Labex CominLabs
Project POSEIDON

2012-2015

Final reporting

POSEIDON in a nutshell

POSEIDON stands for, in French, « **expression et déploiement de POlitiques de SEcurité Intégrées pour DONnées externalisées ». In English, we would say « How to express and deploy integrated security policies for outsourced data ».** From 2012 to 2015, it addressed three main issues:

- a need for sound studies on the efficient combination of different security mechanisms,
- a need to work on the systematic formalization of security properties guaranteed by the involved security mechanisms. In particular when these mechanisms are combined,
- a need to work on automating the choice of mechanisms to be deployed. An issue especially important in the context of growing data/content outsourcing practices.

These issues require a strong synergy between researchers with complementary backgrounds and skills, this is why members of the project have been chosen to get complementary profiles, coming from three different labs and five different institutions: Caroline Fontaine (CNRS and Télécom Bretagne, Lab-STICC, coordinator), Gouenou Coatrieux (Télécom Bretagne, LaTIM), Nora Cuppens (Télécom Bretagne, Lab_STICC), Frédéric Cuppens (Télécom Bretagne, Lab_STICC), Sébastien Gambs (Univ. Rennes 1 and INRIA, IRISA), David Gross-Amblard, (Univ. Renne1, IRISA), and Nicolas Prigent (Centrale Supélec, IRISA). Moreover, 2 PhD students have been funded by the project: Anis Bkakria (2012 - 2015) and Julien Lolive (2012-2015). 1 post-doctoral researcher has been funded by the project: Wei Pan (2013-2014).

Results

During this project, three tracks were followed in parallel, but not independently, as our goal was to converge to a global vision of the security.

1. The first track concerns the **study and improvement of security mechanisms** related to data privacy in the Cloud. It includes studies of some mechanisms combination to address the targeted issues. In particular, we focused on the following ones:

   - **Ensuring confidentiality of outsourced data, while preserving query and processing utilities**

The most usual way to ensure confidentiality of outsourced data consists of encrypting the data on the client machine (which is supposed to be trusted) before uploading it to the cloud storage server. The main issue is then to be able to perform queries on this data while it is encrypted. So, one of our objectives has been to define new solutions allowing to ensure the best compromise between outsourced data confidentiality and utility.

If the encryption scheme is chosen among the most common ones (as AES, etc), then data should be decrypted before on the server side to be processed; in this case, there is a huge problem concerning the delivery to the server of the decryption keys, which may be compromised. Nevertheless, some solutions melting classical encryption and fragmentation can be derived. One of our contributions has been to propose a pragmatical approach [Bkakria et al. 2013a, Bkakria et al. 2013b] allowing the protection of confidentiality of sensitive information in outsourced multi-relational databases by improving existing approaches based on a combination of data fragmentation and encryption. These approaches have a major limitation as they assume that data to be outsourced is represented within a single relation schema (or table), which is too strong and seldom satisfied in real environments. Then we defined a secure and effective technique for querying data hosted on several service providers. Finally, we improved the security of the querying technique in order to protect data confidentiality under a collaborative cloud storage service provider model.

In parallel, we worked on recent and promising enhanced encryption schemes called *Fully Homomorphic Encryption* (FHE) schemes, which enable to perform any process composed of additions and multiplications on the encrypted data, without requiring to decrypt it. Despite numerous partial answers, the problem of designing such a powerful primitive has remained open until the theoretical breakthrough of the FHE scheme published by Gentry in the late noughties. Since 2009, a lot of publications provided variants and improvements. In particular, several so-called *somewhat FHE* cryptosystems have been proposed, which allow any number of additions but a bounded number of multiplications. These schemes are really interesting as they are less complex than the fully homomorphic ones and are able to process a number of multiplications that is sufficient for most applications. Hence, they are considered today as the most promising. But despite these promising characteristics, their overhead remained too high to make them directly usable in practice when our study began. There are mainly two ways to improve their efficiency. The first one is to propose new tricky variants that are less complex. The other one is to find some crafty way to implement them. Unfortunately, very few implementations were published and publicly discussed when our study begin, to measure how far we stood from their use in real applications. We performed between 2012 and 2014 a number of steps towards bridging the gap between non trivial algorithms and their practical, relatively seamless, execution on (somewhat) FHE schemes. We have also provided some experimental results indicating that there was hope, in the near term, to be able to homomorphically execute simple algorithms on BGV-style cryptosystems in reasonable time. Our work also proposed tools to help programmers use such schemes properly. It has been the result of a fruitful Lab-STICC collaboration with academic teams from XLIM and CEA/LIST [Aguilar et al. 2013, Fau et al. 2013]. Nevertheless, some important issues remained, as for example the huge size of the related ciphertexts and keys, and the difficulty to estimate the real security level of the schemes. Between 2014 and 2015, we worked on the issue related with the huge ciphertext size, proposing a new dedicated symmetric encryption scheme that enables to send the encrypted data to the server with a very small size expansion; the data is then transcrypted on the server to get a ciphertext on which the computation can be performed. It is important to notice that the data is never in clear on the server, and that the security of our solution is firmly established. This works has been the result of a fruitful Lab-STICC collaboration with academic teams from CEA/LIST and INRIA, and with CryptoExperts company [Cantaut et al. 2016]. All theses contributions on FHE have be integrated in an experimental platform at CEA/LIST, showing on a small example of medical diagnosis that FHE can be used in practice in some small but real scenario with reasonable performances.

- **Ensuring at the same time privacy and traceability of malicious users in the context of multimedia content delivery.**

Active fingerprinting schemes were originally invented to deter malicious users from illegally releasing an item, such as a movie or a picture. To do so, each time an item is released, a different fingerprint is embedded in it. This fingerprint is generated with the help of an anti-collusion code. Thus, even if several malicious users collude to release a fake copy of the item, it should be possible to identify at least one of them by analyzing the fingerprint that can be extracted from this fake copy. The fingerprinting scheme is generally encapsulated inside an asymmetric distributed fingerprinting protocol, which objective is to prevent both parties (i.e., the merchant and the buyer) from cheating. Some of these protocols also address privacy concerns.

We worked on the design of PIMENTO, the first multimedia distribution protocol mixing optimal traitor tracing anti-collusion codes (Tardos codes) to trace dishonnest users, while preserving privacy issues for honnest users. A proper security analysis and an implementation have also been driven. This first protocol, has been presented in an international conference in 2014 [Fontaine et al. 2014] and an extended version PIMENTO+ which also tackle item unlinkability issues is currently submitted for publication in a journal.

A second protocol called PINOCCHIO has been designed with the company NagraVisions and a common patent is to be applied for this second protocol before publication.

2. The second track focused on the design of a support tool allowing, **for a given security policy, selection of the best mechanism or combination of mechanisms** to enforce this security policy.

Linking this goal with the first track of the project, we first defined a policy-based configuration framework [Bkakria et al. 2014b] for encrypted data allowing the data owner to specify the policy to be applied over the outsourced data. Then, we provided an efficient method allowing to detect conflicts between confidentiality requirements (e.g., the set of sensitive information) and utility requirements (e.g., SQL queries that should be executed over the encrypted data) specified by the data owner, and proposed a heuristic polynomial-time algorithm for finding a combination of encryption schemes that satisfies a near optimal trade-off between confidentiality requirements and utility requirements.

Nevertheless, analyzing some real-life scenarios of applications that need mechanisms to securely outsource data leads to the following: security and utility requirements specified by data owner are different in each scenario. In addition, they are in some cases heterogeneous (e.g., confidentiality requirements, privacy requirements, ownership requirements, etc). Moreover, security mechanisms allowing to enforce those security requirements have recently been the focus of huge interest, especially cryptographic and information hiding techniques. These techniques greatly help in tackling security issues: copyright protection (cryptography, watermarking, fingerprinting), content/data confidentiality (cryptography through encryption, fragmentation, access control), content/ data integrity (cryptography through digital signature or message authentication codes, watermarking), authentication of entities (cryptography), anonymity (anonymous networks or granting), and privacy (cryptography, k-anonymity and its extensions, and the more recent differential privacy). These mechanisms are known to be efficient when used independently. However, in many situations they have to be combined in an appropriate way to provide the security functionalities without one harming the other. Our objective was then to design support tools that allows data owners to easily specify their security requirements and automatically choose the best set of security mechanisms, and the best way to combine them (e.g. the best order in which they are applied) to get the best tradeoff between complexity, security and utility in the final choices. To meet this objective, we proposed the following contributions:

- **Definition of an appropriate language**

  Using an Epistemic Linear Temporal Logic (Epistemic LTL), we defined an expressive language [Bkakria et al. 2014a] allowing to: (1) formally model a system composed of involved entities (e.g., data owner, Cloud Storage server administrator, external adversary, etc.) and the data structure on which the security policy should be enforced. (2) formally express as finely as possible the security policy defined by the data owner. Then, we defined a reasoning method for our formal model allowing identifying the relevant combination of mechanisms to efficiently enforce the defined security policy.

- **Achieving near-optimal trade-offs in choosing mechanisms addressing several security and utility issues**

  In [Bkakria et al. 2014a], we supposed that the security mechanisms that can satisfy a policy are applied in parallel over the target system. However, we have seen that in some cases, some security mechanisms should be applied over the same part of the data to be outsourced to satisfy the required security properties. Obviously, in those cases, we should take into consideration conflicts that may occur between security mechanisms which make finding a combination of security mechanisms that satisfy many security requirements much harder to fulfill. We defined an approach that extends [Bkakria et al. 2014a] and uses a planning graph based method to find the combination of security mechanisms providing the near optimal trade-off between the security and the utility of the data to be outsourced and the complexity of its application over the used system.

3. The third track concerns the **adaptation of security solutions** (watermarking and joint encryption watermarking) **to the particular contexts of Cloud and peer-to-peer networks**. We studied some usage scenarios based on cloud computing while considering the medical domain, an especially the outsourcing of medical imaging data from hospital (images and their associated medical records). We conducted a risk analysis focusing on digital content only, so as to cope with untrusted and roaming servers, and to identify security objectives in terms of confidentiality, traceability and integrity purpose. As far as we know, such analysis has never been addressed previously. We then adapted different security mechanisms ranging from watermarking, partial image encryption and database watermarking. Such solutions were experimented into a simulated platform for medical image sharing through the cloud, as proof of concept [Pan et al. 2015].

Dissemination

Publications

6 articles published in international journals (3 IEEE, 1 Springer) + 2 international journal articles currently submitted (Sept. 2016)

10 articles published in the proceedings of international conferences (2 IEEE, 1 ACM, 7 Springer)

Patents

A new fingerprinting protocol has been designed, involving the company NagraVisions. We are working on a patent submission before publication.

Start-Up creation

Two startups have been created: Lamane – about anonymization techniques and query rewriting for big data processing, and WaToo – about database watermarking.

New projects

Several projects followed POSEIDON's dynamics, providing extension directions concerning particular aspects of the project. One can mention AUSTRAL (2012-2015, French FUI project melting academics and industries about the protection of multimedia content, with a focus on copyright protection, traceability and privacy), FRAG&TAG (French 2012-2016), EV-TRUST (2014-2016, French regional project melting academics and industries about the protection of multimedia content, with a focus on homomorphic encryption), CRYPTOCOMP (2015-2018, French FUI project melting academics and industries about the applications of homomorphic encryption in different scenarios, e.g. biometrics and multimedia content distribution), SuperCloud (H2020 European project melting academics and industries, about the security of data in clouds of clouds), SePEMeD (ANR Labcom– 2014-2019), and PRIVGEN (2016-2019, a Labex CominLabs Project in collaboration with Labex GENMED).

Important facts

PhD defenses

Anis Bkakria defended his PhD in December 2015 (POSEIDON Track 2)
Julien Lolive defended his PhD on May, 13rd 2016 (POSEIDON Track 1)

Internships

Anis Bkakria got a sponsorship from the EIT ICT Labs Doctoral Training Centre which includes the opportunity to make an internship in another Lab or industrial in Europe. In january 2014, he started an internship of three months in SAP AG Karlsruhe under the supervision of Andreas Schaad and Florian Kerschbaum. His main mission during the internship was the application of the approach of specification and deployment of Integrated Security Policies for Outsourced Data developed in track 2 in one of the current security projects in SAP.

During Summer 2014, three summer internships contributes to the implementation of the platform of Track 3. They integrate different security tools, including some of those developed in Track 2.

During Summer 2015, one summer internship helped us implementing privacy preserving fingerprinting protocol PIMENTO developed in Track 1.

During Spring and Summer 2016, a Master student helped us exploring a new way to melt sanitization and fingerprinting for databases in Track 1 (ongoing work).

Main References

(see http://www.poseidon.cominlabs.ueb.eu/ for a complete list)

[Aguilar et al. 2013] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, S. Sirdey: Recent advances in homomorphic encryption: a possible future for signal processing in the encrypted domain. IEEE Signal Processing Magazine, Volume:30 , Issue: 2 : 2013

[Bkakria et al. 2013a] *A. Bkakria, F. Cuppens, N. Cuppens-Boulahia, and J. M. Fernandez. Confidentiality-preserving query execution of fragmented outsourced data. In Information and Communicatiaon Technology - International Conference, ICT-EurAsia 2013, Yogyakarta, Indonesia, March 25-29, 2013. Proceedings, pages 426–440, 2013*

[Bkakria et al. 2013b] *A. Bkakria, F. Cuppens, N. Cuppens-Boulahia, J. M. Fernandez, and D. Gross-Amblard. Preserving multi-relational outsourced databases confidentiality using fragmentation and encryption. JoWUA, 4(2):39–62, 2013.*

[Bkakria et al. 2014b] *A. Bkakria, A. Schaad, F. Kerschbaum, F. Cuppens, N. Cuppens-Boulahia, and D. Gross-Amblard. Optimized and controlled provisioning of encrypted outsourced data. In 19th ACM Symposium on Access Control Models and Technologies, SACMAT '14, London, ON, Canada - June 25 - 27, 2014, pages 141–152, 2014.*

[Bkakria et al. 2014a] *A. Bkakria, F. Cuppens, N. Cuppens-Boulahia, and D. Gross-Amblard. Specification and deployment of integrated security policies for outsourced data. In Data and Applications Security and Privacy XXVIII - 28th Annual IFIP WG 11.3 Working Conference, DBSec 2014, Vienna, Austria, July 14-16, 2014. Proceedings, pages 17–32, 2014.*

[Canteaut et al. 2016] A. Canteaut, S. Carpov, C. Fontaine, T. Lepoint, M. Naya-Plasencia, P. Paillier, R. Sirdey: Stream Ciphers: A Practical Solution for Efficient Homomorphic-Ciphertext Compression Fast Software Encryption 2016, Volume 9783 of the series Lecture Notes in Computer Science : 2016

[Fau et al. 2013] S. Fau, R. Sirdey, C. Fontaine, C. Aguilar-Melchor, G. Gogniat: Towards Practical Program Execution over Fully Homomorphic Encryption Schemes. P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on : 2013

[Fontaine et al. 2014] C, Fontaine, S. Gambs, J. Lolive, C. Onete: Private asymmetric fingerprinting : a protocol with optimal traitor tracing using Tardos codes Third International Conference on Cryptology and Information Security in Latin America (Latincrypt'14), Volume 8895 of the series Lecture Notes in Computer Science : 2014

[Pan et al. 2015] W. Pan, G. Coatrieux, D. Bouslimi and N. Prigent: Secure Public Cloud Platform for Medical Images Sharing Studies in health technology and informatics, 2015, vol. 210 : 2015