Towards Blockchain-based Auditable Storage and Sharing of IoT Data

Hossein Shafagh, Anwar Hithnawi
Department of Computer Science, ETH Zurich
{*shafagh, hithnawi*}*@inf.ethz.ch*

Simon Duquennoy
Inria, France & RISE SICS, Sweden
simon.duquennoy@inria.fr

Abstract

Today the cloud plays a central role in storing, processing, and distributing data. Despite contributing to the rapid development of various applications, including the IoT, the current centralized storage architecture has led into a myriad of isolated data silos and is preventing the full potential of holistic data-driven analytics for IoT data. In this abstract, we advocate a data-centric design for IoT with focus on resilience, sharing, and auditable protection of information. We introduce the initial design of our blockchain-based end-to-end encrypted data storage system. We enable a secure and persistent data management, by utilizing the blockchain as an auditable access control layer to a decentralized storage layer.

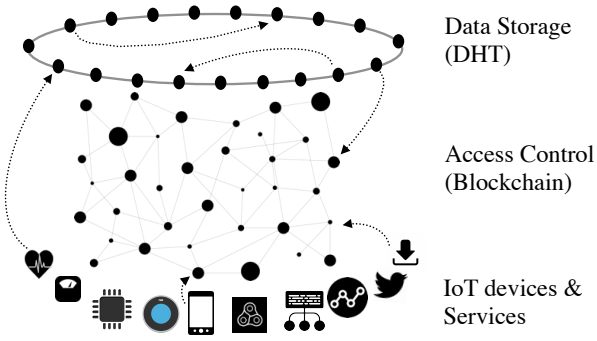

Figure 1: Three layers of our blockchain-based, end-to-end encrypted, and decentralized data storage.

1 Motivation

Internet of Things. With the emergence of networked embedded devices dubbed as the IoT, we witness an ever increasing number of innovative applications in various domains, such as healthcare, fitness, and automation. The current ecosystem of the IoT consists typically of low-power devices equipped with the necessary sensors collecting high-resolution data of their environments. The data is then stored in a third-party cloud storage provider for further processing. In other words, each application service introduces its set of devices and processes the collected data to provide a promised service.

This current approach has resulted into monolithic and isolated data silos, where users have no control over their data. The users have no other option than to trust the cloud and rely on its promises of availability, resilience, and security. The collected data is also valuable beyond one specific application and should be made easily accessible to other services. E.g., one area which could gain from longitudinal health and fitness data is personalized health care. More importantly, in the current model the fate of our data is tied with the lifespan of the service.

Approach. These limitations necessitate a rethinking of the way we currently handle IoT data in its entirety. Instead of storing data centrally in data centers, which are located at the edge of the Internet backbone, we need a data-centric approach which abstracts away the location of data [6]. We advocate a reliable storage and distribution of data streams while empowering the data owner with fine-granular access control. This would facilitate the emergence of a new class of data-driven applications and ensure data ownership. At the same time, our approach must accommodate for a scalable system capable of handling high access throughputs. We envision a separation of data and services, such that services can transparently tap any data source. Inspired by recent blockchain-based technologies [1, 7], we combine a blockchain and an off-chain distributed storage to construct a secure and auditable IoT data management system. An IoT data stream has an append-only nature, where only the data producer has write permissions. On the other hand, several readers (i.e., services) can have simultaneous read rights to the same data stream or one reader can retrieve data from several streams simultaneously. Readers can perform random access on streams.

2 Initial System Design

As illustrated in Figure 1, our system separates access control from the data plane. The former is realized with a public blockchain and the latter with a distributed data storage. The peers of our distributed cloud have financial incentives (similar to Storj [5] or FileCoin [4]) to provide persistent storage. Peers can be either individual users utilizing the excess of their available storage space or commercial cloud storage providers. Data is encrypted end-to-end at the client-side. Hence, the peers have no insights about the hosted data at their side.

Blockchain. We employ a publicly verifiable ledger (blockchain) to create an accountable distributed system and bootstrap trust in an untrusted network, without a central point of trust. Blockchain-based technologies [2] incentivize a network of peers to make computations towards consensus in the network. The consensus agrees on the next valid block of the blockchain. Each block contains validated transactions which remain publicly auditable. In our system, transactions consist of per data stream ownership and access permissions.

Data Plane. The IoT data is a stream where data records are generated continuously. This renders the current distributed storage approaches [4, 5] which primarily target archiving of data not suitable for IoT data. Hence, instead of storing data records, we store data chunks which compose several consecutive data records. To this end, we abstract a data stream into data chunks. Although chunking of data prevents random access at the record level, it has a positive impact on the performance of data retrieval since in time-series most queries require data that is co-located in time (e.g., all records of one day) [3].

IoT data is highly compressible. The compressed data chunks reduce bandwidth and storage requirements. Before a chunk leaves the origin, it is encrypted with an efficient symmetric cipher (AES-CTR). The key is shared with services which are granted read access rights.

We rely on a *Distributed Hash Table* (DHT) as our general-purpose private key-value data store interface. The DHT serves as a scalable, self-managing storage with high availability (i.e., robust against targeted communication outages or malicious attacks in case of central servers). The DHT enforces a randomized storage across a 256-bit address space. Additionally, data replication is used to ensure high availability.

Access Control. We use the blockchain to store access permissions securely. A signed transaction in our case contains the data owner, data readers, the corresponding data stream, and some additional metadata. Access rights are granted per data stream and are limited in time, as expressed in the number of chunks. The data owner can extend or revoke the sharing of a data stream. Moreover, keys are renewed periodically.

For any request to retrieve data, the responsible DHT node first checks the blockchain for access rights. A malicious DHT node could hand out data without permission. However, the impact of this action is limited since *(i)* data is encrypted, *(ii)* each node holds a small random fraction of a data stream due to the nature of DHTs, and *(iii)* access right violation is detectable.

Search. In our off-chain DHT, we store key-value pairs. In our case, the value is the current data chunk in a data stream, whereas the key (i.e., a 256-bit identifier) is the cryptographic hash of the tuple: <stream-ID, device-ID, user-ID, #counter>. The IDs are unique bit strings (i.e., hash digests). The key is used for a lookup in the DHT.

One challenge of such decentralized systems is an efficient search. To address this challenge, we consider building local indices [3] and share these with services. This abstraction allows an authorized party to be able to determine which chunks to retrieve locally.

3 Work Under Progress

Realizing such a system requires addressing research challenges at several fronts. We are currently in the process of finalizing our design and implementing a complete prototype of our system and building several IoT applications on top of it. We tackle *(i)* privacy and security by integrating integrity protection, authentication, and encryption in the design, *(ii)* scalability by the distributed nature of the underlying blockchain and peer-to-peer storage network, *(iii)* fine-grained access control by separating data plane and access control, and *(iv)* durability by giving control over the fate of data to the users.

References

[1] Ali, M., Nelson, J., Shea, R., and Freedman, M. J. Blockstack: A Global Naming and Storage System Secured by Blockchains. In *USENIX ATC* (2016).

[2] Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J. A., and Felten, E. W. SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. In *IEEE Symposium on Security and Privacy* (2015).

[3] Gupta, T., Singh, R. P., Phanishayee, A., Jung, J., and Mahajan, R. Bolt: Data Management for Connected Homes. In *USENIX NSDI* (2014).

[4] Techical Report. Filecoin: A Cryptocurrency Operated File Network. `http://filecoin.io/filecoin.pdf`, 2014.

[5] Techical Report. Storj: A Peer-to-Peer Cloud Storage Network. `https://storj.io/storj.pdf`, 2016.

[6] Zhang, B., Mor, N., Kolb, J., Chan, D. S., Lutz, K., Allman, E., Wawrzynek, J., Lee, E., and Kubiatowicz, J. The Cloud is Not Enough: Saving IoT from the Cloud. In *USENIX HotCloud* (2015).

[7] Zyskind, G., Nathan, O., and Pentland, A. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In *IEEE Security and Privacy Workshops* (2015).