Integrity Management and Access Control of External Storages using Blockchain Technology

Hasan Mohammad Shahriar
Research and Development
Kona Software Lab
Dhaka, Bangladesh
h.m.shahriar@konasl.com

Muhammad Nur Yanhaona
Research and Development
Kona Software Lab
Dhaka, Bangladesh
nur.yanhaona@konasl.com

***Abstract*—**

***Index Terms*—peer-to-peer computing, distributed information systems, document handling**

I. Introduction

Since its inception in 2008 [10], the blockchain technology has gained widespread attention as a transformative technology that can revolutionize many industries [1]. Blockchain based digital currencies such as Bitcoin [10], Ether [15], and Ripple XRP [5] are already considered viable alternatives to existing currencies for trade and commerce for their security and ease of transfer. Blockchain smart contracts [12] [15], on the other hand, have spawned innovative applications in many business and financial sectors due to their capacity of encoding the rules of interaction and ensuring their enforcement.

Central to the appeal of blockchain technology is its maintenance of a distributed ledger of transactions – called the blockchain – in a peer-to-peer network of autonomous and anonymous entities. In a blockchain network, all entities are even and none of them is trusted; still, the security and integrity of the transaction ledger can be guaranteed. This feat is achieved by a complete replication of information in all network participants where each participant validates and executes all transactions. As long as the majority of the network participants are honest, the outcome of the transactions, i.e., the state of the blockchain ledger can be trusted.

The blockchain technology's decentralization of trust through information and processing replication in a theoretically infinitely scalable peer-to-peer network is leading innovations and renovations in many application domains where trust and information security are key concerns. However, problem in one area in particular appears to be a major obstacle for blockchain based application adoption. This is the problem of document storage.

The blockchain technology is inherently unsuitable for storing bulky information such as files and media contents due to the networking and storage cost associated with their management. Peculiarities of blockchain ledger maintenance such as *blockchain reorg* [2] further complicates the situation by making direct integration of existing trusted storage solutions with a blockchain network difficult. Finally, public blockchain technologies find the continual preservation and integrity insurance requirement for trustworthy document storage in conflict with their blockchain transaction ledger maintenance incentive where participants are only being paid for extending the ledger of transactions[1] and they can join or leave the network at any time.

Nevertheless there are some blockchain based or blockchain inspired storage technologies such as Ethereum Swarm [13], Filecoin [7], Storj [14], and IPFS [4] already available. These solutions breakdown a user's file as a series or hierarchy of data chunks then distribute the chunks to the peer-to-peer network. On a broad level, some of these storage solutions are like traditional distributed hash tables [8] [6]. Some others are like peer-to-peer file sharing services such as the popular Bittorrent [11]. These solutions apply some bitcoin-like incentive mechanisms on top of these base technologies to motivate the network participants to retain and serve the data chunks.

The motivation for these solutions is that they protect the users from vendor locked-in and they offer an overall larger storage capacity compared to existing storage alternatives. However, blockchain based solutions have the common problem that the owner (or user) has to take the responsibility of ensuring persistence and integrity of his/her data in the blockchain by retaining document metadata and issuing periodic audits. Furthermore, despite the combined storage capacity being huge, the download bandwidth can be significantly low as the network peers may be running simple commodity hardware behind low-speed network connections. In addition, designing incentive mechanism for long-term persistent of documents in a mining network is difficult

Legacy databases of existing applications are also an obstacle for the applications' migration to the blockchain domain. Data stored in proprietary data centers are often confidential that a typical administrator may not be comfortable to put in the hand of anonymous blockchain participants. Further, when existing cloud storage providers [9] [3] have already solved the storage capacity, scalability, and cost-effectiveness problems

This research is funded by KONA Software Lab, Limited.

[1] by mining transactions into new blocks

for the clients; there is little motivation for moving data into a blockchain storage.

We believe, to steer blockchain application innovations, blockchain technology should be supportive of existing storage solutions instead of being their competitors. In other words, the goal should be integrating existing storage technologies with blockchain – not toppling them. This collaboration of technologies can bring the best of both worlds. The blockchain technology can ensure integrity of external documents and control access to them according to the transparent governance of blockchain smart contracts and leave the actual storage, delivery and capacity scaling to a matured storage technology. Here, the blockchain technology is ideally suited for its part because information corruption in a blockchain network is very difficult and access rules written in the smart contracts are self-enforcing if integrated properly.

In our scheme, location, signature, access control configuration, upload/download fees and so on metadata information about the document is stored in blockchain smart contracts. A user gets access to the externally stored documents by interacting with a *Storage Integration Blockchain Gateway* from a blockchain client application. Any conversation with the gateway involves a series of transactions in the blockchain network and happens following the instructions of some secure interaction protocol. Finally, if the gateway approves the access request then it generates an access token to the external storage that the user uses to upload/download a document with the external storage directly. In case of a download, in particular, the client application verifies document authenticity by locally computing the document signature and matching it against the signature stored in the blockchain before delivering the document to the user.

Figure I depicts a high-level description of the system architecture of our solution. Observant readers will notice that the *Blockchain Network* and the back-end *Document Storage* of Figure I can scale up to meet users' high-availability and other quality of service needs. However, the same cannot be said about the *Gateway*. If proper care is not taken, its failure can make documents unavailable for client access. We avoided this grave potential problem by ensuring that the gateway database only contains information derived from the blockchain network. Hence, the gateway can be limitlessly replicated and the same back-end document storage may be interfaced by many of them at the same time. Furthermore, we support different gateways to be configured differently to charge users for upload and/or download differently based on their quality of service (QoS) and application requirements.

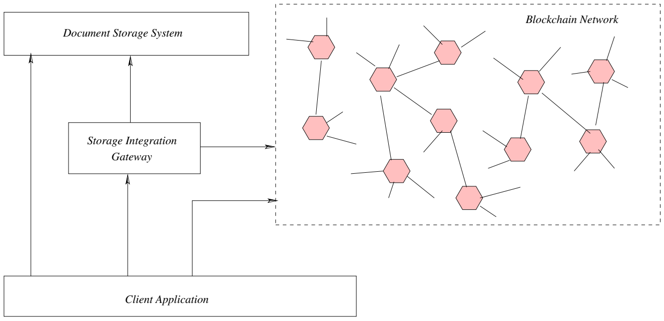

Figure 1. External Storage System Integration Model

The underlying core innovations that make our solution work are as follows:

1) Efficient, secure, and accountable document upload-download protocols that support configurable blockchain based payments.
2) A generic document access control configuration paradigm using blockchain smart contracts.
3) Enforcement of access control rules in the storage integration gateways.
4) Fault-tolerant design of the storage integration gateway against blockchain transaction reversal and back-end document storage failures.

This paper describes these core innovations and discusses some associated concerns. The rest of the paper is organized as follows:

A. Paper Organization

References

[1] Break through with blockchain. https://www2.deloitte.com/us/en/pages/financial-services/articles/blockchain-series-deloitte-center-for-financial-services.html. Accessed: 2019-05-06.

[2] Chain reorganization. https://en.bitcoin.it/wiki/Chain_Reorganization. Accessed: 2019-02-12.

[3] Google cloud storage: Features and benefits. https://cloud.google.com/storage/features/. Accessed: 2019-05-15.

[4] Juan Benet. Ipfs - content addressed, versioned, p2p file system, 07 2014.

[5] D. Stalin David. The ripple protocol consensus algorithm. 2014.

[6] Michael J. Freedman and David Maziéres. Sloppy hashing and self-organizing clusters. In M. Frans Kaashoek and Ion Stoica, editors, *Peer-to-Peer Systems II*, pages 45–55, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[7] Protocol Labs. Filecoin: A decentralized storage network. https://filecoin.io/filecoin.pdf, 2017.

[8] Petar Maymounkov and David Mazières. Kademlia: A peer-to-peer information system based on the xor metric. In *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, IPTPS '01, pages 53–65, London, UK, UK, 2002. Springer-Verlag.

[9] James Murty. *Programming Amazon Web Services*. O'Reilly, first edition, 2008.

[10] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Cryptography Mailing list at https://metzdowd.com*, 03 2009.

[11] Johan Pouwelse, PawełGarbacki, Dick Epema, and Henk Sips. The bittorrent p2p file-sharing system: Measurements and analysis. In *Proceedings of the 4th International Conference on Peer-to-Peer Systems*, IPTPS'05, pages 205–216, Berlin, Heidelberg, 2005. Springer-Verlag.

[12] Nick Szabo. Formalizing and securing relationships on public networks. *First Monday*, 2(9), 1997.

[13] viktor trón, aron fischer, dániel a. nagy, zsolt felföldi, and nick johnson. swap, swear and swindle incentive system for swarm. https://swarm-gateways.net/bzz:/theswarm.eth/ethersphere/orange-papers/1, May 2016.

[14] Shawn Wilkinson, Tome Boshevski, Josh Brandoff, and Vitalik Buterin. Storj a peer-to-peer cloud storage network. https://storj.io/storj2014.pdf, 2014.

[15] D. Wood. Ethereum: a secure decentralised generalised transaction ledger. 2014.