Towards Efficient Cryptographic Group Access Control Systems

Stefan Contiu
Université de Bordeaux, France
stefan.contiu@u-bordeaux.fr

ABSTRACT

Encrypted cloud storage systems allow end-users to cryptographically protect their data locally before sending it to the storage spaces. Therefore, groups of users performing collaborative operations, such as data sharing, need to rely on cryptographic mechanisms as well. This PhD research proposal describes a wide spectrum of research challenges and hypothesis within the context of cryptographic group sharing systems, such as zero-knowledge membership operations, the scalability and anonymity of group access control, the scalability of active revocation, and a multi-admin approach to access control.

KEYWORDS

Group Access Control, Trusted Execution Environment, Data Storage

1 INTRODUCTION

Even though cloud storage providers (such as Google Cloud or Amazon S3) make use of TLS to secure the communication between end-users and the cloud, there are no clear guarantees over what happens to the data once it reaches the cloud storages premises. To mitigate this risk, one can choose to encrypt the data locally, before sending it to the storage provider, by relying on encryption keys known only to the end user. Moreover, to enable collaborative operations on this data, such as file sharing, only the group of authorized users should be able to consume and create the group encrypted data. Therefore both the data and the access control to the data need to be cryptographically protected.

1.1 Context

Fig 1 depicts a model of multiple groups sharing data over an untrusted cloud storage. In order to access the encrypted group data, each group posses an encryption key. Group administrators perform group membership operation (e.g. *add* or *remove* users from groups) and thus modify the group access control definitions. Both group data and access control should be decipherable only by group members.

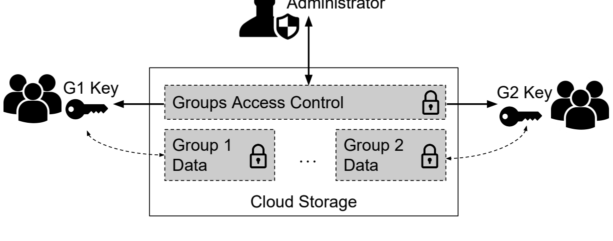


Figure 1: Group Sharing Model Diagram.

A number of factors characterize the context of encrypted cloud storages, mainly when compared to traditional storage mediums. First, cloud storages have a slower response time therefore imposing a low traffic overhead constraint. Second, one should consider the large volumes of data that cloud storages put up with (hundreds of TB for a mid-size organization). Third, as cloud storages can accommodate large sets of users, one needs to consider realistic, possibly very large and dynamic, group membership patterns. Fourth, one should ideally rely on the existing cloud identities and credentials for group membership operations.

The trust model of encrypted cloud storages considers that the group members are trusted. On the other hand, the cloud storage can either have an honest-but-curious behavior, or under toughened models could behave fully arbitrary. We consider that the group administrators are only trusted with performing group membership operations, and manifest an honest-but-curious behavior with respect to the secrets shared by the group. Revoked group members exercise an arbitrary behavior. Under our model, revoked users should lose access immediately to the entire group data (i.e. active revocation). Colluding can happen among any of the un-trusted agents: the cloud, revoked users or curious administrators.

1.2 Sample Scenarios

We illustrate next two real life scenarios in which cryptographic group access control emerges as a necessary component of shared storage systems.

Consider that several governmental institutions desire that their staff work collaboratively on confidential data over the Internet. As these institutions do not posses the specific technological skills to host a shared storage system, they aim at relying on public cloud storages such as Dropbox or Google Cloud. The IT departments of these governmental institutions usually contain few (administrative) staff that carry out typical directory service operations, including access control. Within such a scenario it is required that both the public cloud storage and the mentioned staff administrators do not learn the data content shared by the groups of staff, all while ensuring a correct functioning of the system. The scenario would employ few administrators (e.g. less than 20), a medium size set of users per group (e.g. 15,000) with few TB of shared data.

A second scenario is represented by pay-per-view movie streaming companies that want to outsource their access control services to third parties. In this situation, the groups of users differ based on the viewing rights they have over various movies. The administrators agents are thus represented by the third party subscription service, performing the group membership operations. However, the administrators should not be able to decipher nor consume the video content protected for each group. This scenario envisions a small number of administrator agents, but with very large sets of users (e.g. 1 Million) and very large shared content (e.g. 1PB).

2 RESEARCH CHALLENGES

Considering the aforementioned context, we pose the *conceptual research question* : how to build an end-to-end cryptographic system such that un-trusted agents (e.g. curious administrator, cloud) can perform group membership operations gaining **zero knowledge** over the group secrets? Moreover, can this cryptographic system scale to realistic cloud workloads?

To answer these questions we discuss a number of research dimensions, together with their existing limitations and possible hypotheses to address them.

2.1 *Zero-knowledge* membership operations

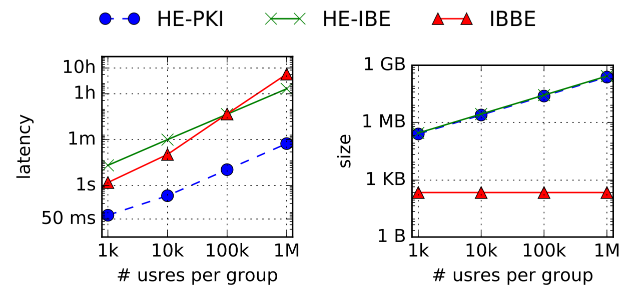
As described earlier, it is necessary that curious agents perform operations on the encrypted data. For example, a curious group administrator should be able to modify the encrypted group access control definitions without gaining knowledge of the actual group key. *Homomorphic encryption* is one cryptographic primitive that supports computations on ciphertexts (i.e. the encrypted data) without gaining knowledge of the plaintext. However, mostly viewed as a theoretical construction, homomorphic encryption suffers from high costs in practical setups.

Differently, one might ask that the un-trusted agent executes all membership operations within a trusted execution environment such as *Intel Software Guard Extensions* (SGX).[1]). Computations performed within SGX enclaves are not visible to the outside, moreover, an attestation mechanism can prove the enclave code as genuine. Relying on an out-of-the-shelf solution like SGX can therefore mitigate the practicability of computations on encrypted data [4].

2.2 *Scalability* of Group Access Control

To recall, the scope of the group access control data structure is to envelope the group key (see Fig. 1) in such a way that it is decipherable only by the group members. A widely used scheme that achieves such guarantee is *Hybrid Encryption* (HE) [5], in which the group key is encrypted separately by using the public key of each group member. However, this approach causes a storage overhead linear in the number of group members, and thus not respecting our self imposed low traffic constraint. A different set of solutions is represented by *Pairing Based Cryptography* (PBC) schemes, such as *Broadcast Encryption* (BE), that can achieve small constant size storage overhead. Moreover, *Identity Based Broadcast Encryption* (IBBE) [3] removes the need of a PKI, and allows the utilization of any arbitrary string (such as the user email or log-in name for the cloud provider) as a user public key. Even though storage wise IBBE seems a perfect fit, its computational performance, just like in the case of homorphic encryption, makes it unusable in practice. Figure 2 lists the latency and storage overhead benchmarks for the three state of the art evaluated schemes. One can observe that even if IBBE produces less than 1KB group metadata expansion, its performance can be up to 420 times worse than HE.

Our *hypothesis* is that, since we can leverage on SGX to solve the *no knowledge* problem (Section 2.2), we can also employ SGX for storing cached secrets that can ease the computations of PBC schemes such as IBBE. While employing such a mechanism, one could mathematically deduce a drop in IBBE's encrypt operation from quadratic to linear. Meaning that, when executed in SGX, IBBE could provide both low storage overhead and fast computational time. However, as SGX would only be employed for administrative operations, regular users will not be able to access nor use the cached secrets, being bounded to the initial un-practical performance of the IBBE scheme. To address this issue, we *hypothesize* that by partitioning the group and deriving a broadcast key for each partition, the user IBBE decryption operation will be bounded to a (much smaller) partition size than the group size.



(b) Latency for group creation. **(c) Group metadata expansion.**

Figure 2: Performance of HE-PKI, HE-IBE and IBBE, without *zero knowledge*.

The proposed solution was fully implemented and evaluated [2]. Results highlight that an extension leveraging on IBBE and SGX can perform membership changes 1.2 order of magnitude faster than the state of the art HE, producing less storage overhead by 6 orders of magnitude.

2.3 *Anonymity* of Group Access Control

Under specific usage cases, it might be desired that secrecy is imposed not only to the content shared by the group, but also for the identities of group members. Within such a context, neither the peers members of the group nor the data storage should be able to infer the identities of group members.

Indeed, current state of the art systems manage to hide the group communication from an untrusted cloud by using improved computational private information retrieval schemes [1]. However, designing a system that can perform access control for groups of users that should not infer the identities of their group peers remains an open challenge.

Our hypothesis is that theoretical constructions such as Anonymous Broadcast Encryption [7], which similarly to IBE and IBBE is utilizing pairing based cryptography, could provide an avenue for investigation. Even if in its raw form such a scheme proves unpractical, we could envision certain assumptions being relaxed by the use of a TEE, similarly to Section 2.2.

2.4 *Scalability* of Active Revocation

Active (or immediate) revocation requires that the data shared by a group is immediately re-encrypted once the user loses access to that group. Differently than *lazy revocation* in which the data is re-encrypted with the first write, active revocation asks for a much

[1] https://software.intel.com/en-us/sgx

higher cost as the whole data has to be re-encrypted at once, as quickly as possible.

To take advantage of data locality, one would want to perform the revocation on the storage side, and thus avoid losing time with moving the data from and to the cloud. Thus, one can envision cloud side agents equipped with SGX enclaves re-encrypting the data without being able to decipher any of it. But since the data to be re-encrypted can grow to very large sizes, a single cloud worker can easily reach a scalability bottleneck.

Li et al. [6] propose to transform the data into multiple packets by using the All or Nothing Transform (AONT) and then over encrypt only one of the packages. The AONT schemes requires that all packages are present in order to reconstruct the data. Therefore, in the case of a revocation, it suffices to re-encrypt the over encrypted package. In addition to executing the AONT scheme in SGX, we *hypothesize* that a distributed set of workers can scale-out the immediate revocation process. However, the challenge is to have a set of untrusted workers cooperate in a decentralized setup, in order to perform the re-encryption, and then provide proof that the re-key was indeed correctly performed and terminated.

2.5 Access Control by Multiple Administrators

A side research dimension arises when considering a different assumption about the set of group administrators. If we consider that a sub-set of administrators could be potentially compromised, we would need to enforce a mechanism in which every decision for each group membership operation is to be taken by multiple administrators rather than a single one. Within such a set-up, for each membership operation, administrators would have to vote (binary decision) if they believe that the group operation should be undertaken or not. We could envision that each group has different policies on the quorum of administrators required per each operation (create or delete group; add or remove user).

The individual votes of each administrator would have to be secret, and therefore hidden from the peers administrators, the hosting (cloud) platform, and the end-users of the group. However the outcome of the voting should be public. Once a vote count (outcome) is final, it is necessary that the group membership change is performed. The (possibly untrusted) agent that performs the group membership should provide proof that it performed the change according to the vote and also that it did not introduce other users or him/herself in the membership change.

Our *hypothesis* is that blockchain technologies could be applicable for this scenario and therefore have the administrators conduct the voting process in a decentralized manner through a smart contract. Moreover, finalizing an immutable transaction that is appended to the blockchain, implies that the access control change was performed per the outcome of the vote. Within such a hypothesized solution, Trusted Execution Environments such as SGX could impact the blockchain technology in beneficial ways. For example, the proof-of-work or the smart contracts could be executed inside SGX enclaves. Their results could be signed by the enclave and published for verification to the rest of the network.

3 CONCLUSION

The presented PhD research proposal introduced a number of research challenges in the context of group sharing over encrypted cloud storages. The proposal introduced the context of the research work by detailing the threat model, the requirements factors and by leveraging on real life examples.

Moreover, the proposal discussed a number of research challenges, together with limitations and hypothesis within the field of cryptographic access control systems.

Acknowledgments

The author is supported by Scille and DGA under contract RAPID-172906010, and hosted by the University of Bordeaux under the supervision of Laurent Réveillère. Parts of the work are being performed in joint collaboration with with the University of Neuchâtel (Pascal Felber, Marcelo Pasin, Rafael Pires and Sèbastien Vaucher) [2] and Université Catholique de Louvain (Etienne Rivière). The work is also partially supported by the CHIST-ERA "DIONASYS" project.

REFERENCES

[1] Sebastian Angel and Srinath TV Setty. 2016. Unobservable Communication over Fully Untrusted Infrastructure.. In *OSDI*. 551–569.

[2] Stefan Contiu, Rafael Pires, Sèbastien Vaucher, Marcelo Pasin, Pascal Felber, and Laurent Réveillère. 2018. IBBE-SGX: Cryptographic Group Access Control over Inquisitive Cloud Storages. In *Dependable Systems and Networks (DSN), 2018 48th Annual IEEE/IFIP International Conference on.* IEEE.

[3] Cécile Delerablée. 2007. Identity-based broadcast encryption with constant size ciphertexts and private keys. In *International Conference on the Theory and Application of Cryptology and Information Security.* Springer, 200–215.

[4] Ben Fisch, Dhinakaran Vinayagamurthy, Dan Boneh, and Sergey Gorbunov. 2017. Iron: functional encryption using Intel SGX. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security.* ACM, 765–782.

[5] William C Garrison, Adam Shull, Steven Myers, and Adam J Lee. 2016. On the practicality of cryptographically enforcing dynamic access control policies in the cloud. In *Security and Privacy (SP), 2016 IEEE Symposium on.* IEEE, 819–838.

[6] Jingwei Li, Chuan Qin, Patrick PC Lee, and Jin Li. 2016. Rekeying for encrypted deduplication storage. In *Dependable Systems and Networks (DSN), 2016 46th Annual IEEE/IFIP International Conference on.* IEEE, 618–629.

[7] Benoît Libert, Kenneth G Paterson, and Elizabeth A Quaglia. 2012. Anonymous broadcast encryption: Adaptive security and efficient constructions in the standard model. In *International Workshop on Public Key Cryptography.* Springer, 206–224.