Safeguarding Distributed Data Storage at the Edge

Cláudio Correia
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
claudio.correia@tecnico.ulisboa.pt

ABSTRACT

This work aims at designing security mechanisms for data storage at the edge, we plan to leverage the use of secure hardware (e.g., Intel SGX) as a means to harden the implementation. However, SGX alone is not enough. We consider two different challenges and focus on each one individually, where the resulting techniques can be combined to support the deployment of edge storage systems. Our first challenge is the user authentication at the edge, that due to its strong locality, can compromise user privacy. We strive to design a scheme based on pseudonyms, similar to previous work, while improving this scheme's storage cost and fault tolerance. In the second challenge, we aim to design a cryptographic proof of data replication, to audit the local storage of an edge provider and detect potential storage oversell.

1 INTRODUCTION

Edge computing is a paradigm that extends cloud computing with storage and processing capacity close to the user, supporting new applications that require low latency [10]. This paradigm assumes the availability of fog nodes that are located close to the edge. However, fog nodes are most likely managed by several local providers and installed in physical locations that are more exposed to tampering [2, 8]. Therefore, fog nodes are substantially vulnerable to be compromised. A compromised fog node can tamper with stored data, leading these applications into a fail state, impacting the quality of service. One or several compromised fog nodes can also obtain private information about the clients, such as their daily routines, habits, and preferences, due to the strong locality between clients and fog nodes. Therefore, it is important to secure the functions fog nodes provide.

This work aims at designing security mechanisms for data storage at the edge. To achieve our goal we plan to leverage the use of Intel SGX enclaves as a means to harden our service implementation in each fog node. In this context, we separate this paper into two different components: 1) Stored data access protection from unauthorized entities. To achieve this, we aim to design a decoupled component dedicated to access control at the edge, where clients leverage pseudonyms to authenticate towards edge resources while maintaining unlinkability between requests. Unlinkability must be enforced even after client access is revoked while [5], reducing the storage costs from linear to logarithmic regarding the granularity of revocation, when compared to previous work. We also plan to design a resilient protocol based on gossip to ensure that no revocation information is lost in the presence of faulty servers or an attacker, guaranteeing that no client is able to evade revocation at the edge. 2) An audit tool that aims to assess whether a storage node at the edge is able to retrieve a data object with a latency smaller than some specified threshold δ. We leverage the availability of enclaves, to ensure that the proof is produced by the fog node being audited. We plan to design a cryptographic time-bounded challenge to detect if a malicious fog node is resorting to remote cheap cloud storage[3].

We expect that our contributions can later be combined to help harden distributed storage systems for the edge.

2 WORK PROGRESS

We now present our current work, that discusses each of our two main components separately. We envision that our contribution to these components will be paramount for the deployment of distributed storage systems at the edge. The first stage of our research has been fulfilled by the design and implementation of a distributed access control service for the edge, aiming for a resilient service while enforcing user anonymity at the edge. We have proposed a novel mechanism to construct revocation material that maintains user anonymity and only suffers a logarithmic storage cost. Our work further explored the gossip-based reliable broadcast protocol, to achieve data and replica resilience at the edge.

The second stage in our research, is an ongoing work and the preliminary results are promising. We have designed and evaluated a storage proof that despite the enclave and network variance is capable of detecting if a storage provider is resorting to remote storage instead of edge storage [9]. Next, we present an overview of the achieved results.

2.1 Distributed Access Control

Due to the strong locality between edge resources and clients, the *anonymity* of clients is a clear concern at the edge [6, 7]. Applications such VANETs [6], V2V, and V2I broadcast routine traffic messages that can disclose vehicle locations and trajectories. The European Telecommunications Standards Institute (ETSI) sets the use of pseudonyms as a requirement to ensure privacy for vehicles in V2X [4]. 5GAA also presents use cases for V2X that require both pseudonyms and low latencies to execute [1], such as the cooperative lane merge that requires latencies of 20 *ms* for autonomous vehicles.

Access control systems based on pseudonyms usually require publishing revocation information, such as Certificate Revocation List (CRL)s, containing all client pseudonyms. An adversary can use this information to discover and link the various pseudonyms of a client, breaking the client's anonymity. The challenge of preserving pseudonym unlinkability after revocation was addressed by Haas *et al.* [5]. Their solution involves associating pseudonyms to time intervals and revoking only pseudonyms of the current and future intervals. However, all pseudonyms used within the current interval still suffer from linkability when revoked.

In our solution, clients generate capabilities based on their pseudonyms to authenticate towards edge resources while maintaining unlinkability between requests. We stand out from the related work by enforcing the unlinkability property even after client access is revoked while, when compared to previous work, reducing the storage costs from linear to logarithmic with regard to the granularity of revocation. Our novel approach is based on a binary tree of digital signatures, where the tree nodes work as non-revocation proofs.

We also design a resilient protocol based on gossip to ensure that no revocation information is lost in the presence of faulty servers or an attacker, guaranteeing that no client is able to evade revocation at the edge. Our protocol is based on epochs, and requires the servers to communicate with $N - f$ servers before advancing to a new epoch. Allowing enclaves to detect if they are isolated from the network or lacking revocation information (while no more than f servers fail).

We experimentally evaluate our system and show that despite our cryptographic mechanism and the use of enclaves, users authenticate with low latencies of $0.5 - 3.5$ *ms* while benefiting from strong levels of unlinkability. We compared our system with the related work [5] over a real data set of vehicle traces, and achieve more than 10× storage savings.

2.2 Proof of Timely-Retrievability

Fog nodes will replicate data files such that clients can access data with low-latency. Yet, as the capacity of edge nodes is limited, providers of edge storage may be tempted to oversell their capacity and to hide this behaviour by fetching, on-demand, data from the cloud instead of serving it with the required low latency [3, 9]. We propose a new proof of storage: a *Proof of Timely-Retrievability* (PoTR) that leverages the enclaves in fog nodes, to ensure that the challenge is executed by the fog node being audited, and not by some other fog node, and to avoid revealing the data to be accessed during an audit, prematurely. PoTR aims to assess whether a storage node at the edge is able to retrieve a data object with a latency smaller than some specified threshold δ.

In PoTR, for each challenge c, the fog node has to read a pseudo-random and unpredictable sequence of N data blocks (each with size s_b), and return a cryptographic hash of the concatenation of all accessed data blocks. The number N of data blocks is a configuration parameter, that influences the challenging accuracy and efficiency. A fog node stores the data in the untrusted part, and the enclave will interactively support the untrusted part to solve our challenge. The auditor can later replay the challenge and verify if the fog node was able to correctly solve the challenge under a required time interval. If the fog nodes have a rational behavior, they may resort to remote storage, and our challenges can detect such behavior in an efficient manner.

We have implemented and evaluated experimentally our PoTR. We varied the location of the remote storage server, between the remote cloud, a different university campus, and another close by fog node. Our results show that our proof can accurately detect a node that is not able to satisfy the target latency constraint δ.

However, our experiments, similar to previous work [3], assume either the data is 100% stored in a remote server or locally at the fog node. In the future, we aim to evaluate our proof under different percentages of local and remote storage and evaluate our detection rate. We believe, such a hybrid adversary will be more difficult to detect, and may require a redesign to our storage proof to deal with such a new type of adversary.

Acknowledgments: This work was partially supported by the Fundação para a Ciência e Tecnologia (FCT) under grant 2020.05270.BD, project NG-STORAGE (with ref. PTDC/CCI-INF/32038/2017), project UIDB/50021/2020 and Ainur (PTDC/CCI-COM/4485/2021).

REFERENCES

[1] 5GAA. 2020. C-V2X Use Cases Volume II: Examples and Service Level Requirements. (White paper).
[2] Cláudio Correia, Miguel Correia, and Luís Rodrigues. 2020. Omega: a secure event ordering service for the edge. In *DSN*. València, Spain.
[3] H. Dang, E. Purwanto, and E. Chang. 2017. Proofs of data residency: Checking whether your cloud files have been relocated. In *ASIACCS*. Abu Dhabi, United Arab Emirates.
[4] ETSI. 2021. ETSI TS 102 941 V1.4.1: Intelligent Transport Systems (ITS); Security; Trust and Privacy Management. (Technical Specification).
[5] Jason J Haas, Yih-Chun Hu, and Kenneth P Laberteaux. 2011. Efficient certificate revocation list organization and distribution. *IEEE Journal on Selected Areas in Communications* (2011), 595–604.
[6] Mohammad Khodaei and Panos Papadimitratos. 2018. Efficient, scalable, and resilient vehicle-centric certificate revocation list distribution in VANETs. In *WiSec*. Stockholm, Sweden.
[7] D. Meyer. 2018. *What the GDPR will mean for companies tracking location.* Retrieved 2021-12-22 from https://iapp.org/news/a/what-the-gdpr-will-mean-for-companies-tracking-location/
[8] Mithun Mukherjee, Rakesh Matam, Lei Shu, Leandros Maglaras, and Mohamed Amine Ferrag et al. 2017. Security and Privacy in Fog Computing: Challenges. *IEEE Access* (2017), 19293–19304.
[9] Rita Prates, Cláudio Correia, Miguel Correia, and Luís Rodrigues. 2021. Prova de Resposta Pontual no Acesso ao Armazenamento Contratado na Periferia da Rede. In *INForum*. Lisboa, Portugal.
[10] Glenn Ricart. 2017. A City Edge Cloud with its Economic and Technical Considerations. In *SmartEdge Workshop*. Kona, HI, USA.