PhD topic proposal

- **Title: Code-Based Secure Schemes in Multi-Clouds.**
- **Advisors: Françoise Levy-dit-Vehel (ENSTA-INRIA) and Abdullatif Shikfa (Alcatel-Lucent Bell Labs France).**
- **Location: INRIA Saclay–Île-de-France, project-team Grace (Daniel Augot) and Alcatel-Lucent, Nozay.**
- **Start: January 2014.**

1 Introduction

Outsourcing data and computations is an ongoing trend which benefits another boost from cloud computing. Cloud computing is an increasingly adopted paradigm which considers computing resources as utilities: the goal is to reach a stage where computing resources would be economic, elastic, instantly available and on-demand as in the case of electricity or water. There is however a major difference: data and computations contain information much richer, more valuable and more sensitive than electricity and water as commodity. Thus security is often cited as the number one concern in the adoption of cloud computing.

Information security is classically defined as confidentiality, integrity and availability of data and these areas offer interesting research challenges in the outsourced setting.

Confidentiality refers to the need of users to ensure secrecy of their data, and is classically obtained through encryption, which prevents access to data in clear. However access to data is required to perform computations on them. A first research challenge is thus to come up with encryption mechanisms that meet these conflicting requirements by enabling computations on encrypted data.

Integrity means that the user is able to detect if data was modified, altered or corrupted, and is usually achieved with a signature or HMAC scheme. Such schemes are passive though in that they only check for integrity (but they do not repair errors), and their operation requires the whole data to be retrieved. In a cloud computing setting, a more desirable property would be an ability to check integrity actively on partially retrieved data.

Finally availability means that data should be available whenever the user requires it, hence it relates to the fact that the system is resilient to failures. Telecommunications systems typically require high-availability which is usually achieved though sufficient redundancy. The general theory of redundancy in data is known as coding theory, which achieves integrity and availability by expanding data redundantly, to obtain benefits like error resilience, etc. The most fundamental problem is keeping the redundancy small while resisting as many errors as possible.

These aspects and many others relating to security in the cloud have been mostly addressed by the cryptography community with recent breakthroughs such as fully homomorphic encryption. These schemes primarily target a single cloud scenario and they are far from efficient or practical. In this thesis we propose to explore the multi-cloud model to take advantage of multiple computing cloud operators to enhance security. Such a setting is ideal for the use of coding schemes. The aim is thus to explore new advanced concepts in coding, such as Locally Decodable Codes, that can substantially improve secure distributed storage and computation schemes, in terms of communication complexity, reliability, or functionality.

2 State of the art

Looking at the big picture and the major principles behind cloud computing, the core issue of distributed and remote storage has already been studied for years producing a rich literature. As mentioned in the introduction, availability in distributed storage is classically solved by the introduction of redundancy and codes such as the Reed-Solomon codes [1]. Confidentiality and integrity are dealt with by using cryptographic mechanisms such as encryption and signatures, but these mechanisms are constraining and do not allow operations on encrypted data; Computations on encrypted data are possible if the cryptographic mechanisms are homomorphic and such mechanisms have received much attention recently in particular with Gentry's breakthrough consisting of a theoretical fully-homomorphic encryption scheme [2] that supports any computation on encrypted data. The scheme is not efficient though and thus cannot be readily applied in real systems, but other mechanisms enabling only specific operations are already practical. For example searchable encryption, which is an encryption mechanisms that enables searches on encrypted data has been shown to be sufficiently efficient [4] although several improvements in terms of flexibility in particular could be envisioned.

It is interesting to note that usually coding schemes and cryptographic ones are considered separately and studying their interactions is already an interesting topic. Moreover, most cryptographic schemes focus on computational security, which means that an attacker with a great but not infinite amount of resources cannot break current algorithms, and security proofs are based on hardness assumptions (e.g. hardness of computing discrete logarithms in appropriate groups). However such assumption can be broken if new algorithms are found or if the computing model changes, e.g. if quantum computing becomes practical. On the contrary codes are usually studied in the framework of information theory, and encryption mechanisms based on codes (e.g. [5]) resist quantum attacks and in a broader sense the use of codes permits to reach unconditional security. Introducing codes to improve security mechanisms beyond encryption is therefore a promising direction.

In particular Locally Decodable Codes (LDCs) present unique characteristics which are appealing from a security perspective. LDCs appeared formally in the literature in 2000 [7]. The idea is to retrieve a particular symbol of a message by looking only at a few positions of its encoding. Such a concept has natural applications in Private Information Retrieval (PIR) schemes as well as multiparty computation and average case complexity [8]. From the theoretical point of view, the main research issues are the construction of LDCs with the smallest redundancy while minimizing the query complexity. Note that there do not exist codes with constant query complexity and positive rate (asymptotically). While constructions exist based on so-called Reed-Muller codes, matching vector codes and multiplicity codes [9], the subject of the optimal trade-off between query complexity and length is still open.

Apparently, PIR schemes and locally decodable codes have not been jointly considered in the context of distributed storage or clouds, with the main concern in the literature being a multi-server setting. Also, the opportunity of doing computation on the data is usually not considered in PIR schemes and LDCs.

3 Directions and expected work

The overall approach is to consider multi-cloud as an opportunity to enhance security in cloud computing and to examine how new coding schemes, in particular LDCs, could contribute to enhance the security of distributed remote storage and computations.

Our concerns are mainly confidentiality, availability, and advanced features like computation

on private data. We will start by focusing on storage as it is the most emblematic cloud computing service. Building on a preliminary work showcased during the open days of Alcatel-Lucent Bell Labs France, we will introduce coding schemes as a replacement for the linear encoding used, to determine how we gain in communication complexity, storage space, availability and functionalities. We will then extend our work to computation on encoded data by studying the feasibility of combining the previous schemes with privacy features.

While multi-cloud is considered here as a case study, we expect broader results, in more generic distributed storage settings, in constructions of LDCs, and in terms of their application to security features such as private information retrieval.

4 Organization of the work

We envision the progress of the work as follows :

Short term objectives :

Conduct a Thorough investigation of Locally Decodable Codes. By this we mean review the state-of-the-art constructions of LDCs, as well as implement the major families of such codes - namely Reed-Muller codes and multiplicity codes - with a detailed analysis of their performances as a function of their parameters.

Make existing code constructions practical in the distributed storage context. We are here thinking of using and tuning the codes that have been studied above in the context of distributed storage provided by Alcatel-Lucent.

Investigate ways to obfuscate the encoded data stored in the clouds. More specifically, study how to inject privacy features in the coding constructions made before. This is strongly linked to the practical setting given by Alcatel-Lucent.

Long term objectives :

Investigate more complex LDC constructions. We here have in mind two lines of research. One is to examine constructions based on algebraic-geometry codes, namely Hermitian codes. The other is to construct LDCs from liftings of affine-invariant codes, following the work of Guo, Kopparty and Sudan [3].
We will make the relevant tuning of such codes in the context of concern.

Generalize the context, i.e. consider more generic settings than multi-clouds.

Enlarge the functionalities, for example to Private Information Retrieval Schemes, searchable encryption, partial sharing, proofs of retrievability...

References

[1] Irving S. Reed and Gustave Solomon. Polynomial Codes Over Certain Finite Fields. *Journal of the Society for Industrial and Applied Mathematics (SIAM)*, **8** (2): pages 300–304, doi:10.1137/0108018.

[2] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st annual ACM symposium on Theory of computing (STOC '09). ACM, New York, NY, USA*, pages 169–178.

[3] Alan Guo, Swastik Kopparty, Madhu Sudan. New affine-invariant codes from lifting. *ECCC report no. 149, nov. 2012*

[4] Reza Curtmola, Juan Garay, Seny Kamara and Rafail Ostrovsky. Searchable symmetric encryption: Improved definitions and efficient constructions *Journal of Computer Security*, **19** (5): pages 895–934. 2011.

[5] Robert J. McEliece. A Public-Key Cryptosystem Based On Algebraic Coding Theory. *DSN Progress Report: 42–44*, pages 114–116, 1978.

[6] Mohammad Iftekhar Husain, Steve Ko, Atri Rudra, and Steve Uurtamo. Almost universal hash families are also storage enforcing. *CoRR*, abs/1205.1462, 2012. `http://arxiv.org/abs/1205.1462`

[7] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *STOC 2000, Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 80–86. ACM, 2000.

[8] Sergey Yekhanin. Locally Decodable Codes and Private Information Retrieval Schemes. Information Security and Cryptography. Springer, 2010.

[9] Swastik Kopparty, Shubangi Saraf and Sergey Yekhanin. High-Rate Codes with Sublinear-Time Decoding. In *STOC 2011, Proceedings of the Fourty-third Annual ACM Symposium on Theory of Computing, June 6-8, 2011, San Jose, CA, USA*, ACM, 2011.