains suggest potential latency issues, which can impact real-time applications and services.

In conclusion, the analysis indicates that while public blockchains offer the advantages of decentralization and security, they come with higher costs and slower performance compared to private blockchains. Private blockchains like Hyperledger Besu present a compelling case for applications where cost efficiency and performance are paramount. Organizations must carefully consider these factors when choosing a blockchain platform, weighing the benefits of decentralization against the need for cost-effective and performant solutions.

B. Multi-Worker Execution Scenario

In this scenario, we studied lineage event storage using Hyperledger Besu private Blockchain. Three different consensus algorithms were used in this experimentation respectively Ethash, QBFT and IBFT2.0 with a block period of 1 and then 2 seconds. We use multiple workers to execute 100 transactions each, simultaneously. We test configurations with 4, 8, 12, and 16 workers.

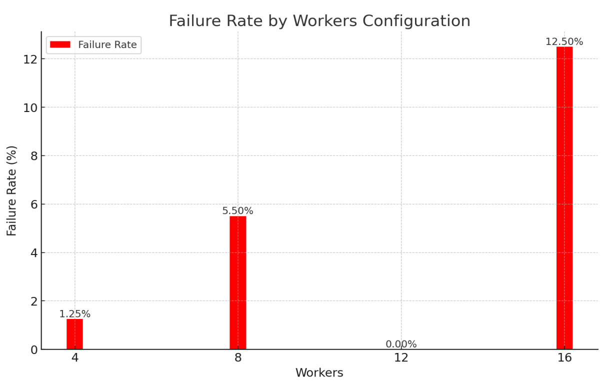

Fig. 4. Ethash Failure Rate Transactions

1) Ethash Data Presentation, analysis and observations: The chart presented in 4 illustrates the failure rates of transactions as a function of the number of workers used in the blockchain network. Failure rate, expressed as a percentage, is plotted against the different configurations of workers (4, 8, 12, and 16). This analysis is crucial for understanding the reliability and robustness of the blockchain network under varying loads. The data shows a clear trend of increasing failure rates with a higher number of workers, except for the configuration with 12 workers, which interestingly reports a 0% failure rate. With 4 workers, the failure rate is relatively low at 1.25%. However, when the number of workers increases to 8, the failure rate rises significantly to 5.50%. The configuration with 16 workers exhibits the highest failure rate at 12.50%, indicating substantial reliability issues under heavy load.

The following charts provide a detailed view of the lineage event storage time across the Hyperledger Besu Ethash network with varying numbers of workers. The lineage event storage time, measured in milliseconds (ms), is plotted against the number of transactions for each worker configuration. This analysis helps in understanding how the storage time varies with the load on the blockchain network.

Fig. 5. Ethash Lineage Storage Time Variation with 4 workers

In the first chart, which depicts the performance with 4 workers, there is noticeable variability in storage times, with some transactions experiencing significant delays. The average median time is around 5000 ms, but several transactions exceed this, indicating periods of high latency. As we increase the number of workers to 8, 12, and finally 16, the distribution of storage times shows increased dispersion. This suggests that the system experiences more significant delays as the number of concurrent transactions increases, possibly due to network congestion or resource contention. There is a noticeable

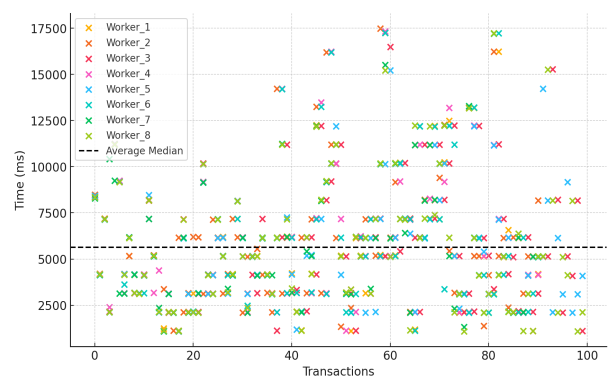

Fig. 6. Ethash Lineage Storage Time Variation with 8 workers

increase in the failure rate as the number of workers increases from 4 to 16, suggesting that higher concurrency levels stress the network, leading to more frequent transaction failures. Across all configurations, there is considerable variability in

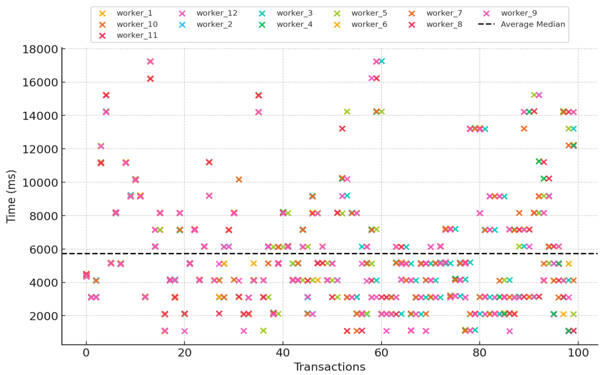

Fig. 7. Ethash Lineage Storage Time Variation with 12 workers

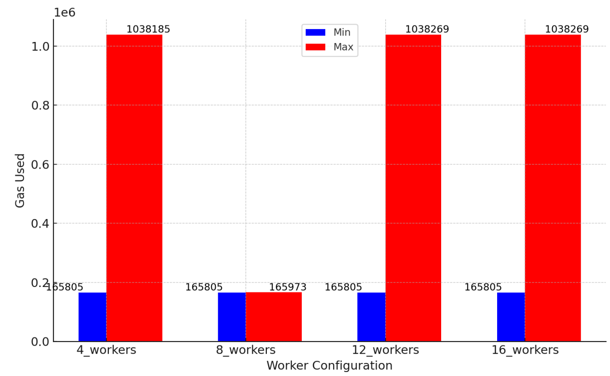

Fig. 9. Ethash Lineage Storage Cost

the storage times. This variability increases with the number of workers, suggesting that higher concurrency levels introduce more significant delays. Average Median Time: The average median storage time remains relatively stable around the 5000 ms mark across different worker configurations. However, the presence of outliers indicates that some transactions are disproportionately delayed. As the number of workers increases from 4 to 16, the frequency of high-latency transactions also increases. This trend highlights the scalability challenges faced by the blockchain network when handling higher transaction volumes. The charts reveal several high-latency outliers, particularly in configurations with more workers. These outliers may result from temporary network bottlenecks, peak usage periods, or inefficiencies in the consensus mechanism under high load.

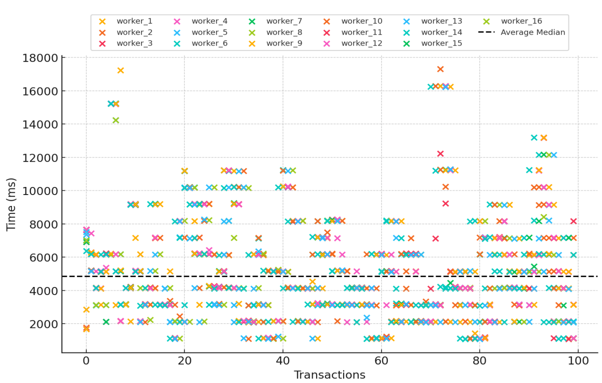

Fig. 8. Ethash Lineage Storage Time Variation with 16 workers

2) IBFT2.0 Data Presentation, analysis and observations: The charts presented provide a comprehensive view of the lineage event storage time across an IBFT (Istanbul Byzantine Fault Tolerant) blockchain with different worker configurations (4, 8, 12, and 16 workers) and for two different setups (BP1 and BP2). The lineage event storage time, measured in milliseconds (ms), is plotted against the number of transactions for each worker configuration. This analysis aims to elucidate the performance variations in storage time under varying loads and worker configurations.

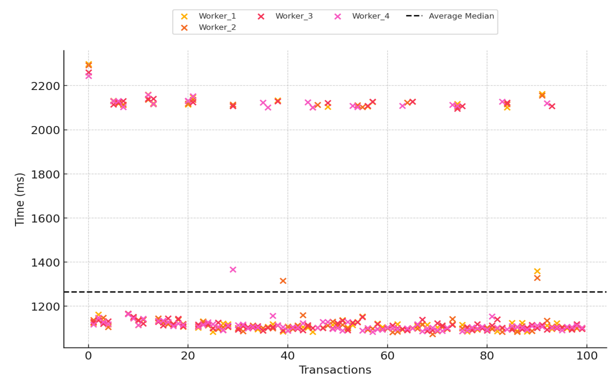

Fig. 10. IBFT2.0 Lineage Storage Time Variation with 4 workers/1s Block Period

In the first chart, showing the performance with 4 workers for BP1, there is a consistent clustering of storage times around the median of approximately 1200 ms. Some outliers do occur, reaching up to 2200 ms, indicating occasional latency spikes. As the number of workers increases to 8, 12, and 16 for both BP1 and BP2, a pattern emerges where the majority of transactions maintain relatively low storage times, but the presence of outliers becomes more frequent and pronounced, especially in the higher worker configurations.

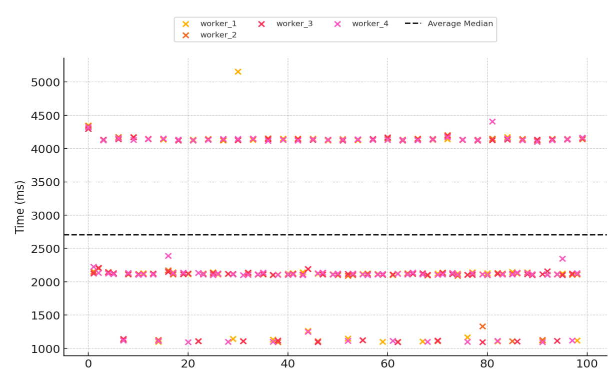

Fig. 11. IBFT2.0 Lineage Storage Time Variation with 4 workers/2s Block Period

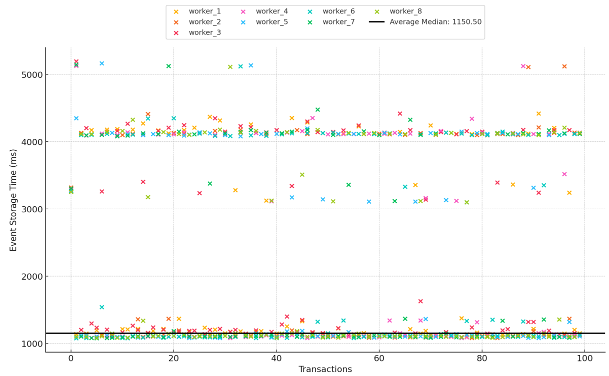

Fig. 12. IBFT2.0 Lineage Storage Time Variation with 8 workers/1s Block Period

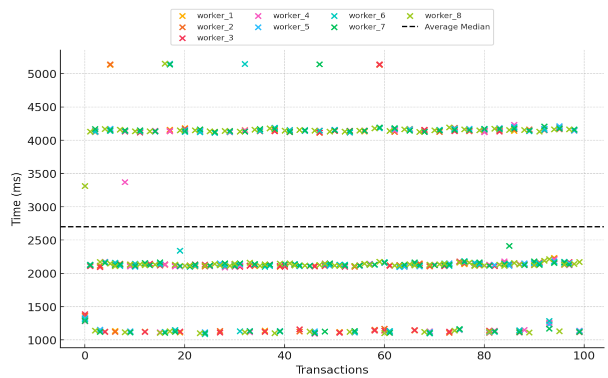

Fig. 13. IBFT2.0 Lineage Storage Time Variation with 8 workers/2s Block Period

Fig. 14. IBFT2.0 Lineage Storage Time Variation with 12 workers/1s Block Period

The chart presented in 18 provides a comparison of the gas usage for different worker configurations (4, 8, 12, and 16 workers) on an IBFT2.0 blockchain platform. The gas usage is a critical metric for evaluating the cost efficiency of transactions on the blockchain. The chart displays the minimum (blue) and maximum (red) gas used for each worker configuration, providing insights into the variability and consistency of gas consumption under different loads.

For the 4-worker configuration, the minimum gas usage is recorded at 165,805 units, while the maximum is substantially higher at 1,038,185 units. This significant disparity indicates high variability in gas consumption. As the number of workers increases to 8, the minimum gas usage drops to 39,360 units, yet the maximum remains consistently high at 1,038,185 units. Similar trends are observed for the 12 and 16 worker configurations, with minimum gas usages at 165,805 and maximums at 1,038,269 units and 1,038,017 units respectively. Across

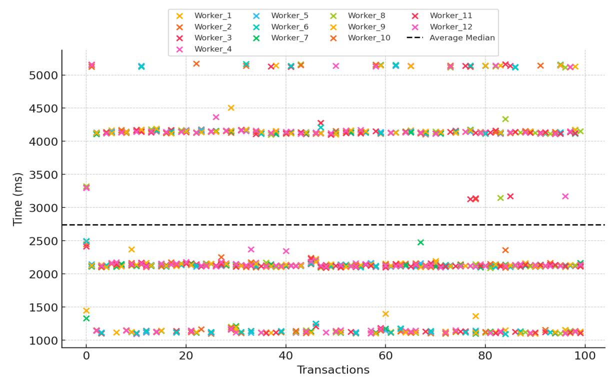

Fig. 15. IBFT2.0 Lineage Storage Time Variation with 12 workers/2s Block Period

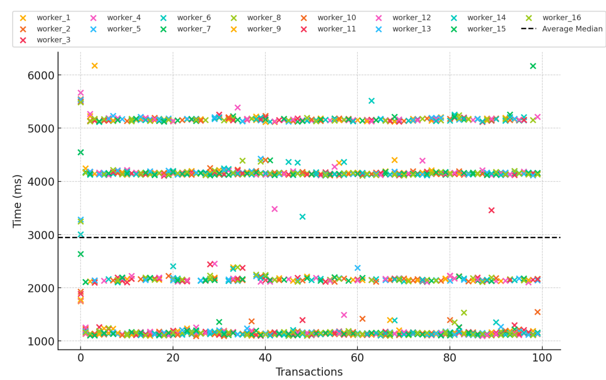

Fig. 16. IBFT2.0 Lineage Storage Time Variation with 16 workers/1s Block Period

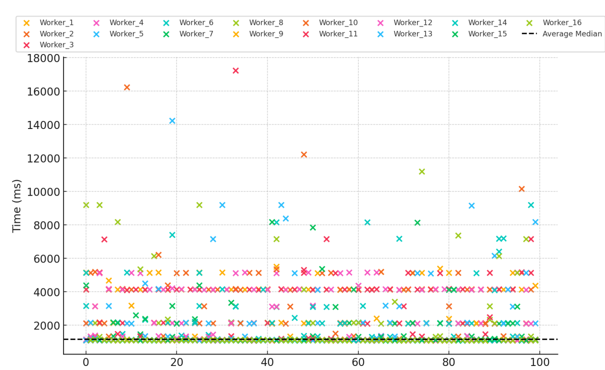

Fig. 17. IBFT2.0 Lineage Storage Time Variation with 16 workers/2s Block Period

all worker configurations and both setups (BP1 and BP2), the median storage time remains relatively stable, hovering around 1200 ms to 2500 ms. This consistency suggests that the IBFT blockchain maintains a predictable performance baseline even as the load increases. Increasing the number of workers from 4 to 16 generally does not significantly degrade the median performance, but it does lead to a higher number of outliers. This indicates that while the system can handle increased concurrency without major degradation in average performance, it occasionally experiences significant delays. The occurrence of outliers is more pronounced in higher worker configurations. These outliers may result from network congestion, resource contention, or inefficiencies in the consensus mechanism under high load. Notably, these spikes can reach up to 18000 ms, especially in configurations with 16 workers. Both BP1 and BP2 setups exhibit similar trends in performance, suggesting that the underlying consensus mechanism handles increased worker loads in a comparable manner across different configurations. Across all worker configurations, the maximum gas usage remains consistently high, slightly above 1,038,000 units. This consistency indicates that certain transactions incur significant gas costs regardless of the number of concurrent workers.

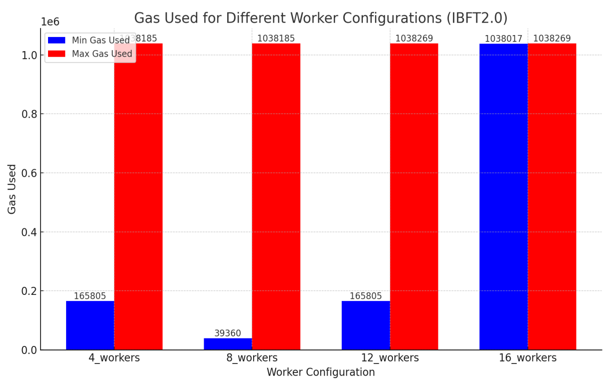

Fig. 18. IBFT2.0 Lineage Storage Cost

In conclusion, the IBFT blockchain platforms demonstrate robust performance with consistent median storage times across varying worker configurations. However, the presence of significant outliers in higher worker configurations highlights potential scalability challenges. The analysis of gas usage reveals a pattern of consistent high maximum gas usage and variable minimum gas usage. This suggests that while the blockchain can handle transactions efficiently under certain conditions, it also experiences periods of high gas consumption, potentially due to network congestion or the complexity of specific transactions.

3) QBFT Data Presentation, analysis and observations: The charts provided display the storage times for lineage events on the QBFT configuration with different worker setups (4, 8, 12, and 16 workers). Each worker configuration involves executing 100 transactions, and the times are measured in milliseconds (ms).

For the QBFT configuration, we observe a relatively consistent performance across varying numbers of workers. The median storage times remain stable, suggesting that QBFT efficiently handles the increase in worker count and transaction load. However, as the number of workers increases, there is a slight increase in variability, but it is less pronounced compared to the Ethash configuration. The chart in 27 illustrates

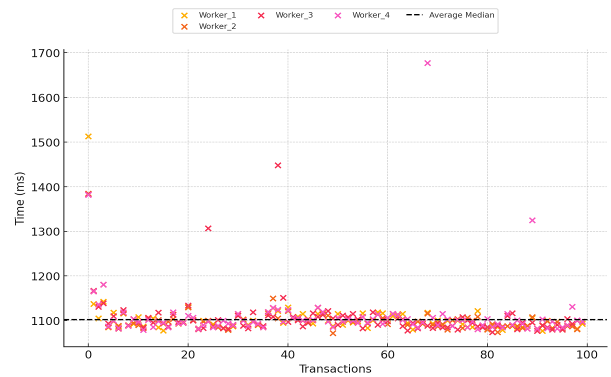

Fig. 19. QBFT Lineage Storage Time Variation with 4 workers/1s Block Period

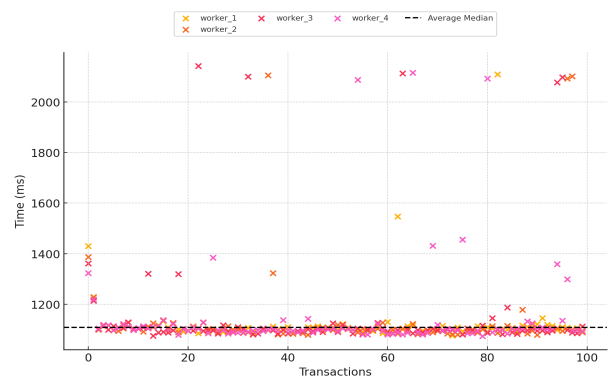

Fig. 20. QBFT Lineage Storage Time Variation with 4 workers/2s Block Period

the gas consumption for deploying lineage events using QBFT consensus algorithm across different worker configurations. Specifically, it contrasts the minimum and maximum gas usage for configurations involving 4, 8, 12, and 16 workers. The gas usage demonstrates a significant variability contingent on the worker configuration. For configurations involving 4 and 8 workers, the gas usage remains relatively low, with a negligible difference between the minimum and maximum values, indicating consistent performance. Notably, the 4-worker configuration shows a minimum gas usage of 39,360 and a maximum of 39,528, while the 8-worker setup maintains the same minimum gas value and exhibits a slightly increased maximum value of 39,528. As the number of workers increases to 12 and 16, the gas consumption escalates markedly. The 12-worker configuration experiences a sharp rise in maximum gas usage, peaking at 1,038,269 compared to its minimum value of 165,805. Similarly, the 16-worker configuration shows a significant range with a minimum gas usage of 165,805 and a maximum of 166,057. This substantial increase in gas usage suggests a higher computational overhead as the number of concurrent transactions scales, reflecting the complexity of maintaining consensus across more nodes. For the 4

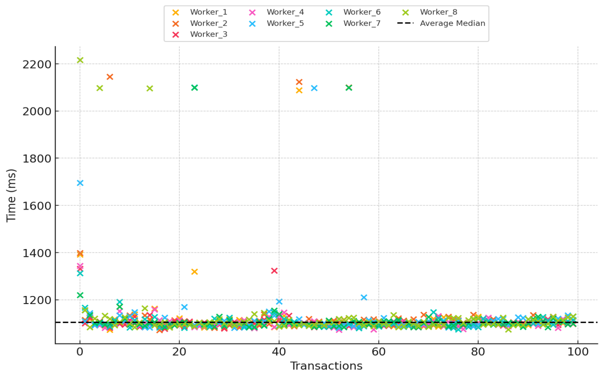

Fig. 21. QBFT Lineage Storage Time Variation with 8 workers/1s Block Period

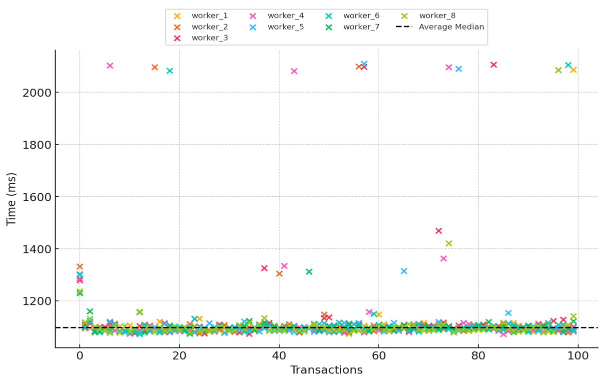

Fig. 22. QBFT Lineage Storage Time Variation with 8 workers/2s Block Period

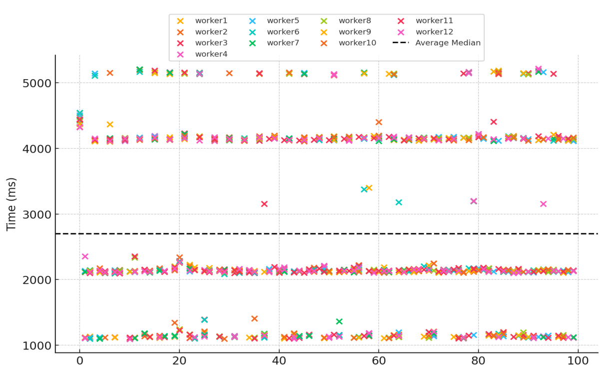

Fig. 23. QBFT Lineage Storage Time Variation with 12 workers/1s Block Period

Workers (BP1 and BP2) configurations, the median storage times are consistent, with most transactions completing around 1,200 ms. There are minimal outliers, indicating a stable

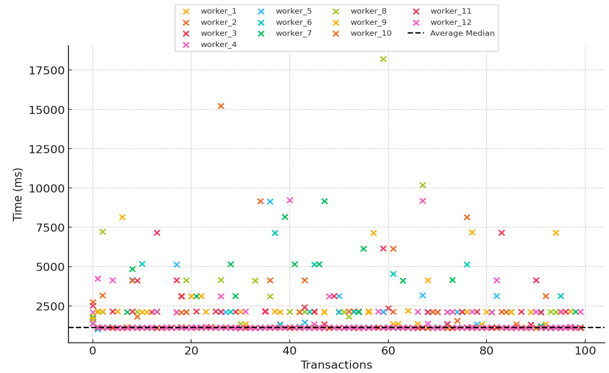

Fig. 24. QBFT Lineage Storage Time Variation with 12 workers/2s Block Period

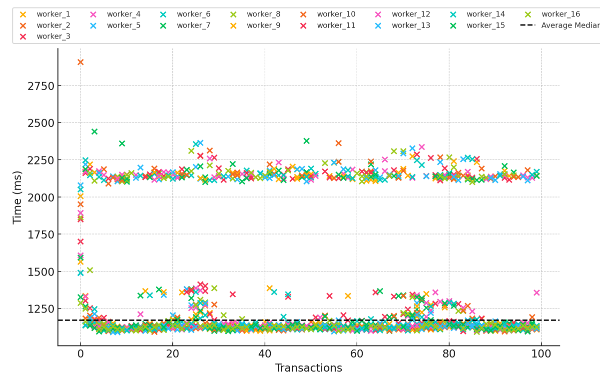

Fig. 25. QBFT Lineage Storage Time Variation with 16 workers/1s Block Period

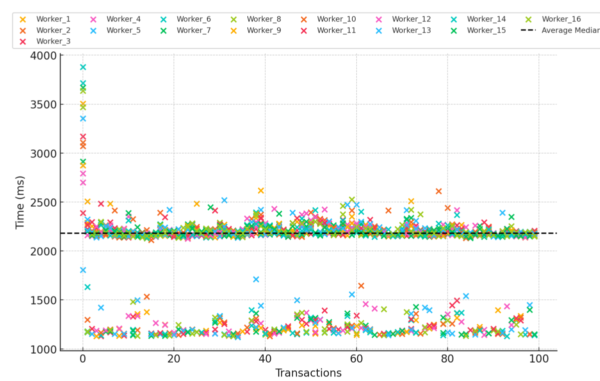

Fig. 26. QBFT Lineage Storage Time Variation with 16 workers/2s Block Period

performance with low concurrency. Then, when we multiplied the number of workers to 8, The median storage times remain stable around 1,200 ms. There are a few more outliers compared to the 4-worker setup, but the overall variability is still within acceptable limits. Using 12 Workers, the median times continue to hover around 1,200 ms. There is a noticeable increase in outliers, suggesting that while the system remains efficient, there are occasional delays possibly due to network or processing bottlenecks. For our last configuration, the median storage times are stable, but there is a marked increase in the number of outliers. Some transactions take significantly longer, indicating that the system is starting to experience more contention or resource limitations at higher concurrency levels.

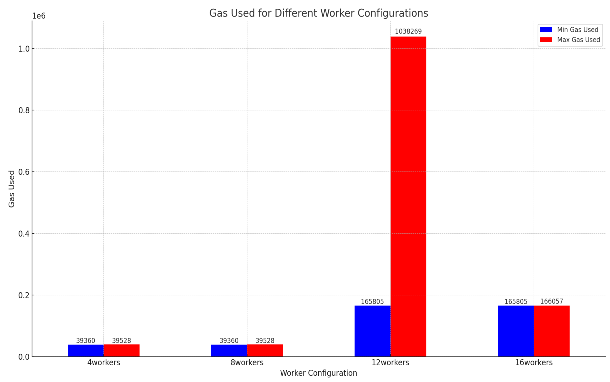

Fig. 27. QBFT Lineage Storage Cost

The QBFT configuration demonstrates robust performance under varying levels of concurrency. The median storage times remain consistently low across all worker setups, indicating that QBFT is well-suited for high-throughput environments. However, as the number of workers increases, there is a slight increase in variability and outliers, suggesting that while the system can handle increased load, there are occasional delays that may need to be addressed. Overall, QBFT's performance in lineage event storage is efficient and stable, making it a viable choice for scenarios requiring high transaction volumes and concurrency.

4) Storage Time Comparison Data Presentation, analysis and observations: The comparison charts for the average mean storage time across different worker configurations (4, 8, 12, and 16 workers) and different block periods (1 second and 2 seconds) for ETHASH, IBFT2.0, and QBFT consensus algorithms reveal several insights. ETHASH consistently shows the highest storage times across all worker configurations, indicating that it is the slowest among the three algorithms. For a block period of 1 second, ETHASH storage times range from 5228 ms to 5830 ms. When the block period is increased to 2 seconds, ETHASH storage times are similar, ranging from 5057 ms to 5830 ms. This suggests that ETHASH is relatively insensitive to changes in the block period, maintaining high latency regardless. IBFT2.0 and QBFT demonstrate significantly lower storage times compared to ETHASH. For a block period of 1 second, IBFT2.0 times range from 1293 ms to 2950 ms, while QBFT times range from 1105 ms to 1375 ms. This trend holds for the 2-second block period, where IBFT2.0 ranges from 1141 ms to 2679 ms and QBFT ranges from 1117 ms to 2007 ms. Both IBFT2.0 and QBFT show an increase in storage time with a higher block period, but the increase is more pronounced for IBFT2.0. The observations from the data are clear: QBFT consistently outperforms both ETHASH and IBFT2.0 in terms of storage time efficiency. ETHASH, while robust and secure due to its Proof-of-Work nature, is the

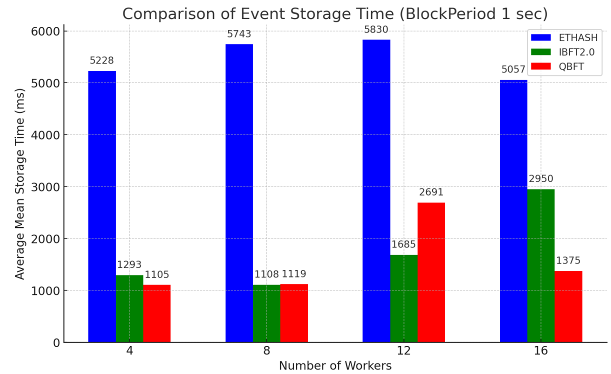

Fig. 28. Lineage Event Storage Time Comparison (IBFT2.0/QBFT BP 1)

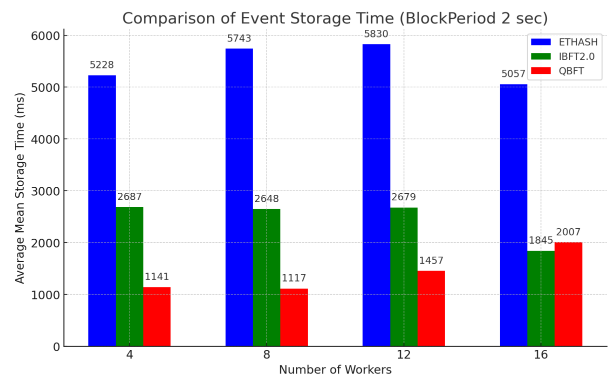

Fig. 29. Lineage Event Storage Time Comparison (IBFT2.0/QBFT BP 2s)

least efficient for storage operations, making it less suitable for scenarios requiring rapid transaction processing. The performance difference between IBFT2.0 and QBFT is also notable; QBFT not only has lower storage times but also demonstrates better scalability across increasing worker configurations. The increase in block period to 2 seconds impacts IBFT2.0 more significantly than QBFT, further highlighting QBFT's superior efficiency and adaptability.

In conclusion, for applications prioritizing low latency and efficient storage times, QBFT is the preferred consensus algorithm, offering the best performance in the tested configurations. ETHASH, despite its widespread use, is not optimal for storage-intensive operations due to its high latency. IBFT2.0, while better than ETHASH, still lags behind QBFT in both efficiency and scalability.

References

[1] BAIRD, Leemon, HARMON, Mance, et MADSEN, Paul. Hedera: A public hashgraph network and governing council. White Paper, 2019, vol. 1, no 1, p. 9-10.

[2] CHOI, Sang-Min, PARK, Jiho, NGUYEN, Quan, et al. Fantom: A scalable framework for asynchronous distributed systems. arXiv preprint arXiv:1810.10360, 2018.

[3] https://harmony.one/whitepaper.pdf

[4] BJELIC, Mihailo, NAILWAL, Sandeep, CHAUDHARY, Amit, et al. POL: one token for all polygon chains. 2017.

[5] BUTERIN, Vitalik et WOOD, Gavin. Ethereum Whitepaper.

[6] https://whitepaper.io/document/695/klaytn-whitepaper

[7] FAN, Caixiang, LIN, Changyuan, KHAZAEI, Hamzeh, et al. Performance analysis of hyperledger besu in private blockchain. In : 2022 IEEE international conference on decentralized applications and infrastructures (DAPPS). IEEE, 2022. p. 64-73.
[8] KALODNER, Harry, GOLDFEDER, Steven, CHEN, Xiaoqi, et al. Arbitrum: Scalable, private smart contracts. In : 27th USENIX Security Symposium (USENIX Security 18). 2018. p. 1353-1370.
[9] ARMSTRONG, Matthew. Ethereum, Smart Contracts and the Optimistic Roll-up. 2021.