

Immutable Log Storage as a Service on Private and Public Blockchains

William Pourmajidi¹, Lei Zhang², John Steinbacher³, Tony Erwin⁴, and Andriy Miransky⁵

^{1,2,5} Department of Computer Science, Ryerson University, Toronto, Canada

³ IBM Canada Lab, Toronto, Canada

⁴ IBM Watson and Cloud Platform, Austin, USA

william.pourmajidi@ryerson.ca, leizhang@ryerson.ca, jstein@ca.ibm.com,
aerwin@us.ibm.com, avm@ryerson.ca

Abstract

Service Level Agreements (SLA) are employed to ensure the performance of Cloud solutions. When a component fails, the importance of logs increases significantly. All departments may turn to logs to determine the cause of the issue and find the party at fault. The party at fault may be motivated to tamper with the logs to hide their role. We argue that the critical nature of Cloud logs calls for immutability and verification mechanism without the presence of a single trusted party.

This paper proposes such a mechanism by describing a blockchain-based log storage system, called Logchain, which can be integrated with existing private and public blockchain solutions. Logchain uses the immutability feature of blockchain to provide a tamper-resistance platform for log storage. Additionally, we propose a hierarchical structure to address blockchains' scalability issues. To validate the mechanism, we integrate Logchain into Ethereum and IBM Blockchain. We show that the solution is scalable and perform the analysis of the cost of ownership to help a reader select an implementation that would address their needs.

The Logchain's scalability improvement on a blockchain is achieved without any alteration of blockchains' fundamental architecture. As shown in this work, it can function on private and public blockchains and, therefore, can be a suitable alternative for organizations that need a secure, immutable log storage platform.

1 Introduction

In the majority of Cloud offerings, there are two parties involved. A Cloud Service Provider (CSP) owns a pool of computing resources and offers them at predefined prices to a Cloud Service Consumer (CSC) via the Internet.

The CSP uses continued monitoring to ensure that the current Quality of Service (QoS) provided to the CSC matches with the one in the signed Service Level Agreement (SLA). When a technical issue arises, the CSC will also become interested in reviewing and assessing the logs, because logs hold the truth about the delivered QoS.

While the full control over monitoring systems allows CSPs to monitor Cloud services efficiently, it gives them a controversial power over evidential resources essential to CSCs. That is, logs are generated and stored on a platform, which is built, managed, and owned by a CSP. Hence, CSPs have full permission on all collected logs. Such situations cause many trust-related issues.

Logs are evidential documents [1]. They contain all the details and QoS metrics related to the operation of software, network components, servers, and Cloud platforms. As a key element in computer forensic investigations, logs are presentable in the court of law [2, 3] only if they satisfy requirements, such as authenticity and reliability.

Log tampering include adding, removing, and manipulating a log partially or entirely. We provide three log tampering examples in Appendix A. Moreover, log tampering may affect CSCs financially and technically.

If a CSP tampers with the logs related to resource usage and overcharges the customer, or if a CSP hides the breach of one or more criteria of an SLA, the CSC is in immediate need of finding a method or a tool to verify the integrity of data provided by the CSP.

Given the existence of many motivations to tamper with logs, the negative consequences of log tampering for the CSCs, and the inadequacy of existing monitoring solutions, we conclude that an immutable log system, which is capable of storing the logs and verifying their integrity, can be genuinely beneficial for the CSCs and can be used to establish trust among Cloud participants.

To address the trust issue among Cloud participants, our **objective** is to create an immutable log system called Logchain (LC). We choose blockchain as our immutable storage option. The proposed solution collects logs from various platforms and stores them in blocks of blockchains. To make the LC more accessible, we implement an API module for LC that allows participants to submit logs to the LC and verify the integrity of the submitted logs to the LC. As the API allows the LC to be used as a service, from here onward, we use the term Logchain as a service (LCaaS) to refer to the API-enabled LC.

Traditionally, blockchains use two types of blocks. The genesis blocks mark the start of a blockchain, and the data blocks store data in the blockchain. Additional details about these blocks are provided in Section B.1. We introduce a conceptual model in which we extend the definition of the data blocks to allow us to construct additional types of blocks while keeping the internal architecture of the data blocks intact. We design a special type of data block named Terminal Block (TB). The terminal block terminates a blockchain, and the terminated blockchain can no longer accept data blocks. The definition and technical details of additionally constructed block types are provided in Section 3.1. In our definition of special types of blocks, we keep the internal architecture of data blocks intact to ensure blockchains can work with LCaaS in the future.

The LCaaS API receives the logs (or their digest) and sends them to LCaaS *Blockify* module to convert them to the blocks and store the blocks in the blockchain. Blockchains are known to have scalability issues [4]. Hence, to address the scalability issues of blockchains, we build a hierarchical ledger that uses inter-related blockchains to form a scalable log storage platform. In LCaaS, blocks are constantly added to a lower-level blockchain until a time- or size-bound threshold is triggered. Once the trigger is sent to the *Blockify* module, it needs to terminate the current lower-level blockchain, summarize its important data into a TB, and submit the content of the TB to a data block of a higher-level blockchain. TB contains the most important information needed for verification of data stored in all the blocks of a lower-level blockchain that it has terminated. Therefore, verifying the integrity of a single TB proves that none of the blocks of a lower-level blockchain have been tampered with, saving time and resources on the verification process. The details of LCaaS design are captured in Section 3.

The main **contribution** of this paper is to extend our work¹ on the applicability of LCaaS as a hierarchical ledger [5, 6] from a proprietary proof-of-concept blockchain to public and private commercial blockchains. We selected Ethereum [7] as a public blockchain vendor and IBM [8] as a private blockchain vendor. We chose them as they are popular and representative blockchains vendors.

We compare the results of LCaaS integration to these two blockchain platforms. We also assess and compare the cost of ownership for the two implementations.

We focus on the primary integration point between LCaaS and blockchain vendors and assess integrated solutions' performance by measuring the impact of all relevant factors. The incoming transactions per second (*tps*) defines the incoming load for the LCaaS. The higher the *tps*, the more work there is for LCaaS and its hashing methods. Nevertheless, the *tps* does not directly impact the number of outgoing transactions to blockchain vendors. The number of outgoing transactions is defined by the length of lower-level blockchains, also known as circled blockchains. The longer circled blockchains result in a lower frequency of submission of blocks to blockchain vendors. While this lower submission makes the integration more feasible (as there are fewer transactions to be paid for), it creates a longer window of time for logs to be stored locally before they can be sent to blockchain vendor, hence a higher risk.

While there are solutions that recommend the use of blockchain for storage of critical data (e.g., [9, 10, 11]), to the best of our knowledge, none of them have tried to address the scalability issues of blockchains for large-scale log storage. As a result, most existing solutions have settled for partial storage of logs, such as audit logs [12, 13]. While this solves a portion of the problem, we argue that the entire log storage system

¹In [5, 6], we mainly focused on design of LCaaS and its interaction with Ethereum blockchain.

has to be immutable.

We design the proposed solution as a service so that multiple customers can use it concurrently. Moreover, given that CSCs are already using many “X-as-a-Service” solutions, they are accustomed to such a service offering model.

The clients of LCaaS can use its API to interact with the solution and submit and verify logs. The source code of LCaaS can be accessed via [14].

CSPs can use LCaaS to promote visibility, trust, and accountability. CSPs can choose to submit their operational logs to LCaaS and offer a verification mechanism to their CSCs. Similarly, the CSCs can confirm the authenticity of the generated logs for their consumed services using the verification API. The CSPs can decide what components or portions of their infrastructure need to send logs to the LCaaS. Ideally, all the Cloud platform components, from the core network to the user-facing services, should submit their operational logs to the LCaaS. The more components of the Cloud platform that submit logs to the LCaaS, the more trustable the relationship between the CSPs and the CSCs will become. However, as submission of data to blockchain incurs a cost, a justifiable feasibility assessment is suggested. Additionally, suppose CSPs are concerned about the confidentiality and secrets that submission of the actual logs to the LCaaS may reveal. In that case, they can choose to submit the digest of the logs. While submission of the digest of logs offers peace of mind in terms of any potential breach of privacy or sensitive information, LCaaS can not offer the original version of the logs if they are tampered with.

The rest of this paper is structured as follows. In Section 2, we provide related works and a brief literature review on both Cloud computing and blockchain. In Section 3, we introduce the methodologies that we use to build LCaaS, our prototype, and the approaches for enhancing blockchain and its capability, followed by the implementation of the hierarchical ledger. In Section 4, we present implementation and integration details of LCaaS to blockchain vendors, namely, Ethereum and IBM Blockchain. In Section 5, we analyze the results of the integration of the LCaaS model with Ethereum and IBM Blockchain. In Section 6, we discuss practical aspects of implementing LCaaS. In Section 7, we summarize potential threats to the validity of the proposed solution. Finally, in Section 8, we conclude the paper by providing a summary and a direction towards future work.

2 Literature Review

In Section 2.1, we review the importance of Cloud monitoring and its related challenges. In Section 2.2, we provide an overview of blockchain and its characteristics, concluding with a review of blockchain capacity limitations and their solutions. We also refer the reader to Appendix B for background details on blockchain and its taxonomies.

2.1 Cloud Logs and Related Challenges

The legal system relies on a range of forensic investigation and identification. In the digital era, digital evidence such as operational logs, transaction logs, and usage logs replace physical evidence. This new type of evidence requires new forensic investigation methods [15]. In an attempt to regulate and standardize digital forensics practices, the Digital Forensic Research Workshop (DFRWS) [16] has developed a forensic framework. The framework identifies preservation as a crucial step and indicates that it must be a guarded principle across “forensic” categories.

Another major challenge related to the authenticity and reliability of digital evidence is that digital data are much more easily tampered with compared to physical evidence. To effectively use logs as digital evidence, many have recommended using a Log Management System (LMS) [17, 18]. The majority of LMSs promise a set of desirable features, such as tamper-resistance, verifiability, confidentiality, and privacy [18]. To secure the collected logs, specifically audit logs, Waters et al. [13] build a platform that uses hash encryption to protect the audit logs from unauthorized parties by encrypting the content of them. However, LMSs are managed by IT personnel; thus, requiring the presence of a trusted third party (TTP), limiting LMSs applicability.

In recent years, many verification-as-a-service platforms offer integrity control for the data that are uploaded by the user, but they do need a TTP. For example, arXiv [19] provides a repository that ensures document’s integrity.

The problem of trusting a third-party can be alleviated by a self-contained solution that does not rely on a TTP integrity verification service. We argue that blockchain is among the most promising solutions that can be used to replace the requirement for a TTP. It was initially designed to support a cryptocurrency known as Bitcoin that does not require a TTP (such as banks or other financial institutes) for verification or maintenance of financial transactions. A correctly implemented distributed blockchain is an adequate alternative to address the TTP issue [20, 21].

2.2 Blockchain

A significant component of blockchain that allows transactions to be immutable is the use of the Proof of Work (PoW) verification schema. PoW involves running iterations for finding a particular value that, when it is hashed in conjunction with other elements of a block, the calculated hash begins with a certain number of zero bits. The number of zeros is proportional to the time required to complete a PoW. The higher the number of zeros, the longer it will take to complete the PoW. Once the computational effort is dedicated and the hash value is found, all items along with the found value, known as *nonce*, are kept in a block. The content of a block cannot be changed unless the whole PoW process is repeated. Chaining blocks together using hash binding or hash chaining [12, 22, 23] significantly increases the amount of computational effort that is needed for changing the content of an earlier block. In a hash binding relationship, the current hash of the previous block is used as the previous hash of the current block. This chain makes any attempt to change the blockchain computationally unfeasible as one needs to re-process PoW for all the blocks in order to tamper with any of the earlier blocks [24].

Current blockchain implementations of the distributed ledgers already have notary proof-of-existence services [20]. For example, Poex.io [25], launched in 2013, verifies the existence of a computer file at a specific time, by storing a timestamp and the SHA-256 [26] of the respective file in a block that eventually will be added to a blockchain. Proof-of-existence solutions cannot be used as scalable LMSs, as they consider files individually, with no search function to locate the appropriate file or block. Moreover, Cloud solutions consist of thousands of components, each of which generates a large volume of logs [27]. The current solutions are not designed to handle the scale that is required to store Cloud-generated logs. Furthermore, the current public blockchains support limited concurrent transactions [20].

Although blockchain technology has great potential and can be used in many disciplines, it is dealing with a number of challenges. The scalability remains the most critical challenge [28]. Blockchain heavily relies on consensus algorithms, like PoW, and such algorithms are computationally expensive. To overcome the scalability issues, a novel cryptocurrency scheme is suggested by [29] where old transactions are removed from the blockchain, and a database holds the values of removed transactions. Although this solution reduces the size of the blockchain, it introduces the same trust issue that traditional databases are suffering from.

Eyal et al. [30] suggest redesigning the current structure of a blockchain. In the redesigned model, known as Bitcoin-NG (next generation), conventional blocks are decoupled into two parts: the key block and microblocks. The key block is used for leader election, and the leader is responsible for microblock generation until a new leader appears.

3 Design of LCaaS

In this section, we introduce the methodologies that we use to build LCaaS and explore the implementation of the hierarchical ledger. Here, we provide a general overview of LCaaS, its details, and the additional components that are added to the blockchain to enhance LCaaS' scalability. LCaaS API, its signatures and methods are given in Appendix C.

Current blockchain consensus protocols require every node of the network to process every blockchain block, hence a major scalability limitation. PoW is among the most common consensus algorithms at the time of writing this work [31]. Thus, to ensure that our proposed solution is useful for the majority of blockchains, we use PoW as the main consensus algorithm for the LCaaS. However, our modular design allows one to easily replace PoW with any other consensus algorithms such as Proof-of-Stake, Proof-of-Importance, or newer ones based on the Byzantine Fault Tolerance (BFT), which aim to solve the problem of reaching consensus when nodes can generate arbitrary data [32].

As blockchain consists of various components, different solutions have been offered for scalability issues of each specific component. For instance, On-chain, Off-chain, and Inter-chain solutions are used to improve the overall performance of blockchains [33]. LCaaS is mainly concerned with the performance issues caused by the size of the blockchain (number of blocks) and uses an approach known as Child-chain [34], where a parent-child structure is promoted and continuous transactions are stored in a child-chain (similar to our lower-level blockchain), and the results are stored in a parent-chain (similar to our high-level blockchain).

The performance of a blockchain platform is mainly impacted by the number of its blocks, the size of these blocks, and the consensus algorithm it uses. While choosing an appropriate consensus algorithm has a major impact on the blockchain performance [35], our focus is on the performance issues caused by a large number of blocks in a never-ending blockchain.

We overcome this size issue by segmenting a portion of a blockchain and locking it down in a block of a higher level blockchain, i.e., we create a two-level hierarchy of blockchains. Validating the integrity of a high-level block confirms the integrity of all the blocks of the lower-level blockchain. This segmentation leads to a more efficient validation process. The hierarchical structure of LCaaS is graphically shown in Appendix D.

LCaaS resides on top of a basic blockchain and converts it to a hierarchical ledger. Our primary goal is to bring scalability to blockchain for the situations in which the number of data items stored in a blockchain is large (e.g., operational logs of a Cloud platform).

As the name implies, the LCaaS offers the hierarchical ledger as a service. Cloud participants can create an account and receive a unique API key for all corresponding API calls. Clients also need to configure two main settings on their instance before they can use it. The first key configuration is the *difficulty_target*, which is defined as the number of required zeros at the beginning of an acceptable hash. The LCaaS will continue to generate new hashes and new *nonces* until a hash is generated that matches the difficulty target. The second key configuration is defining a limit for the number of blocks in a circled blockchain. This constraint acts as a size-limit and controls how many blocks are accepted in each circled blockchain. Once the limit is reached, the LCaaS takes the blockchain and pushes it to the hierarchical ledger. Let us now look at the key components of the LCaaS.

3.1 Enhancements on Blockchain Structure

While common key components of blockchains are necessary to implement a blockchain, LCaaS requires additional components. We have introduced absolute genesis block, relative genesis block, terminal block, circled blockchains, super block, and super blockchains. These advancements allow the LCaaS to provide the hierarchical structure that improves the scalability of blockchains.

Absolute Genesis Block (AGB): Similar to the Markle root in a Markle tree [36], the absolute genesis block is placed as the first block of the first circled blockchain. An AGB is the first block that is created in the LCaaS and has the same characteristics as GB, with *index* and *previous_hash* set to zero and the *data* element set to null.

Relative Genesis Block (RGB): Relative genesis block is placed at the beginning of every subsequent circled blockchain after the first circled blockchain. The *previous_hash* of an RGB is set to the *current_hash* of the terminal block of the previous circled blockchain.

Terminal Blocks (TB): Terminal Blocks are added at the end of a blockchain to “close” it and produce a circled blockchain. The terminal block’s *index* and *current_hash* are calculated similarly to any other block. The part that differentiates a terminal block from a genesis block or a data block is its *data* element. Data blocks of blockchains are designed to be data-type agnostic. The majority of Cloud monitoring tools allow data to be submitted to external sources for further analysis. At the time of writing this work, JSON format is the de facto data exchange standard for such data exchange. Hence, we assume that the CSP can submit data to our API using the JSON format. As a result, the terminal block’s *data* element stores a JSON object that contains details about the terminated circled blockchain. Practitioners who adopt LCaaS may choose other machine-readable formats, such as XML, YAML, or a user-defined one. The details and the processes to create a TB are as follows. The *aggr_hash* is created by collecting and hashing *current_hash* values of all blocks in that circled blockchain, from the AGB or RGB to the block before the terminal block. The *data* element also store four additional values, namely *timestamp_from*, *timestamp_to*, *block_index_from*, and *block_index_to*.

Circled Blockchains (CB): Circled blockchains are blockchains that are capped. In other words, there is a limit on the number of blocks that they can include before a terminal block “caps” the blocks. Once a circled blockchain is terminated, it can not accept any new block.

Super Blocks (SB): Super blocks exhibit the features of regular data blocks and have *nonce*, *index*, *timestamp*, *data*, *previous_hash*, and *current_hash*. The only difference between a super block and data block is that super block’s *data* element stores all of the field of a terminal block of a circled blockchain. In order to accept terminal block elements, the *data* element consists of a JSON object. The elements of this JSON object are as follows: *index*, *timestamp*, *data*, *current_hash*, *previous_hash*, and *nonce*.

Super Blockchain (SBC): Super blockchain is a blockchain where each of its blocks is a super block. The super blocks are “chained” together by hash binding. In other words, super blocks that are linked together will result in a super blockchain. An i -th super block in a super blockchain relies on *current_hash* of its previous super block. If data in an earlier super block m is tampered with, the link among all the subsequent super blocks, from $m+1$ to the most recent super block, denoted by super block i , will be broken. Then one has to recompute *current_hash* and *nonce* values of each super block from super block m to super block i .

Figure 9 shows the relationship between a TB and all other blocks in a CB. All elements of a SB are identical to the ones of a data block. Thus, it can be implemented by any other blockchain framework.

Considering that a SB’s data element includes all the elements of a TB, changing any block in a CB, not only breaks the CB but also breaks the SBC. The relationship between a TB and the data element of a SB is shown in Appendix D.

The data element of the SB provides a hash tree structure and enhances the immutability the CBs, while decreasing the computational resources required to verify blocks in a circled blockchain.

The above novel enhancements allow the LCaaS to provide the hierarchical structure that is needed to overcome scalability limitations of the blockchains.

4 Validation Case Study

Here, we review the validation of the LCaaS. Section 4.1 provides rationale for choosing Ethereum and IBM Blockchain. Sections 4.2 and 4.3 cover details of integration with the Ethereum and IBM Blockchains, respectively. Section 4.4 discusses controlling factors of Ethereum and IBM Blockchain integration. Finally, Section 4.5 depicts the workload drivers.

4.1 Integration Platforms - Rationale

4.1.1 Rationale for choosing Ethereum

While the focus of some public blockchains (such as Bitcoin and Litecoin) is on financial transactions, other public blockchains such as Ethereum try to provide different use cases for blockchains. Coherently, Ethereum provides the developer with an end-to-end system for building various distributed applications [37].

The smart contracts are autonomous pieces of code [38]. They are deployed over the blockchain and upon being called, can interact with the user data or data stored in the blockchain. Ethereum has developed Solidity [39], a high-level language for smart contracts, and Remix [40], an Integrated Development Environment (IDE) for Solidity.

Given the popularity of Ethereum, we have selected Ethereum as the public blockchain platform for integration with the LCaaS.

4.1.2 Rationale for choosing IBM Blockchain

IBM Blockchain provides developers with an end-to-end platform for designing, building, and implementing various applications based on the underlying blockchain technology [41].

Another major feature of IBM Blockchain is its extensive support for the smart contracts. Within the IBM Blockchain, smart contracts are known as Chaincode [42]. To simplify the development process, IBM Blockchain is equipped with an extension [43] for Visual Studio code(VSC) [44] that can be used for building and deploying smart contract on the IBM Blockchain platform. The extension uses direct access to IBM

Blockchain (of course, after successful authentication) and can directly edit, run, and deploy smart contracts on the IBM Blockchain. Smart contracts for IBM Blockchain can be developed in Java, JavaScript, and Go [45].

Given its popularity and its wide range of use cases [41], we have selected IBM Blockchain as the private blockchain platform for integration with the LCaaS.

4.2 Case study setup: Ethereum

Figure 1 depicts the relationship between LCaaS and Ethereum. The *API* module receives API calls from clients and passes them to the *Logchain* Module. The *Logchain* module employs the *Blockchain* module to convert the received data (digest or raw logs) to blocks and pushes a copy of the blocks to Firebase and another copy to the *API* module. The *API* module pushes the received data to the Ethereum and informs the client of the successful submission of data to the blockchain. We use Google Firebase real-time database [46] to store all types of blocks as additional permanent storage. Integration with the Firebase can be disabled without affecting the normal operation of the LCaaS.

LCaaS is built on top of a private blockchain. In order to replace it with a public blockchain (e.g., Ethereum), integration points have to be designed. We propose a composite structure, in which receiving logs and converting them to blocks happens at the LCaaS side and storing the hashes and digitally signing them happens over the Ethereum blockchain. Using blockchain terminology, data collection and blockification of logs happen off-chain, and the block storage on the Ethereum blockchain is handled by Ethereum smart contract and will be on-chain.

Within the Ethereum blockchain, economics is controlled by an execution fee called gas. The gas is paid by Ether—the Ethereum intrinsic currency [37]. Gas measures, in computation resource terms, the effort that is needed to process the transaction. A smart contract consists of one or more operations, and each operation has an associated gas cost, which is defined by the Ethereum protocol [47]. For instance, a SHA-3 operation costs 30 units of gas. The higher the gas price, the more appealing the transaction would become for the miners. Hence, if a transaction needs to be executed faster, the higher gas price will motivate a miner to consider the transaction and mine it in the upcoming block. The current implementation of LCaaS does not use a lot of computational resources for each block that is pushed to the blockchain; thus, the main bottleneck is at the blockchain provider side. As indicated above, one can increase the performance of the blockchain by increasing the gas price for the desired transaction.

In the light of the above economics and the fact that each transaction incurs a cost, we limited the submissions to the Ethereum blockchain to super blocks. Super blocks include complete elements of a terminal block of a circled blockchain and can be used to verify the integrity of all the blocks in that circled blockchain. Based on the preimage resistance property of hash functions (mentioned in Section B.2), it would be computationally infeasible to construct an entire circled blockchain such that its hash matches the *current_hash* of a super block. However, if one desires, minor changes to the current implementation of the *Ethereum* module can be made to allow the LCaaS to push data blocks to the Ethereum blockchain as well.

Since the main Ethereum blockchain is used for production, Ethereum team has allocated a test network for development purposes [48]. In this work, we have used the Ethereum test network (as opposed to the real one) for the integration of the LCaaS and Ethereum. The cost of transactions on the Ethereum test network is paid with a special type of Ether, known as test Ether which does not have any real monetary value. To obtain test Ether, we use MetaMask Ether Faucet [49].

Smart contracts on the Ethereum blockchain allow users to interact with the blockchain. The storage of SBs in the blockchain requires a smart contract. It is important to mention that a SB’s data element contains the terminal block of a CB. Hence, SB is the most efficient candidate to be stored in a public blockchain, as one can easily verify the integrity of a SB and conclude the integrity of all the blocks in the CB that the TB has terminated.

We use Solidity [39] to develop the smart contract and name it *Superblock.sol* (available at [14]). Once the smart contract is developed, it has to be published on the Ethereum blockchain. We use the Remix to publish the smart contract. Publishing a smart contract on the blockchain is considered a transaction and is a chargeable service. Thus, we use the test Ether that we have stored in MetaMask vault to pay the transaction fee.

The published smart contract stores the SBs on the Ethereum blockchain and, upon successful submission,

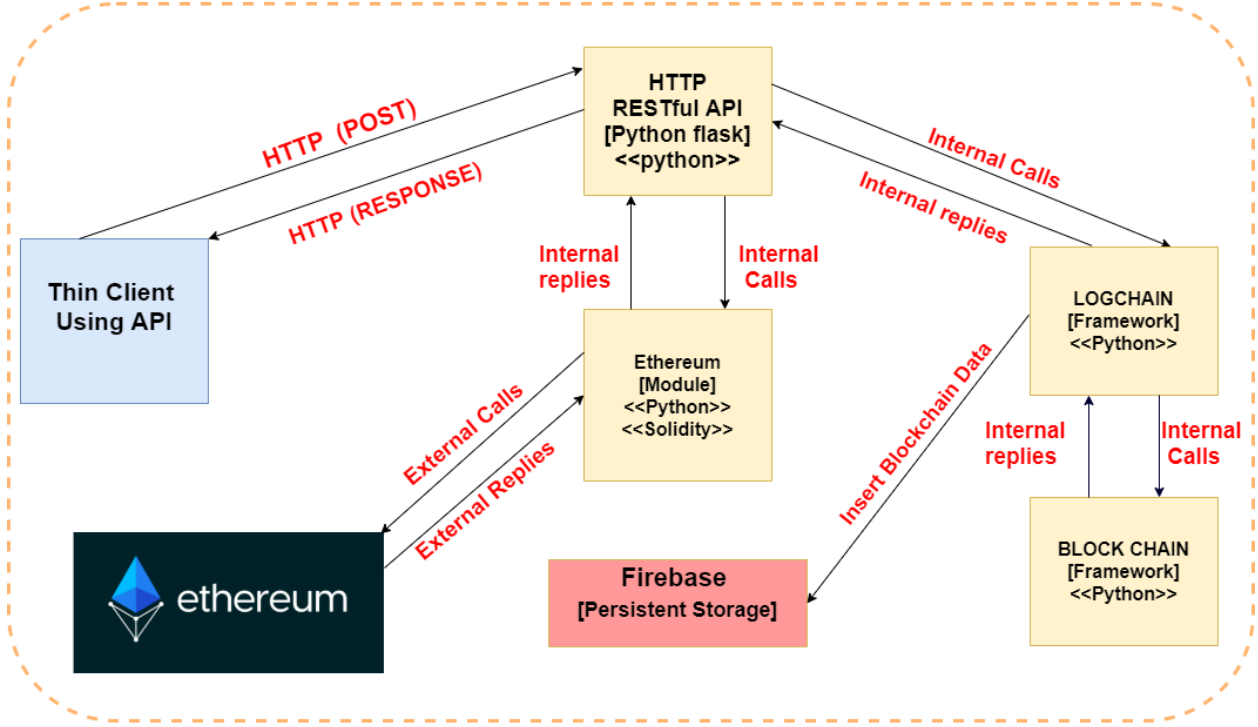


Figure 1: LCaaS and Ethereum integration

returns a receipt back to the *Ethereum* module. The receipt includes details of the transaction, such as the sender address, the content, the transaction hash, and the block number.

All interactions with the Ethereum blockchain can be traced using Etherscan [50], a web dashboard connected to Ethereum blockchain. Known as Ethereum block explorer, Etherscan allows anyone to look up transactions' details by using the sender or recipient address, transaction hash, or block number.

Using the Ethereum *blockNumber* field or the sender address, one can verify the transaction. If the transaction is found and the submitted SB matches with the one at hand, the integrity of the TB in the data element of the SB is confirmed; thus, confirming the integrity of all blocks in the CB that the TB has terminated. An example of a successful transaction of LCaaS on Etherscan can be seen in [51].

4.3 Case study setup: IBM Blockchain

The relationship between LCaaS and IBM Blockchain is depicted in Figure 2. The process of receiving data, processing it, and converting it to blocks remain the same as the processes explained in Section 4.2 except that data are pushed to IBM Blockchain instead of Ethereum.

As for the integration between LCaaS and IBM Blockchain, receiving logs and converting them to blocks happens at the LCaaS side and storing the hashes and digitally signing them happens over the IBM Blockchain.

Unlike Ethereum, there is no concept of gas price on IBM Blockchain. However, the economics are controlled, on hourly bases, and based on virtual processor core (VPC) allocation. This simplified model is based on the amount of CPU (or VPC) that the IBM Blockchain Platform nodes are allocated on an hourly basis. To further clarify the concept of VPC, is important to mention that a VPC is a unit of measurement that is used to determine the licensing cost of IBM products and is based on the number of virtual cores (vCPUs) that are available to the product. A vCPU is a virtual core that is assigned to a virtual machine or a physical processor core. Within the IBM Blockchain platform, the platform cost estimation for 1 VPC = 1 CPU = 1 vCPU = 1 Core [52].

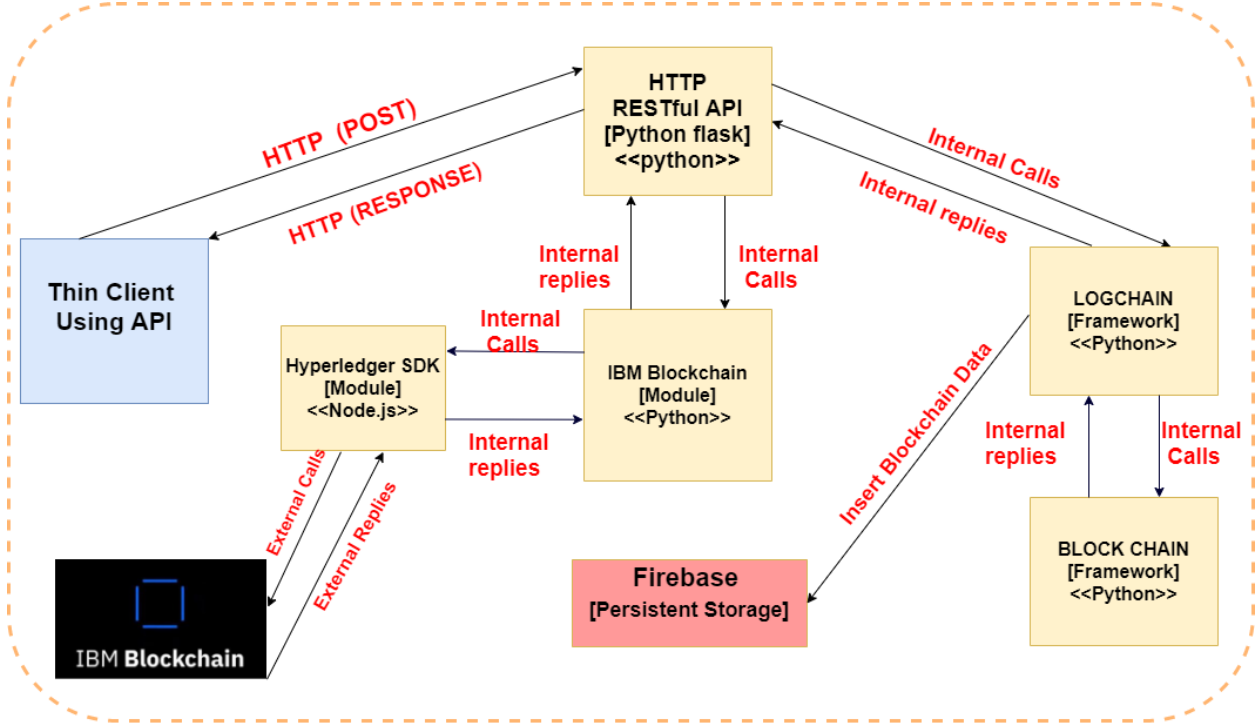


Figure 2: LCaaS and IBM Blockchain integration

There are some benefits of using this pricing model for IBM Blockchain. To begin with, it resembles Cloud pricing mode, namely hourly, pay-as-you-go model. Additionally, it brings clarity to estimations and as there is no minimum-investment requirement, developers and early adopters can try out the platform

Given that this is a paid service where each transaction incurs a cost, we limited the submissions to the IBM Blockchain to super blocks as they include complete elements of a terminal block of a circled blockchain and can be used to verify the integrity of all the blocks in that circled blockchain. Figure 2 depicts the relationship between LCaaS and IBM Blockchain.

The IBM Blockchain is a paid-service, and can be deployed on the IBM Cloud [53] as a service. We create an account on the IBM Cloud and follow the guidelines listed on the IBM Blockchain developer support website [54] to deploy and configure an instance of the IBM Blockchain. The latest version of VSC was downloaded and the IBM Blockchain add-on for the VSC was installed. As reviewed in Section 4.1.1, smart contracts allow external actors to interact with a blockchain. In the case of IBM Blockchain, smart contracts, known as chaincodes, are used for the same purpose.

The IBM Blockchain uses Hyperledger fabric [55] as its blockchain framework and work based on the open source tools hosted by the Linux Foundation [56]. Thus, we created a local development/deployment environment similar to the IBM Blockchain by instantiating a local version of the Hyperledger fabric blockchain. Once the smart contract code was up and running, we ported it to the IBM Blockchain.

In this work, to develop smart contracts, our first attempt was to implement them using Python; however, the Hyperledger Python SDK [57] was not ready at the time of this experiment; hence, we chose JavaScript (running on the Node.js backend). The next step was to package the smart contract using the IBM Blockchain Platform Extension for VSC and export it so it can be imported to the IBM Blockchain network.

The VSC IBM Blockchain extension allows direct interaction with the IBM Blockchain as long as a private blockchain with credentials (permissioned) is set up, and its connection details are configured on the VSC side. To instantiate an instance of The IBM Blockchain platform, one has to create a network [58] onto an IBM Kubernetes [59]. After deploying the network, we created the required Membership Service Provider (MSP) and a Certificate Authority(CA) for the private blockchain and associated them with a channel. Each

Table 1: Configurable factors. Note that g is specific to Ethereum integration; 1Ether = 10^9 gwei.

Factor	Values
Transactions per second (tps)	[0.1, 1, 10, 100]
Length of circled blockchain (n)	[1, 10, 100]
Gas price (g) measured in gwei	[6, 9, 20]
Number of sample log files	200 or 1000
Size of sample log files measured in bytes	64
$difficulty_target$ measured in bytes	000 prefix

transaction on a network is executed on a channel that is a private sub-network of the main blockchain and only allowed participants (known as organizations) can communicate and conduct transactions. Once the channel is created, the developed chaincode is imported and associated with it. The deployed chaincode constructs the business logic that receives data from LCaaS and submits it to the IBM Blockchain. At this stage, the IBM Blockchain is configured, and there is a channel with authorized participants who can submit data as blocks to the private instance of IBM Blockchain instantiated on the IBM Cloud. The only remaining part for the end-to-end integration is a module that connects LCaaS to the instantiated instance of IBM Blockchain as the private blockchain.

Many distributed blockchains, such as Ethereum and Bitcoin, are not permissioned, which means that any node can participate in the consensus process, wherein transactions are ordered and bundled into blocks. Because of this fact, these systems rely on probabilistic consensus algorithms which eventually guarantee ledger consistency to a high degree of probability, but which are still vulnerable to divergent ledgers (also known as a ledger “fork”), where different participants in the network have a different view of the accepted order of transactions. As for the Hyperledger Fabric, the open source platform behind the IBM Blockchain, things work differently. Hyperledger Fabric features a kind of a node called an orderer (it’s also known as an “ordering node”) that does this transaction ordering, which along with other nodes forms an ordering service. Because Fabric’s design relies on deterministic consensus algorithms, any block a peer validates as generated by the ordering service is guaranteed to be final and correct. Ledgers cannot fork the way they do in many other distributed blockchains.

4.4 Test bench and control factors

4.4.1 Test bench

To test the performance of LCaaS and its integration with blockchain vendors, we designed a load test which we run on our test computer with Intel i7-7500U CPU and 16 GB of RAM. The main goal of the load test is to evaluate the impact of the configurable factors: incoming transactions per second and length of circled blockchains.

For Ethereum, we use publicly available Ethereum test network (as was discussed in Section 4.2).

IBM Blockchain can be deployed in two different environments. The first, and preferred method is to deploy it on the IBM Cloud while the second option is to install and configure it on-premises [60]. We chose the Cloud deployment as it was the recommended deployment model. The default configuration of computing resources of the IBM Blockchain [61] seem adequate for our test scenarios, but one can increase the computing resources for a larger storage and/or a higher performance.

To be able to compare the results of submission of super blocks to Ethereum and IBM Blockchain, we kept the rest of the configurations alike. Table 1 provides a summary of controlling factors and workload drivers for Ethereum and IBM blockchain respectively. Below, we will provide the details of each of the factors.

4.4.2 Common configurable factors

Incoming transactions per second (tps): The number of tps depends on the velocity of log generation at the CSP side and the way that CSP decides to push their logs into LCaaS. The transaction in this case may

represent a log file. It is important to mention that the change in *tps* does not directly lead to a change of frequency of submissions to the Ethereum blockchain. This is because LCaaS will store the received logs in blocks of a circled blockchain and then will push the generated super block to the Ethereum. By varying the value of *tps*, we mimic workloads with different intensities, from the least intense at *tps* = 0.1 to the most intense at *tps* = 100.

Length of circled blockchains (*n*): Circled blockchains store a genesis block, one or many data blocks, and one terminal block. In other words, the number of data blocks in circled blockchains is configurable. A larger number of data blocks in a circled blockchain will result in a lower frequency of submissions to the Ethereum. Since each submission has a transaction fee, configuring a larger value for the number of data blocks in a circled blockchain seems more feasible. While this is true, it brings an additional risk. The risk is caused by the fact that LCaaS stores generated blocks internally and waits until the circled blockchain is terminated before sending the respective super block to the Ethereum blockchain. Hence, transactions in a longer circled blockchain are stored for a longer period, and this gives a longer window to an adversary to tamper with the logs (we will further discuss this in Section 6.2). Therefore, by varying the value of *n*, we mimic LCaaS handling from the most sensitive data (at *n* = 1) to the least sensitive one (at *n* = 100). We submit the digest of a log file (64-byte long) using *submit_digest* function. Thus, the length of the file is always constant. We did not measure the performance of *submit_raw* in this series of experiments, but we know that most of the time in this function is spent in computing the digest, which is proportional to the length of the file. On our test computer, it takes 200 ms to compute SHA-256 digest for 1 MB file, 1.5 seconds — for 10 MB file, and 15 seconds — for 100 MB file. The time to transfer the raw file from the user to LCaaS will also be proportionate to the length of the file. However, if the internal network is fast, then the transfer time will be small and can be ignored. Note that the time needed to process a super block is independent of the length of a raw file, as we are dealing with digests of the files at that stage.

We set the difficulty target for internal computations to 000. To ensure that there are enough submissions to Ethereum for each of the 36 experiments, for the setups with *n* ∈ [1, 10], we submit 200 digests representing 200 log files to the LCaaS; and for the setups with *n* = 100, we submit 1000 digests representing 1000 log files to the LCaaS.

4.4.3 Ethereum-specific control factors

In addition to *tps* and *n*, Ethereum has gas price (*g*) as another controlling factor.

Gas price (*g*): The gas price is the amount paid per unit of gas and is defined by the initiator of the transaction. The higher the gas price, the more appealing the transaction would become for the miners. Hence, if a transaction needs to be executed faster, the higher gas price will motivate a miner to consider the transaction and mine it in the upcoming block. The ETH Gas Station [62] keeps track of all submitted transactions and their processing times and suggests a value of *g* for different processing speeds. For our experiments, we tried setting *g* ∈ [6, 9, 20] gwei. These three values of *g* on September 10, 2018 corresponded to processing times of less than 30, 5, and 2 minutes, respectively (based on [62]).

4.4.4 Test scenarios

Based on the configurable items and suggested values for gas price², for Ethereum, we designed 36 scenarios consist of combination of each one of the following possible values: *tps* = [0.1, 1, 10, 100], *n* = [1, 10, 100], and *g* = [6, 9, 20].

Unlike Ethereum, IBM Blockchain does not have any currency or gas price, hence, we deal with a subset of the controlling factors, namely, *tps* and *n*, leading to 12 distinct setups.

4.5 Workload Drivers

Since the LCaaS receives its incoming data through API, we use Postman [63] to generate incoming transactions (i.e., log files) to the LCaaS. Using the *Runner* function of Postman, one can set the number of iterations and delay for each submitted API calls. For example, the number of iterations set to 200 and the delay of 1000 milliseconds will send 200 transactions to LCaaS at the rate of 1 *tps*.

²Suggested gas prices as per [62], gathered on Sep. 10, 2018 were as follows. For SafeLow tier executed in (< 30 minutes) — 6 gwei; for Standard tier (< 5 minutes) — 9 gwei, and for Fast tier (< 2 minutes) — 20 gwei.

5 Results

In Section 5.1, we discuss the performance of the the LCaaS itself, comparing the processing time for each block type. In Sections 5.2 and 5.3, we discuss integration performance with Ethereum and IBM Blockchain, respectively.

5.1 LCaaS performance test analysis

LCaaS component, integrated with Ethereum and IBM Blockchain, is kept identical in both integration scenarios (so that we can fairly assess the performance of each integration). That is, the number of transactions, the length of circled blockchains, and the size of submitted logs to LCaaS are identical in both integration scenarios. Therefore, the only difference between the two setups, from the LCaaS point of view, is the blockchain vendor that is chosen as the endpoint for submissions of super blocks.

Internally, we track the time needed to create a block of each block type (discussed in Section 3). The results of the timing for each block type are listed in Figure 3 and Table 2. As expected, by construction, the creation of internal blocks on LCaaS (namely, AGB, DB, RGB, and TB) is much faster than of the SB, which has to be submitted to blockchain vendors. Internal blocks creation time ranges³ between 10^{-6} and 0.38 seconds, while creation of the SB — from 10.17 seconds to 23 minutes for Ethereum and from 0.78 to 3.63 seconds for IBM Blockchain.

The time to create an internal block depends mainly on the *difficulty_target* and the processing speed of a computer on which the test is performed. This is why the processing times for all types of internal blocks are similar. Moreover, eyeballing of Figure 3 suggests that the processing time of the AGB, DB, RGB, and TB blocks remains similar, independent of the values of *tps* and *n*. This suggests that our test bed is capable to absorb both low- and high-intensity workloads without reaching the saturation.

To create a super block, one needs to parse and fetch all of the field of a terminal block of a circled blockchain and send it out to an external vendor, hence the additional processing time. Let us look at the SB submission time in details.

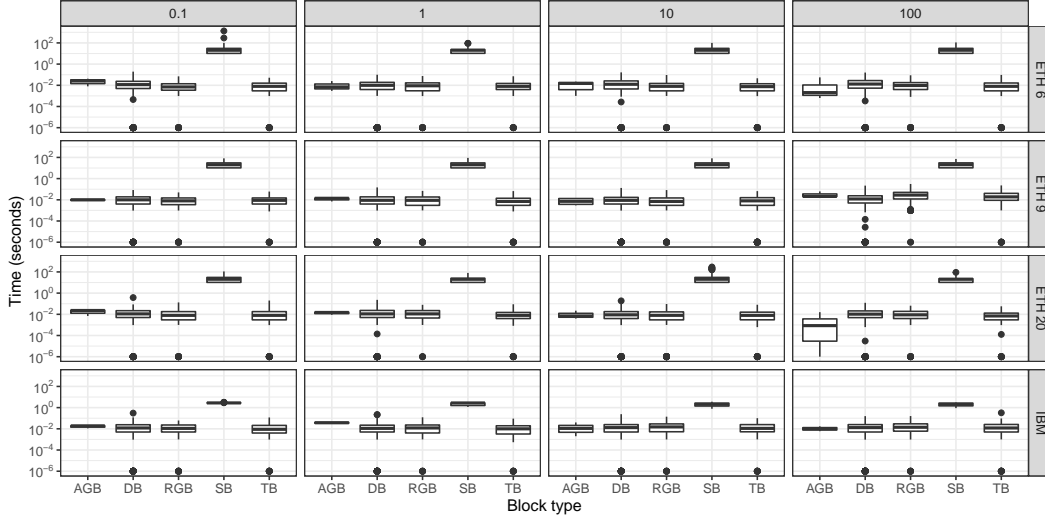


Figure 3: Processing time of different types of blocks for different experiments. Columns portray different values of *tps*, namely, 0.1, 1, 10, and 100. Rows show different gas prices in the case of Ethereum LCaaS implementation (denoted ETH), namely, 6, 9, and 20, and no gas price in the case of IBM LCaaS implementation (denoted IBM).

³Technically, the 10^{-6} should be interpreted as $\leq 10^{-6}$, as 10^{-6} is the smallest duration that we could measure.

Table 2: Summary statistics for processing time of different types of blocks. The columns of ‘1st qu.’ and ‘3rd qu.’ denote first and thirds quantiles, respectively. SB (ETH) and SB (IBM) denote super blocks submitted to Ethereum and IBM Blockchain, respectively.

Block	min	1st qu.	median	mean	3rd qu.	max
AGB	1E-6	0.01	0.01	0.02	0.02	0.06
DB	1E-6	0.00	0.01	0.02	0.02	0.38
RGB	1E-6	0.00	0.01	0.02	0.02	0.31
TB	1E-6	0.00	0.01	0.01	0.02	0.32
SB (ETH)	10.17	10.33	20.34	23.10	30.44	1353.37
SB (IBM)	0.78	1.62	2.72	2.28	2.79	3.63

5.2 Ethereum integration performance test analysis

We conduct 36 experiments (one for every permutation of the values of the factors listed in Table 1).

To see whether the performance will be affected by n , g and tps , we performed Pearson and Spearman correlation analysis as well as linear regression analysis on the raw data (i.e., per SB timing), the mean, the median, and the 95th percentile timing⁴ of SBs for each experiment. We found that none of the factors or the composite factors have any statistically significant relation to the response times, based on the low (< 0.15) values of correlations and high (> 0.1) p -values of linear models. This implies that the time needed to process SB block is dependent mainly on external factors, e.g., saturation of the Ethereum network and availability of the miners.

We show a distribution of processing times for SB block in Figure 4. Eyeballing of the distributions suggests that the lower the gas price is, the more SB blocks have higher processing time (> 32 seconds), even though the difference is not dramatic. Based on Kolmogorov-Smirnov test, the distribution of $g = 20$ case differs significantly (p -value < 0.001) from the cases when $g = 9$ or $g = 6$. However, the difference between $g = 6$ and $g = 9$ cases is less pronounced: p -value ≈ 0.08 . We were anticipating a stronger difference between all three cases; probably our usage of the test network rather than a production one lead to this anemic difference.

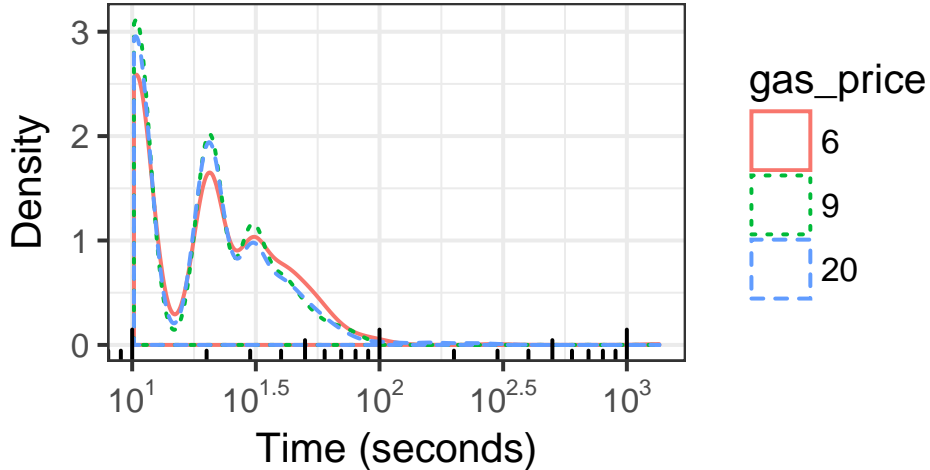


Figure 4: Density estimate of the processing time for SB block on Ethereum test network.

The Ethereum production network strives to keep the average processing time of a transaction at ≈ 15 seconds [50]. As shown in Table 2, we observe an average of 23 seconds and the median of 20 seconds, making it close to the target (even on the test network). Essentially, our findings show that the network has enough capacity to “absorb” the changes in our workload even in the intense cases, such as $tps = 100$ and

⁴Aggregate statistics are chosen to reduce the amount of noise in the data.

$n = 1$. However, in rare cases, the processing time is high: out of 3089 processed SBs, 5 (0.16%) had been processed in between 3 and 5 minutes, and 1 (0.03%) in 23 minutes. As we can see, these cases are rare, but they do exist and we have to be aware of such events. From practical perspective, if LCaaS workload deals with sensitive data, one may setup an alerting mechanism that will notify stakeholders of such events.

Our tests show the overall feasibility of the approach: the smart contract required to submit an SB is low enough ($\approx 335K$ units of gas, as shown in a sample SB submission [51]) to fit into an Ethereum block (i.e., it can be executed on the production network). We also know that the production Ethereum network is scalable (currently handling ≈ 1.2 million transactions per day [64]). Thus, adding hundreds or thousands of SB submissions per day will not saturate the network.

The processing time is strongly driven by gas prices on the production network, as shown empirically by [62]. We will further discuss gas prices and their effect on the cost of ownership in Section 6.4.1—onward.

5.3 IBM Blockchain integration performance test analysis

As the concept of gas price does not apply to the IBM Blockchain integration, our controlling factors and their subsequent scenarios are limited to 12 experiments (one for every permutation of the values of the factors listed in Table 1).

For each experiment, we have introduced submission timers for each block that is submitted to the IBM Blockchain. To measure the most relevant timer, we have enabled this timer on the Hyperledger Node.js component of the IBM Blockchain integration (see Figure 2).

As shown in Figure 3 and Table 2, the submission time of an SB ranges between 0.78 and 3.63 seconds. Kolmogorov-Smirnov test shows that *tps* and n affect the duration of the submission time. However, from practical perspective, their impact is minor as the submission time, even in the worst-case scenario is less than 4 seconds.

For the IBM Blockchain, the main factor controlling the SB processing time is the performance of underlying hardware running the IBM Blockchain. Furthermore, as IBM Blockchain is implemented over Kubernetes containers, additional clustering configuration can be used to scale up the Blockchain platform performance, if needed. The results published by the IBM Blockchain team show that the platform can handle 128 peers and 325 channels using LevelDB ledger and achieve around 13K *tps* [65].

6 Discussion

In the previous section, we have analyzed the results of integration between LCaaS and two blockchain vendors, namely, Ethereum and IBM Blockchain. We choose these two vendors on purpose as each one represents a different and vital type of blockchain. The private blockchains are designed for businesses that want to employ blockchain technologies but without a publicly accessible ledger. While this can be an applicable scenario for many businesses, many others, such as online asset tracking systems, need publicly available ledgers, hence the public blockchains.

Which solution is the best? As usual, it depends on various aspects of a particular use-case. Below, we will discuss these aspects. Specifically, implementation of the LCaaS part of our solution in production is discussed in Section 6.1, security — in Section 6.2, timing — in Section 6.3, and the cost of ownership — in Section 6.4.

6.1 Platform for CBs

We built a prototype implementing core elements (covering CBs) of the LCaaS [14] ourselves. However, for the production implementations, we recommend that practitioners build their solution on top of the enterprise-grade blockchain services, such as the IBM Blockchain that was used in this work, or similar solutions, such as AWS Blockchain [66] and Azure Blockchain [67]. For the private implementation, one can use one of the Hyperledger frameworks, such as Hyperledger Fabric [68], and build a hierarchical ledger on top of it. Furthermore, public blockchain services, such as Ethereum, can be used but may not be financially feasible for a large number of logs. Essentially, one will need to create a single blockchain for Super blockchain and additional ones for each of the CBs. Note that a personal instance of the Ethereum

network can be deployed on the Cloud. For example, Microsoft Azure Cloud provides a template that can create the infrastructure needed to deploy components needed for the creation of the network [69].

6.2 Security

6.2.1 Log submission intensity

How should we set a policy of submitting the logs to the LCaaS? Two control variables at our disposal are the time interval between submissions to the LCaaS and the maximum length (or the number of records) that a log file has to reach before it gets submitted to LCaaS. Intuitively, if we submit a log file to LCaaS every week, then it will give a perpetrator sufficient time to alter this log. Submitting logs every second (or every time a new log record comes in) will mitigate this risk, but will probably be economically infeasible and may also lead to scalability challenges.

What is the “goldilocks zone” then? The answer will depend on the nature of the logs. Log files with sensitive information, such as security and audit logs (e.g., recording event of a user logging into a system) may have to be submitted to LCaaS individually, i.e., a log file would contain a single log record. Typically, the intensity of arrival of such events is low and will not overwhelm computational resources and budget. Moreover, these are the types of records that an analyst may want to preserve as-is, without hashing them (assuming this does not violate confidentiality). Log files with less sensitive information, e.g., the operational logs containing performance metrics, can be submitted to the system every six hours (or sooner in case the log gets full and is truncated and archived by an operating system, based on policies set by IT personnel). In this case, four transactions per day will be submitted by a single logger, making archiving scalable and budget-friendly. Note that these are the types of records that are good candidates for hashing rather than storing in the raw format.

6.2.2 Public vs. private blockchain

The security of blockchain implementations is mainly dependent upon the security of underlying software and hardware as well as the protocols and settings required for the blockchain to function [70].

A public blockchain is more decentralized, with a large number of participating nodes. In contrast, a private blockchain is more centralized, and is designed to be used by one or more groups of users with a common goal and inherently, a fairly smaller number of participating nodes. The number of participating nodes, the consensus protocol in place, and the type of implementation play a significant role in security aspects of blockchain-based solutions [71]

6.3 Timing

As we can see from Figure 3 and Table 2, the processing of a transaction on average takes two seconds on IBM Blockchain, while on the Ethereum network, it takes at least ten seconds.

For IBM Blockchain, we are using the production network and thus can readily use these statistics. However, in the case of Ethereum, we are using the test network. For a proper comparison, we need to assess the time required to sign a transaction on the Ethereum blockchain.

The timing on the production network will depend on multiple factors, such as the number of miners and the size of their computing resources, the intensity and the size of the transactions submitted to the network. All of these factors are outside of our control.

The Ethereum production network strives to keep the average processing time of a transaction at ≈ 15 seconds [50]. We can treat it the lowest time limit to process and sign a block. In practice, the lowest threshold is a bit higher, as a user needs to spend time to send the transaction to the pool of the transactions and then distribute the transaction between the miners. Thus, signing an SB is faster on the IBM Blockchain than on the Ethereum Network. Let us now look at the cost of ownership of these two solutions.

6.4 Cost of ownership

The costs associated with the creation and maintenance of the CBs will be the same, independent of the solution that we build to sign the SBs. Let us examine the costs of signing the SBs below.

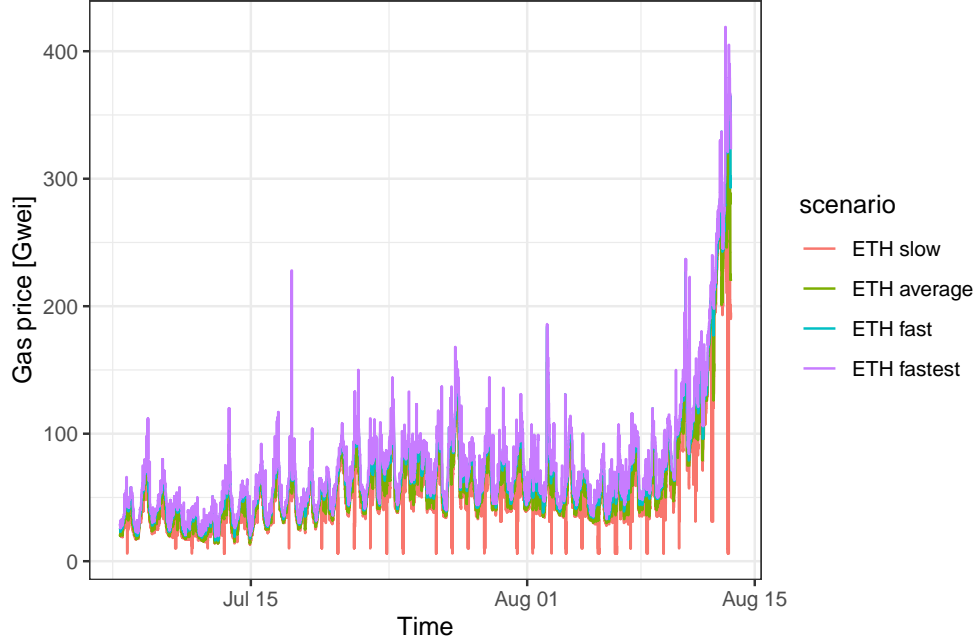


Figure 5: The price of gas required to sign the transaction on the production Ethereum blockchain under a certain time limit, based on [62]. The data are gathered for a ≈ 37 days period between 2020-07-06 22:35 UTC and 2020-08-13 12:50 UTC. Legend: *slow* — less than 30 minutes are required to sign the transaction, *average* — less than 5 minutes are needed to sign the transaction, *fast* — less than 2 minutes are required to sign the transaction, and *fastest* — less than 30 seconds are required to sign the transaction.

6.4.1 Ethereum Blockchain

In the case of Ethereum, SB submissions do not incur fixed cost, as we do not need any specialized infrastructure for Ethereum. However, the cost will depend on the number of transactions submitted (as we have to pay for every transaction) and the price of gas.

We know that the gas price, which a user sets, controls the speed with which a transaction is processed by a miner (the higher the price is, the faster it is executed). However, how does the gas price change with time? Let us explore the empirical data⁵ gathered from the Ethereum Gas Station service[62], shown in Figure 5. The figure shows the price of gas a user needs to set to sign the transaction in less than 30 seconds (*fastest*), 2 minutes (*fast*), 5 minutes (*average*), and 30 minutes (*slow*). The figure also shows that the faster the transaction is, the more we have to pay.

Our data analysis, based on the daily collection of data through the exposed API by ETH GAS Station [73], also shows that the prices exhibit daily seasonality and a non-zero trend. We also see that the gas daily price is at the lowest at 23:50 UTC.

Figure 6 and Table 3 show the daily costs of the SB submissions to the Ethereum blockchain. The details of the computations are given in Appendix E.2. As we can see, the prices will increase with the growth of the speed with which we would like the transaction to be processed, as well as the increase of intensity of the SB submissions. The daily cost of ownership ranges from \$4.07 USD for 1 transaction per day in the *slow* scenario to \$2226.10 USD for 1 transaction per 5 minutes in the *fastest* scenario.

6.4.2 IBM Blockchain

In the case of the IBM Blockchain, we need to pay for the compute and storage resources required to process the transactions. Let us use the reference architecture recommended by the manufacturer [74]. In this case, the solution consists of the costs needed to run IBM Cloud Kubernetes cluster, IBM Blockchain Platform,

⁵While ETH Gas Station predictions are not 100% accurate [72]; they are sufficient to illustrate the concept.

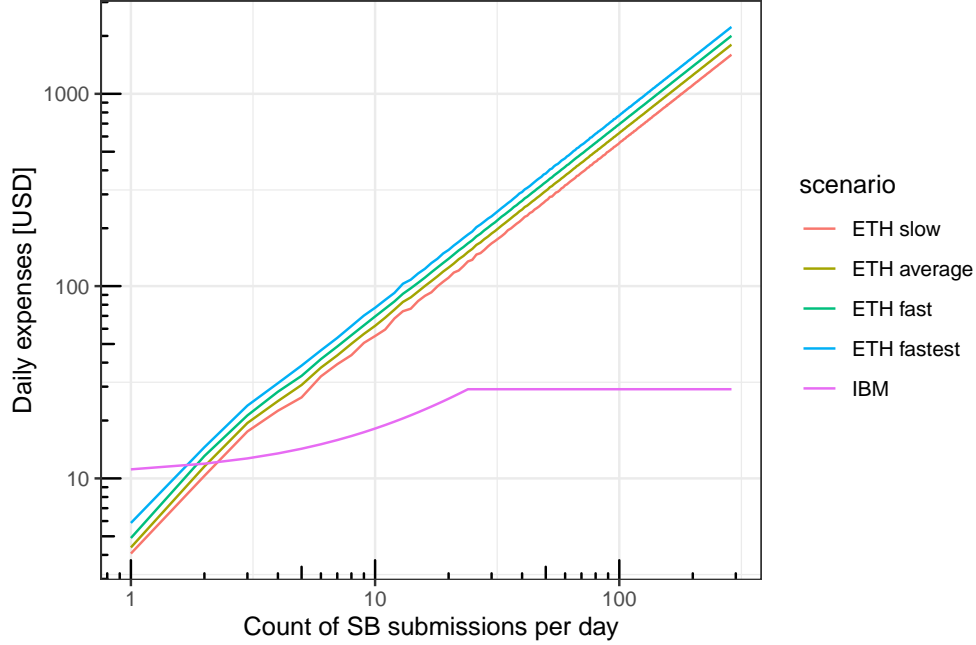


Figure 6: Daily cost of ownership. The intensity ranges from 1 SB submission per day to 288 submissions per day (i.e., one submission per five minutes).

Table 3: Daily costs for submitting SBs in USD

Submissions per day	IBM	ETH slow	ETH average	ETH fast	ETH fastest
1	11.14	4.07	4.37	4.90	5.86
2	11.92	10.31	11.54	13.07	14.52
3	12.70	17.54	19.41	21.26	23.91
...
288	29.09	1597.05	1800.87	2000.94	2226.10

and IP address allocation, as well as compute and storage resources, costing us $\approx \$1.21$ USD per hour. The cost can be reduced to $\approx \$0.43$ USD per hour by reducing computing resources. This reduction will be applicable only if the intensity of SB submission is higher than 1 SB per hour (as the resources are billable by the hour).

The cost per day will depend on the intensity of submissions of SBs and is shown in Figure 6 and Table 3. The details of the computations are given in Appendix E.1. As expected, the more transactions are submitted, the higher the cost (as we need a higher amount of computing resources): the daily cost of ownership ranges from \$11.14 USD to \$29.09 USD for 1 transaction per day and per 5 minutes, respectively.

6.4.3 IBM Blockchain vs. Ethereum blockchain

To reiterate, the cost of ownership will depend on the intensity of the submissions of SBs. When the intensity is low, Ethereum is cheaper. When the intensity increases, IBM Blockchain becomes more economical. Quantitatively, Ethereum blockchain will be more economical if the number of SB submissions per day will be less than 2 (for the *fast* and *fastest* scenarios) and less than 3 (for the *average* and *slow* scenarios). These thresholds may change in the future (as the underlying costs are non-constant). In contrast, the underlying cost associated with running LCaaS and producing CBs is constant (the sum of computation cost for mining and the cost for storing CBs).

Moreover, the IBM Blockchain can be used by multiple workloads. Thus, an organization may reduce the cost per workload by sharing blockchain infrastructure.

Note that we do not consider the overhead associated with the maintenance of the IBM Blockchain infrastructure. Also, the ETH–USD exchange rate is volatile (in the last 12 months it ranged between $\approx \$111$ USD and $\approx \$400$ USD for 1 ETH [75]), adding additional currency conversion risks.

To summarize, the pros and cons are as follows. Ethereum implementation is decentralized, has higher redundancy, and a higher degree of trust (due to the large number of miners involved) than the IBM Blockchain. However, its budget is less predictable (due to fluctuation in gas price and ETH–USD conversion rate). Moreover, if the number of transactions is high, the Ethereum-based LCaaS may become prohibitively expensive.

7 Threats to Validity

In this section, we discuss the threats that affect validity structured as per [76, 77].

Internal validity. Our integration of LCaaS with the blockchain vendors may be suboptimal and should be treated as a proof of idea rather than production-grade integration. Nevertheless, we were able to achieve strong performance results.

We conduct performance testing on the test network of Ethereum rather than the production network. The experiments on the test network show the computational feasibility of our approach (i.e., a transaction with SB can be submitted to a block). To extrapolate the timing results to the production network, we perform additional analysis based on the empirical data for gas prices and timing.

In the case of the IBM Blockchain integration test, we have used the trial version of IBM Blockchain that is running on the lite infrastructure with shared resources and declared limitations [78]. The performance of the non-trial infrastructure may be higher due to more powerful resource allocation [61].

Construction validity. Manual gathering of performance data may lead to errors; thus, we automate the execution and data gathering of the experiments to minimize the risk of human errors.

Statistical validity. We complete several statistical tests to analyze potential factors that affect the performance of the LCaaS and its integration with blockchains. The conclusions are drawn based on the statistical validity of the tests.

External validity. It is important to mention that while we have introduced additional components to LCaaS (namely AGB, RGB, TB, SB, and SBC), we have not altered the key elements of the actual blocks, so that LCaaS can be easily integrated with any other blockchain, which is proven by the successful integration of LCaaS to Ethereum and IBM Blockchain. With minor modifications, the proposed integration architecture can be used for integration between LCaaS and any other blockchain networks that supports smart contracts. Thus, LCaaS and its integration with Ethereum and IBM Blockchain can be treated as “critical cases” (in a case study sense [77]) showing the feasibility of our approach.

An additional threat to the validity is that one could ask why the CSCs who do not fully trust CSPs, would trust the logs provided by them? We argue that the tampering attempts typically happen after a client complains about the service. If the time difference between the submission time of logs and the rendered service is minimal, from milliseconds to a few seconds, the window of time at which the provider can tamper with the logs is narrow.

8 Conclusion

We described Cloud-based immutable log storage solution, LCaaS, based on the blockchain technology. This solution prevents log tampering, ensuring a transparent logging process and establishing trust between all Cloud participants (providers and users). Thus, the solution is of interest to practitioners. We provided detailed design of the solution and showed the implementation of the LCaaS on a public blockchain (Ethereum) and a private blockchain (IBM Blockchain). Performance test suggests that the solution is scalable (dealing with 100 *tps*) and is capable of fast “sealing” of the records (from seconds to minutes, depending on the implementation). This work is also of interest to academics, as it describes building blocks, which may be leveraged in a general scalable and immutable storage platform, leading to novel solutions for storing data.

The proposed LCaaS can act as a hierarchical ledger and a repository for all logs generated by Cloud solutions and can be accessed by all Cloud participants (namely, providers and consumers) to establish trust among them. Using verification services, a Cloud user can verify the Cloud provider's logs against the records in the hierarchical ledger and finds out if the logs were tampered with or not.

In the future, we plan to test LCaaS with other existing blockchain solutions to find integration points that can be used to implement LCaaS on top of such solutions. Additionally, we would like to review the capacity and scalability challenges of blockchain and define key parameters that affect the LCaaS performance and cost. At the final state, we would like to use such parameters to design a framework that can reduce the total cost of ownership of an end-to-end solution for the secure storage of logs using blockchains.

Acknowledgment

We wish to thank the funding agencies profusely. This research is funded in part by IBM Center for Advanced Studies grant No. 1046, NSERC Discovery Grant No. RGPIN-2015-06075, and NSERC CRD Grant CRDPJ 538493-18.

References

- [1] R. Accorsi, "Log data as digital evidence: What secure logging protocols have to offer?" in *33rd Annual IEEE Int. Computer Software and Applications Conference, COMPSAC'09.*, vol. 2. IEEE, 2009, pp. 398–403.
- [2] S. R. Selamat, R. Yusof, and S. Sahib, "Mapping process of digital forensic investigation framework," *International Journal of Computer Science and Network Security*, vol. 8, no. 10, pp. 163–169, 2008.
- [3] D. Reilly, C. Wren, and T. Berry, "Cloud computing: Forensic challenges for law enforcement," in *Int. Conference for Internet Technology and Secured Transactions (ICITST)*. IEEE, 2010, pp. 1–7.
- [4] A. Chauhan, O. P. Malviya, M. Verma, and T. S. Mor, "Blockchain and scalability," in *2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*. IEEE, 2018, pp. 122–128.
- [5] W. Pourmajidi and A. Miranskyy, "Logchain: Blockchain-assisted log storage," in *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*. IEEE, 2018, pp. 978–982.
- [6] W. Pourmajidi, L. Zhang, J. Steinbacher, T. Erwin, and A. Miranskyy, "Immutable log storage as a service," in *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*. IEEE, 2019, pp. 280–281.
- [7] "Ethereum project," <https://www.ethereum.org/>.
- [8] "Ibm blockchain - enterprise blockchain solutions & services — ibm," <https://www.ibm.com/blockchain>.
- [9] H. Shafagh, L. Burkhalter, A. Hithnawi, and S. Duquennoy, "Towards blockchain-based auditable storage and sharing of iot data," in *Proceedings of the 2017 on Cloud Computing Security Workshop*, 2017, pp. 45–50.
- [10] Y. Chen, S. Ding, Z. Xu, H. Zheng, and S. Yang, "Blockchain-based medical records secure storage and medical service framework," *Journal of medical systems*, vol. 43, no. 1, p. 5, 2019.
- [11] Q. Xu, K. M. M. Aung, Y. Zhu, and K. L. Yong, "A blockchain-based storage system for data analytics in the internet of things," in *New Adv. in the Internet of Things*. Springer, 2018, pp. 119–138.
- [12] J. Kelsey and B. Schneier, "Minimizing bandwidth for remote access to cryptographically protected audit logs," in *Recent Advances in Intrusion Detection*, 1999, pp. 9–9.
- [13] B. R. Waters, D. Balfanz, G. Durfee, and D. K. Smetters, "Building an encrypted and searchable audit log," in *NDSS*, vol. 4, 2004, pp. 5–6.

- [14] “Logchain source code,” <https://github.com/WilliamPourmajidi/LCaaS>.
- [15] H. Farid, “Image forgery detection,” *IEEE Signal processing magazine*, vol. 26, no. 2, pp. 16–25, 2009.
- [16] “Digital forensics research workshop,” <https://www.dfrws.org/>.
- [17] R. Marty, “Cloud application logging for forensics,” in *Proceedings of the 2011 ACM Symposium on Applied Computing*, ser. SAC ’11. New York, NY, USA: ACM, 2011, pp. 178–184.
- [18] I. Ray, K. Belyaev, M. Strizhov, D. Mulamba, and M. Rajaram, “Secure logging as a service—delegating log management to the cloud,” *IEEE systems journal*, vol. 7, no. 2, pp. 323–334, 2013.
- [19] “arxiv.org e-print archive,” <https://arxiv.org/>.
- [20] S. Underwood, “Blockchain beyond bitcoin,” *Commun. ACM*, vol. 59, no. 11, pp. 15–17, Oct. 2016.
- [21] M. Mainelli and M. Smith, “Sharing ledgers for sharing economies: an exploration of mutual distributed ledgers (aka blockchain technology),” *J. of Financial Persp.*, vol. 3, no. 3, pp. 38–58, 2015.
- [22] S. Haber and W. S. Stornetta, “How to time-stamp a digital document,” in *Conf. on the Theory and Appl. of Crypt.* Springer, 1990, pp. 437–455.
- [23] B. Schneier and J. Kelsey, “Cryptographic support for secure logs on untrusted machines.” in *USENIX Security*, vol. 98, 1998, pp. 53–62.
- [24] R. Matzutt, M. Henze, J. H. Ziegeldorf, J. Hiller, and K. Wehrle, “Thwarting unwanted blockchain content insertion,” in *IEEE Int. Conf. on Cloud Engineering (IC2E)*. IEEE, 2018, pp. 364–370.
- [25] “Poex.io - the original blockchain notary service,” <https://poex.io/>.
- [26] P. Gallagher, “Secure hash standard (shs),” *FIPS PUB*, pp. 1–27, 2008.
- [27] W. Pourmajidi, J. Steinbacher, T. Erwin, and A. Miranskyy, “On challenges of cloud monitoring,” in *Conf. of the Center for Adv. Studies on Collab. Research (CASCON)*. IBM, 2017, pp. 259–265.
- [28] Z. Zheng, S. Xie, and H.-N. Dai, “Blockchain challenges and opportunities: A survey,” 2016.
- [29] J. Bruce, “The mini-blockchain scheme,” 2014. [Online]. Available: <http://cryptonite.info>
- [30] I. Eyal, A. E. Gencer, E. G. Sirer, and R. Van Renesse, “Bitcoin-ng: A scalable blockchain protocol.” in *NSDI*, 2016, pp. 45–59.
- [31] L. M. Bach, B. Mihaljevic, and M. Zagar, “Comparative analysis of blockchain consensus algorithms,” in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018, pp. 1545–1550.
- [32] S. Bouraga, “A taxonomy of blockchain consensus protocols: A survey and classification framework,” *Expert Systems with Applications*, vol. 168, p. 114384, 2021.
- [33] S. Kim, Y. Kwon, and S. Cho, “A survey of scalability solutions on blockchain,” in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2018, pp. 1204–1207.
- [34] M. K. Pawar, P. Patil, and P. Hiremath, “A study on blockchain scalability,” in *ICT Systems and Sustainability*. Springer, 2021, pp. 307–316.
- [35] S. M. H. Bamakan, A. Motavali, and A. B. Bondarti, “A survey of blockchain consensus algorithms performance evaluation criteria,” *Expert Systems with Applications*, p. 113385, 2020.
- [36] P. Tasca and C. J. Tessone, “Taxonomy of blockchain technologies. principles of identification and classification,” *arXiv preprint arXiv:1708.04872*, 2017.

- [37] G. Wood, “Ethereum: A secure decentralised generalised transaction ledger,” *Ethereum project yellow paper*, vol. 151, pp. 1–32, 2014.
- [38] N. Szabo, “Nick szabo – the idea of smart contracts,” <http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/idea.html>.
- [39] “Solidity — contract-oriented language,” <http://solidity.readthedocs.io/en/v0.4.24/>.
- [40] “Remix — solidity ide,” <https://remix.ethereum.org>.
- [41] “Blockchain use cases - ibm blockchain,” <https://www.ibm.com/blockchain/use-cases/>.
- [42] “Glossary,” <https://cloud.ibm.com/docs/blockchain?topic=blockchain-glossary#glossary-chaincode>.
- [43] “Ibm blockchain platform - visual studio marketplace,” <https://marketplace.visualstudio.com/items?itemName=IBMBlockchain.ibm-blockchain-platform>.
- [44] “Visual studio code - code editing. redefined,” <https://code.visualstudio.com/>.
- [45] “Developing smart contracts with visual studio code extension,” <https://cloud.ibm.com/docs/blockchain?topic=blockchain-develop-vscode>.
- [46] “Firebase,” <https://firebase.google.com/>.
- [47] “Account types, gas, and transactions — ethereum homestead 0.1 documentation,” <http://ethdocs.org/en/latest/contracts-and-transactions/account-types-gas-and-transactions.html>.
- [48] “Test networks — ethereum,” <http://ethdocs.org/en/latest/network/test-networks.html#details>.
- [49] “Test ether faucet,” <https://faucet.metamask.io/>.
- [50] “Ethereum average block time chart — etherscan,” <https://etherscan.io/chart/blocktime>.
- [51] “Ropsten transaction hash (txhash) details: Etherscan,” <https://bit.ly/3xMs9ie>.
- [52] “Pricing for ibm blockchain platform for ibm cloud,” <https://cloud.ibm.com/docs/blockchain?topic=blockchain-ibp-saas-pricing>.
- [53] “Ibm cloud — ibm,” <https://www.ibm.com/cloud>.
- [54] “Get started with blockchain – ibm developer,” <https://developer.ibm.com/technologies/blockchain/gettingstarted/>.
- [55] “Hyperledger – open source blockchain technologies,” <https://www.hyperledger.org/>.
- [56] “Hyperledger - open source blockchain for business - ibm blockchain — ibm,” <https://www.ibm.com/blockchain/hyperledger>.
- [57] “Github - hyperledger/fabric-sdk-py: Hyperledger fabric python sdk,” <https://github.com/hyperledger/fabric-sdk-py>.
- [58] “Build a network,” <https://cloud.ibm.com/docs/blockchain/howto?topic=blockchain-ibp-console-build-network#ibp-console-build-network>.
- [59] “Kubernetes service - overview — ibm,” <https://www.ibm.com/cloud/container-service/>.
- [60] “Blockchain platform pricing - canada ibm,” <https://www.ibm.com/ca-en/cloud/blockchain-platform/pricing>.
- [61] “Pricing for ibm blockchain platform for ibm cloud,” <https://cloud.ibm.com/docs/blockchain?topic=blockchain-ibp-saas-pricing#ibp-saas-pricing-default>.

- [62] “Eth gas station — consumer oriented metrics for the ethereum gas market,” <https://ethgasstation.info/>.
- [63] “Postman api development environment,” <https://www.getpostman.com/>.
- [64] “Ethereum daily transactions chart — etherscan,” <https://etherscan.io/chart/tx>.
- [65] “Does hyperledger fabric perform at scale? - blockchain pulse: Ibm blockchain blog,” <https://www.ibm.com/blogs/blockchain/2019/04/does-hyperledger-fabric-perform-at-scale/>.
- [66] “Blockchain on aws,” <https://aws.amazon.com/blockchain/>.
- [67] “Blockchain technology and applications — microsoft azure,” <https://azure.microsoft.com/en-us/solutions/blockchain/>.
- [68] Hyperledger Fabric a blockchain framework. [Online]. Available: <https://www.hyperledger.org/projects/fabric>
- [69] “Azure blockchain workbench,” <https://azuremarketplace.microsoft.com/en-ca/marketplace/apps/microsoft-azure-blockchain.azure-blockchain-workbench?tab=Overview>.
- [70] W. Gao, W. G. Hatcher, and W. Yu, “A survey of blockchain: techniques, applications, and challenges,” in *27th int. conf. on computer comm. and networks (ICCCN)*. IEEE, 2018, pp. 1–11.
- [71] C. C. Agbo and Q. H. Mahmoud, “Comparison of blockchain frameworks for healthcare applications,” *Internet Technology Letters*, vol. 2, no. 5, p. e122, 2019.
- [72] G. Antonio Pierro, H. Rocha, R. Tonelli, and S. Ducasse, “Are the gas prices oracle reliable? a case study using the ethgasstation,” in *2020 IEEE International Workshop on Blockchain Oriented Software Engineering (IWBOSE)*, 2020, pp. 1–8.
- [73] “Eth gas station api overview - eth gas station api,” <https://docs.ethgasstation.info/>.
- [74] “Pricing for ibm blockchain platform for ibm cloud,” <https://cloud.ibm.com/docs/blockchain?topic=blockchain-ibp-saas-pricing#ibp-saas-pricing-scenarios>.
- [75] “Ethereum usd historical data — coingecko,” https://www.coingecko.com/en/coins/ethereum/historical_data/usd?end_date=2020-08-13&start_date=2019-08-13#panel.
- [76] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*, ser. Computer Science. Springer Berlin Heidelberg, 2012.
- [77] R. Yin, *Case Study Research: Design and Methods*, ser. Applied Social Research Methods. SAGE Publications, 2009.
- [78] “Benefits and service offerings,” https://cloud.ibm.com/docs/containers?topic=containers-cs_ov#cluster_types.
- [79] P. Mell, T. Grance *et al.*, “The nist definition of cloud computing,” 2011.
- [80] C. Decker and R. Wattenhofer, “Information propagation in the bitcoin network,” in *Peer-to-Peer Computing (P2P), 2013 IEEE Thirteenth International Conference on*. IEEE, 2013, pp. 1–10.
- [81] P. Rogaway and T. Shrimpton, “Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance,” in *International workshop on fast software encryption*. Springer, 2004, pp. 371–388.
- [82] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC press, 1996.

- [83] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, “An overview of blockchain technology: Architecture, consensus, and future trends,” in *2017 IEEE international congress on big data (BigData congress)*. IEEE, 2017, pp. 557–564.
- [84] “Blockchain technology overview,” <https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8202.pdf>.
- [85] “Flask (a python microframework),” <http://flask.pocoo.org/>.
- [86] “Openssl dgst,” <https://www.openssl.org/docs/manmaster/man1/dgst.html>.
- [87] R. J. Hyndman and Y. Khandakar, “Automatic time series forecasting: the forecast package for R,” *Journal of Statistical Software*, vol. 26, no. 3, pp. 1–22, 2008. [Online]. Available: <http://www.jstatsoft.org/article/view/v027i03>
- [88] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O’Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen, *forecast: Forecasting functions for time series and linear models*, 2020, r package version 8.12. [Online]. Available: <http://pkg.robjhyndman.com/forecast>

A Log tampering: examples

Below, we provide three examples of situations where there may be motivation to tamper with logs related to private, community, and public Clouds.

Example 1. Case for private Cloud tamper-motivation In a private Cloud, a unique type of tamper-motivation may exist. Imagine a company that has established a private Cloud. The management team has requested from the Information Technology (IT) department a full second-by-second backup for all of the company’s financial data. The IT department configures its backup systems and ensures that there is enough space for continuous backup of transactions. Months after the initial setup, the company’s primary storage is affected by a hardware failure. The IT department finds out that the real-time backup system has stopped working a few days ago and had sent several alerts that no one in the IT department noticed. The IT department is the only department that has access to all the logs. Thus, the department may be motivated to tamper with the logs to cover up the problem.

Example 2. Case for community Cloud tamper-motivation A community Cloud [79] requires a clear definition of responsibilities for each partner. In case of an unfortunate incident, the party at fault may be motivated to tamper with the logs that identify them as the party responsible for the issue. Even worse, they may try to tamper with the logs and fabricate a scenario where another party becomes the main reason behind failure. Having access to the logs for one or more of the parties in a community Cloud may cause trust issues that call for the logs’ immutability.

Example 3. Case for public Cloud tamper-motivation As for the public Cloud, the platform’s operational health, performance, generated metrics, and even charge-back reports (that are consolidated in the form of monthly invoices to CSCs) are entirely managed by CSPs. Without having access to the actual logs or the actual infrastructure, CSCs of the public Clouds are in a very unfair, dependent position. Consider a scenario in which a CSC deploys an application on an elastic Cloud environment with an auto-scaling feature and defines a rule that when the memory usage exceeds 80%, the CSP should allocate 20% extra memory space to the deployed application. Imagine that the CSC receives complaints related to the application performance from its users. The CSC asks the CSP to send a detailed report of the elastic memory allocation. The CSP’s IT team checks their logs and finds out that the auto-scaling feature has worked intermittently, hence the performance issue. If the CSC finds out the truth, there may be a lawsuit on the horizon. Thus, the IT team may be motivated to tamper with the log before sending it to the CSC.

B Background

We introduce some basic components and characteristics of blockchains in this section. In Section B.1, we present a brief overview of common components of all blockchains and describe what each component does. In Section B.2, we provide an overview of the mining and immutability of blocks in a blockchain.

B.1 Common Key Components of Blockchains

Here, we introduce the components that are common among all implementations of the blockchain.

Genesis Block (GB): Genesis block is the first block of any blockchain. Genesis block has predefined characteristics. Its *index* and *previous_hash* are set to zero, as there are no prior blocks. The primary purpose of a genesis block is to indicate the start of a new blockchain [80].

Data Block (DB): A data block, more commonly known as a block, contains the following variables: *index*, *timestamp*, *data*, *current_hash*, and *previous_hash*. The first element, *index*, is a unique sequential ID for each block; it uniquely identifies each block. The *timestamp* indicates the time at which the block is created and is usually stored in the Coordinated Universal Time format. The *data* is the most important element of a data block. It contains valuable information that blockchain has promised to be immutable. The *nonce* is an arbitrary random number that is used to generate a specific *current_hash*. To achieve hash binding, each block includes a *previous_hash* element. The *previous_hash* is the exact duplicate of the *current_hash* of

Algorithm 1: Generation of *current_hash* and *nonce* for a block.

Input : *block_index*, *timestamp*, *data*, *previous_hash*
Output: *current_hash*, *nonce*

```

1 content = concatenate(index, timestamp, data, previous_hash);
2 content = Hasher(content); // to speedup computing
3 nonce = 0;
4 repeat
5   | nonce = nonce + 1;
6   | current_hash = Hasher( concatenate(nonce, content) );
7 until prefix of current_hash = difficulty_target;
8 return current_hash, nonce;
```

the previous block. In other words, the *current_hash* of the m -th block becomes the *previous_hash* of block $m + 1$. We use SHA-256 to generate *current_hash* and *nonce* for a block as illustrated in Algorithm 1.

Blockchain (BC): Blocks that are linked together via hash binding will result in a blockchain. If data in an earlier block (say, block m) are tampered, the link among all the subsequent blocks, from $m + 1$ to the most recent block i will be broken. Then one has to recompute *current_hash* and *nonce* values of each block from block m to block i of the BC.

B.2 Mining and Hash Binding

Relying on key characteristics of cryptographic hash function [81], blocks are cryptographically sealed and mined. Mining is the process running through all possible values of an integer variable⁶, known as *nonce*, to find a value for *nonce*, such that if the value is added to the rest of elements of a block and then hashed, the hash matches the imposed *difficulty_target*. Once the desired value for *nonce* is found, it resides in the *nonce* element of the block, and the calculated hash resides in the *current_hash* element of a block. At this stage, the block is mined. The *difficulty_target* is often defined as the number of required zeros at the beginning of the desired hash. The more zeros, the more computational power is needed to generate a hash that matches the difficulty target. Blocks are linked together based on a hash binding relationship.

Once a block is mined, its content can no longer change unless the whole PoW process is repeated for every block in the blockchain. In addition to hash binding, blockchains take advantage of the hash function basic properties, namely first and second preimage resistance and collision resistance, which make it extremely difficult to tamper with the hash values (ensuring their uniqueness), see [82] for details.

In addition to hash binding, blockchains take advantage of the hash function properties. Most cryptographic hash functions are designed to take an input of any size and produce a fixed-length hash value. Menezes et al. [82] indicate the following three basic properties of a hash function h with inputs x or x' and outputs y or y' .

1. “Preimage Resistance: for all predefined outputs, it is computationally infeasible to find any input which hashes to that output, i.e., to find any preimage x' such that $h(x') = y$ for any y for which a corresponding input is not known. In other words, for a given hash, it would be computationally infeasible to reverse the hash function and find the value that was hashed.” [82]
2. “Second preimage resistance: it is computationally infeasible to find any second input which has the same output as any specified input, i.e, given x , to find a second preimage $x' \neq x$ such that $h(x) = h(x')$.” [82]
3. “Collision resistance: it is computationally infeasible to find any two distinct inputs x and x' which hash to the same output, i.e., such that $h(x) = h(x')$. Collision resistance implies second preimage resistance but does not guarantee preimage resistance.” [82]

⁶Typically implemented as an unsigned integer.

Table 4: Differences between Private and Public blockchains

	Public	Private
Network	Decentralized	Centralized
Security	Open Network	Approved Participants
Access	Permission-less	Permissioned
Identity	Anonymous or Pseudonymous	Known Identities
Speed	Slower	Faster

B.3 Blockchain Taxonomies

Here we provide a summary of different types of blockchains and how they are compared. We will start by looking at the deployment models, including private and public blockchain and then move to the administrative models, including permissioned and permission-less blockchains.

B.3.1 Private and Public Blockchains

Blockchains, as tamper-evident and tamper-resistant distributed digital ledgers, can be deployed in various ways. They can be implemented privately, inside an organization, or can be launched publicly for a wider range of customers.

In a private blockchain, nodes who can participate are often part of the same organization that controls the centralized network and offer permissions to various players. In public blockchains, since they are open to the public, all records are visible to everyone, and the network is open for new nodes to join and participate. In contrast, a public blockchain is not limited to an organization and joining the network is open to public, so is the access to data stored on blocks.

In recent years, consortium blockchains have evolved as a hybrid model based on public blockchains. A consortium blockchain is very similar to a public blockchain except that only a group of pre-selected nodes would participate in the consensus process of the blockchain [83]. Table 4 lists the differences between private and public blockchains.

B.3.2 Permissioned and Permission-less Blockchains

As highlighted in NIST’s Blockchain Technology Overview [84], blockchain implementations are often inspired by a specific purpose or function. The purpose plays a significant role in finding and adopting the right model of blockchain. Blockchains have been categorized into two high-level categories: permission-less and permissioned (also known as permission-based). The permissionless blockchain allows any arbitrary user to interact with the blockchain and read and write blocks without the need of a central authority or getting approval from any party. In contrast, permission-based blockchains limit participants to specific people or organizations and provide a systematic authentication and authorization approach.

C LCaaS API

As reviewed in Section 1, to simplify the interaction with the LC and to allow its users to submit and verify logs easily, we introduce an API that converts the LC to the LCaaS. In other words, for each CSC, an instance of LC can be instantiated and can be used as a service through its API interfaces. Therefore, LC users do not need to integrate LC into their monitoring platform and can simply use it as a service via LCaaS API interfaces. The CSPs can efficiently use this API and interconnect the LCaaS with their monitoring systems and store all their logs, or the hash of their logs, in the Logchain. Similarly, CSCs can search and verify provided logs against the data in the Logchain and, therefore, be assured that the logs provided by the CSPs are not tampered with. However, the API is not provided to CSCs. In the current implementation, the application receives logs or their hashes, adds them to the data blocks and mines the blocks by finding a *nonce*. Like all other blockchains, our implementation links the blocks to their previous blocks by inserting the *current_hash* of the previous block into the *previous_hash* of the current block. Figure 7 illustrates

the sequence of actions withing the LCaaS API module from the submission of the logs to its verification. Additionally, we designate an actor for each interaction.

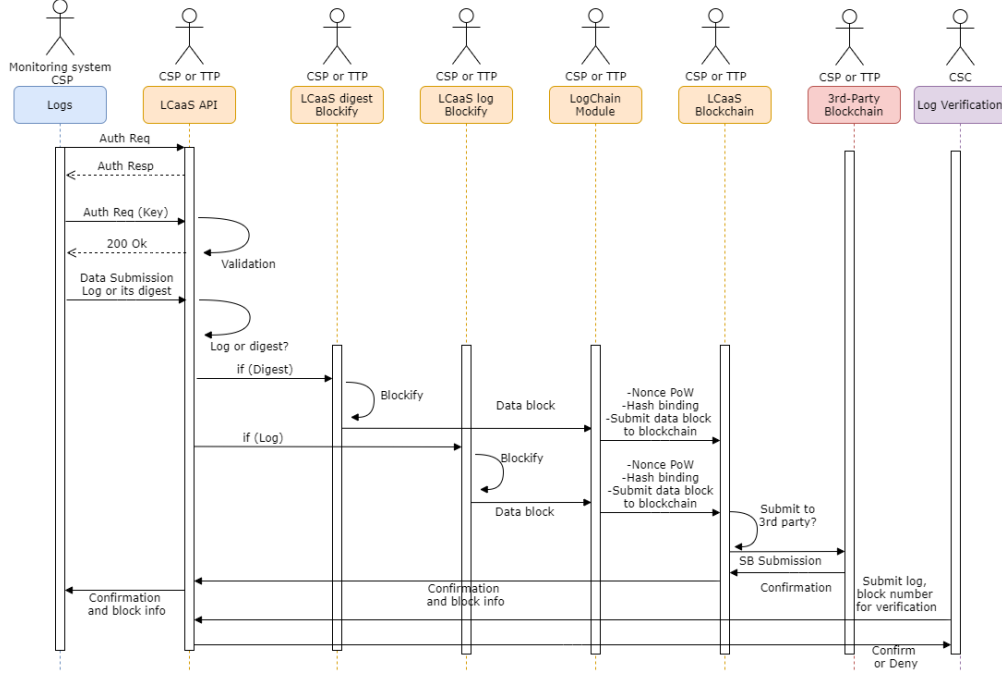


Figure 7: The Sequence Flow Diagram for LCaaS and its Internal Components

C.1 Submission Methods

The API is designed and implemented using Flask [85], a micro-framework for Python. There are two data submission methods: *submit_raw* and *submit_digest*. The former allows the client to submit the actual logs, while the latter — just the file’s digest (e.g., SHA-based digest computed using OpenSSL *dgst* [86]), thus, preserving the privacy of the log and reducing the amount of transmitted data. Both methods return, on success, *timestamp*, *block_index* and other details of the created block and, on failure, details of the error. For privacy reasons, the CSPs or CSCs may decide to generate a digest locally and submit it to LCaaS by using the *submit_digest* method. The Client of the LCaaS, have full control over the amount of log included in each submission and their interval, Therefore, LCaaS API can be really customized based on the presented scenario and the CSPs and CSCs needs.

C.2 Verification Methods

There are three verification methods: *verify_raw*, *verify_digest*, and *verify_tb*. The first one allows the client to verify the existence of actual logs in the LCaaS. The second one allows the client to verify the digest of the logs. The last method allows a user to verify the existence of a terminal block with a specified hash in the LCaaS, hence, proving the integrity of all the blocks in the circled blockchain of the submitted terminal block with one verification operation. All methods return, on success, the details of the found values in the LCaaS and, on failure, details of the error.

For the verification of the actual log content, one should use method *verify_raw*. The method would return the status of submission and number of blocks that match the submitted data; if no block is found, the API will return a message informing the user that no match has been found. In case of an error, the API will return the failed status along with the error’s description.

To improve the scalability of our solution, we introduce the *verify_tb* method. It provides an assurance (in the cryptographic sense [81]) that the sequence of blocks, from *index_from* to *index_to* are not tampered

with. By comparing the generated hash value from all the *current_hash* values of a circled blockchain to the *aggr_hash* value in the data element of a TB, one can verify the integrity of all the blocks in the circled blockchain.

It is important to mention that while we have introduced these additional components, we have tried not to alter the key element of the actual blocks. Avoiding any alteration on block's structure is intentional, because any modification in the blocks format (e.g., adding new elements) will result in a proprietary implementation of blocks and blockchains and will reduce the applicability of the proposed hierarchical structure to other existing blockchain platforms.

D Hierarchical Structure of LCaaS

To overcome blockchain performance issues, LCaaS uses a two-level hierarchy, but the number of levels can be increased if a use case requires it, making it a cascading blockchain. Figure 8 depicts this hierarchy and the internal relationship among its components.

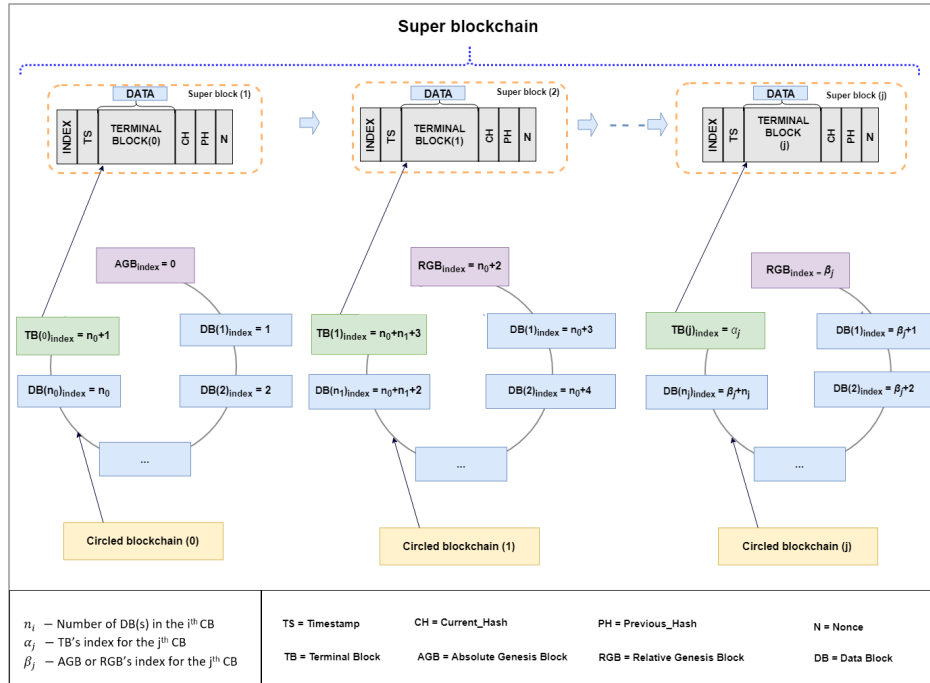


Figure 8: Two-level Hierarchy as implemented by LCaaS

As mentioned in the legend of Figure 8, n_i refers to the number of data blocks in the i -th circled blockchain and α_j is the index of the terminal block of the j -th circled blockchain. β_j is the index for the absolute or relative genesis block of the j -th circled blockchain. The value of α_j will be calculated by

$$\alpha_j = \begin{cases} n_0 + 1, & \text{if } j = 0 \\ n_0 + 1 + \sum_{i=1}^j (n_i + 2)n_i, & \text{if } j \geq 1 \end{cases} \quad (1)$$

and the value of β_j will be calculated by

$$\beta_j = \begin{cases} 0, & \text{if } j = 0 \\ \alpha_{j-1} + 1, & \text{if } j \geq 1 \end{cases} \quad (2)$$

Figure 9 shows the relationship between a terminal block and all other blocks in a circled blockchain. Figure 10 depicts the relationship between a TB and the data element of a SB.

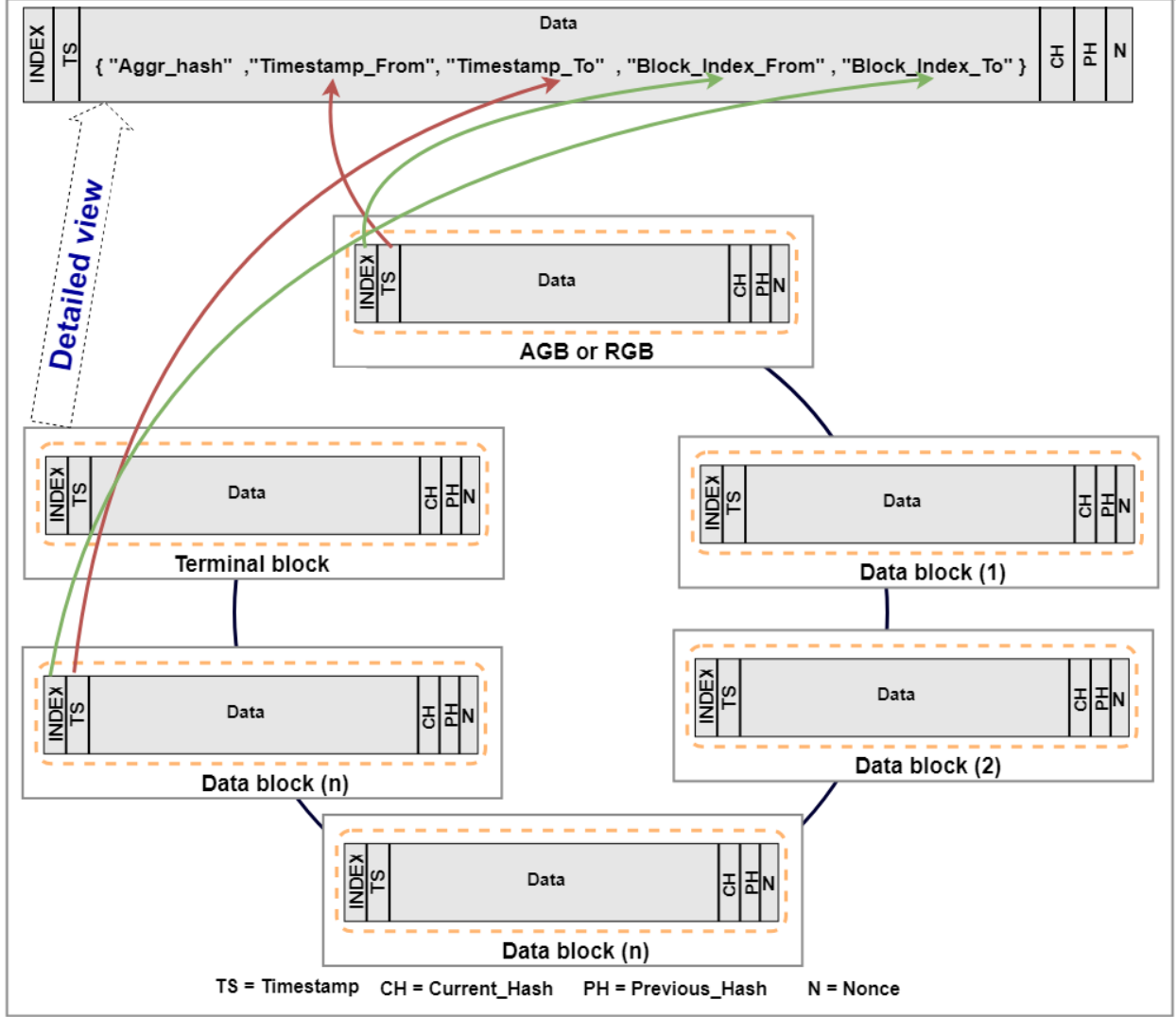


Figure 9: The Relation Between Terminal Block and All other Blocks in a Circled Blockchain

E Details of computing the cost of SB submission

Details of computing the cost of SB for IBM Blockchain are given in Appendix E.1, for Ethereum — in Appendix E.2.

E.1 IBM Blockchain

A typical small-scale, production-grade setup of the IBM Blockchain costs \$1.19 USD per hour [74]. The recommended setup requires three VMs (two for the peers to ensure high availability, and one to act as a Certificate Authority), an access to a production IBM Blockchain Platform, storage, and a Kubernetes cluster (to tie it all together).

To the variable costs above, we also need to add IP address allocation cost of \$16.00 USD per month (or $\approx 16.00/30/24 = \$0.02$ USD per hour. This will increase hourly cost to $\approx \$1.21$ USD per hour.

When the system is inactive, we can scale down the CPU allocation to almost zero to reduce the cost [74]. In this case, the hourly price will be reduced to \$0.43 USD per hour.

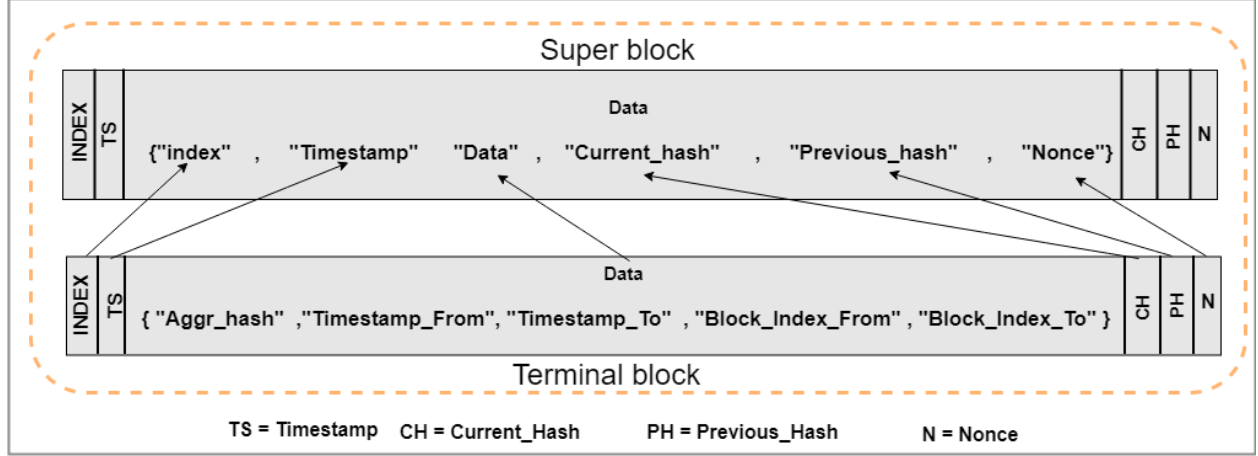


Figure 10: The Relationship Between Terminal Block and Super Block

Assuming that we would like to equidistribute SB submissions around the clock-dial, we can take the advantage of the cost reduction. The cost in this case will be computed as per Algorithm 2.

Algorithm 2: Computation of daily cost for IBM Blockchain.

Input : s ;	// SB count per day
Output : d ;	// Daily cost in USD
1 $c = 16.00/30/24$;	// IP address hourly cost
2 $f = 1.19 + c$;	// Full cost per hour
3 $r \approx 0.41 + c$;	// Reduced cost per hour
4 if $s < 24$ then	
5 $d = s \cdot f + (24 - s) \cdot r$;	
6 else	
7 $d = 24 \cdot f$;	
8 end	
9 return d ;	

E.2 Ethereum Blockchain

We have gathered Ethereum gas prices every 5 minutes for 37 days. The results are shown in Figure 5.

These empirical data are passed to Algorithm 3 to compute daily price for a given scenario and SB submission intensity. In a nutshell, the algorithm works as follows. Based on our analysis (example of gas prices time series decomposition is shown in Figure 11), the lowest ETH gas price is, consistently, at 23:50 UTC. Thus, this will be our starting time. Similar to the approach in Appendix E.1, we will equidistribute SB submissions around the clock-dial. For example, if we need to submit one SB per day the submission will happen at 23:50 UTC. In the case of two SB submissions, they will be sent off at 23:50 UTC and 11:50 UTC, in the case of three SB submissions — at 23:50 UTC, 7:50 UTC, and 15:50 UTC, and so on. For each scenario, we will then average out gas prices for a given minute of the day and sum them up. To convert this value into US dollars, we will need to multiply the average gas price by the number of units of gas needed to execute SB submission contract (namely, $\approx 335K$ gwei) and then convert gwei to USD. In our case, we will use the average ETH-USD exchange rate between 2020-07-06 and 2020-08-13 (i.e., the time interval when the empirical gas prices were collected), which was equal to 302.05 USD per 1 ETH [75] or 3.02×10^{-7} USD per 1 gwei.

Algorithm 3: Computation of daily cost for Ethereum Blockchain.

Input : s, r, p, c ; // SB count per day, the name of the Ethereum scenario, spot prices of gas in gwei, and gwei to USD conversion rate, respectively

Output: d ; // Daily cost in USD

- 1 Transform p to retain three columns for each observation: *minute_of_the_day*, *scenario_name*, *gas_price*;
 - 2 Retain the observations in p only for the scenario r ;
 - 3 Perform the following aggregation: “select *minute_of_the_day*, average(*gas_price*) from p group by *minute_of_the_day*”;
 - 4 For a given s , equidistribute the submission times, starting at 23:50 UTC and identify the closest *minute_of_the_day* to each submission time;
 - 5 z = Sum up the average *gas_price* for each of the closest *minute_of_the_day*;
 - 6 $g = 335000$; // Number of gas units needed for SB submission (measured in gwei)
 - 7 $d = z \cdot g \cdot c$;
 - 8 return d ;
-

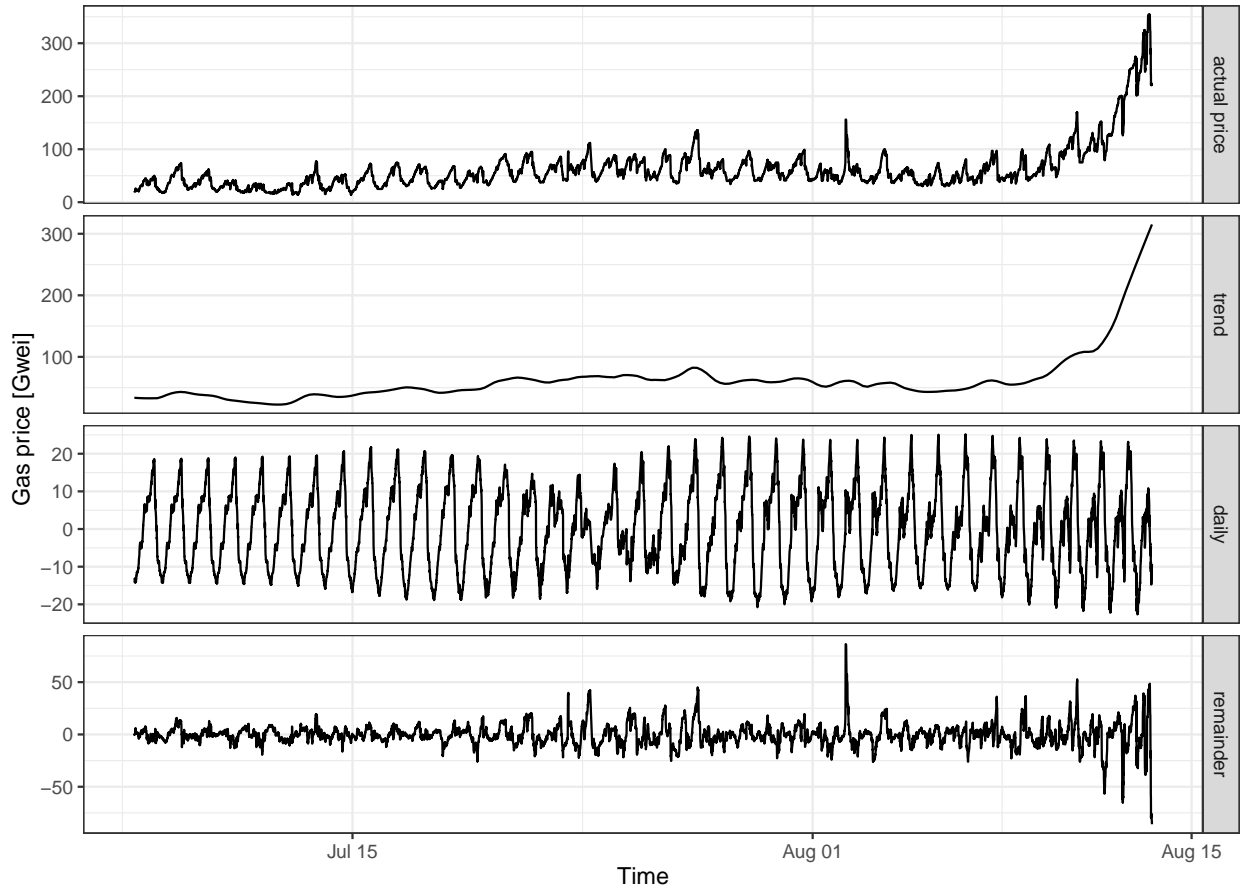


Figure 11: Example of gas price decomposition (*average* scenario) performed using R forecast package [87, 88].