
IN PARTNERSHIP WITH:
**Institut national des sciences appliquées de Lyon**

Activity Report 2018

# Project-Team PRIVATICS

## Privacy Models, Architectures and Tools for the Information Society

IN COLLABORATION WITH: Centre of Innovation in Telecommunications and Integration of services

RESEARCH CENTER
**Grenoble - Rhône-Alpes**

THEME
**Security and Confidentiality**

# Table of contents

# Project-Team PRIVATICS

*Creation of the Team: 2013 January 01, updated into Project-Team: 2014 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A1. - Architectures, systems and networks
A3. - Data and knowledge
A4. - Security and privacy
A9. - Artificial intelligence

**Other Research Topics and Application Domains:**

B2. - Health
B6. - IT and telecom
B8. - Smart Cities and Territories
B9. - Society and Knowledge

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Claude Castelluccia [Team leader, Inria, Senior Researcher, HDR]
Cédric Lauradoux [Inria, Researcher]
Daniel Le Metayer [Inria, Senior Researcher, HDR]
Vincent Roca [Inria, Researcher, HDR]

**Faculty Members**

Antoine Boutet [INSA Lyon, Associate Professor]
Mathieu Cunche [INSA Lyon, Associate Professor]

**Post-Doctoral Fellow**

Raul Pardo Jimenez [Inria, from Feb 2018]

**PhD Students**

Supriya Sreekant Adhatarao [Inria, from Mar 2018]
Coline Boniface [Univ Grenoble Alpes, from Sep 2018]
Guillaume Celosia [INSA Lyon]
Theo Jourdan [INSERM, from Sep 2018]
Raouf Kerkouche [Univ Grenoble Alpes]
Victor Morel [Inria]
Mathieu Thiery [Inria]
Clement Henin [MTES-ENPC, from Sep 2018]

**Technical staff**

Gabor Gulyas [Inria, until Jun 2018]
Belkacem Teibi [Inria, until Sep 2018]

**Interns**

Louis Beziaud [Inria, from Feb 2018 until Jun 2018]
Bastien Durand [INSA Lyon, from Jun 2018 until Aug 2018]
Romain Fournier [INSA Lyon, from Jun 2018 until Aug 2018]
Konstantinos Stefanidis Vozikis [Inria, from Jul 2018 until Sep 2018]
Alexandre Van Beurden [INSA Lyon, from Jun 2018 until Aug 2018]

**Administrative Assistant**
Helen Pouchot-Rouge-Blanc [Inria]

**Visiting Scientist**
Gergely Acs [Budapest University of Technology and Economics, from May 2018 until Jul 2018]

# 2. Overall Objectives

## 2.1. Context

**The promises of new technologies**: Many advances in new technologies are very beneficial to the society and provide services that can drastically improve life's quality. A good example is the emergence of reality mining. Reality mining is a new discipline that infers human relationships and behaviors from information collected by cell-phones. Collected information include data collected by the sensors, such as location or physical activities, as well as data recorded by the phones themselves, such as call duration and dialed numbers. Reality mining could be used by individuals to get information about themselves, their state or performances ("quantified self"). More importantly, it could help monitoring health. For example, the motions of a mobile phone might reveal changes in gait, which could be an early indicator of ailments or depression. The emergence of location-based or mobile/wireless services is also often very beneficial. These systems provide very useful and appreciated services, and become almost essential and inevitable nowadays. For example, RFID cards allow users to open doors or pay their metro tickets. GPS systems help users to navigate and find their ways. Some services tell users where their friends are or provide services personalized to their current location (such as indicating the closest restaurant or hotel). Some wireless parking meters send users a text message when their time is running out. The development of smart grids, smart houses, or more generally smart spaces/environments, can also positively contribute to the well-being of the society. Smart-grids and smart houses attempt to minimize energy consumption by monitoring users' energy consumptions and applying adequate actions. These technologies can help reducing pollution and managing energy resources.

**Privacy threats of new technologies**: While the potential benefits provided by these systems are numerous, they also pose considerable privacy threats that can potentially turn new technologies into a nightmare. Most of these systems leave digital traces that can potentially be used to profile or monitor users. Content on the Internet (documents, emails, chats, images, videos etc) is often disseminated and replicated on different peers or servers. As a result, users lose the control of their content as soon as they release it. Furthermore most users are unaware of the information that is collected about them beyond requested data. It was shown that consumption data provided by smart meters to electricity providers is so accurate that it can be used to infer physical activities (e.g. when the house occupant took a shower or switched-on TV). Also, a picture taken by a user may reveal additional contextual information inferred from the background or the style of any associated text. For example, photos and videos taken with smart phones or cameras contain geo-location information. This may be considered as a potential source of information leakage and may lead to a privacy breach if used for location tracking or in conjunction with data retrieved from OSN (Online Social Networks). The risk becomes higher as the border between OSN and LBS (Location Based Services) becomes fuzzier. For instance, OSN such as FourSquare and Gowalla are designed to encourage users to share their geolocated data. Information posted on social applications such as Twitter can be used to infer whether or not an individual is at home. Other applications, such as Google Latitude, allow users to track the movements of their friends' cellphones and display their position on a map. In addition to social applications, there are other public sources of information that can be exploited by potential adversaries, such as the free geographic data provided by Google Maps, Yahoo! Maps and Google Earth. The danger is to move into a surveillance society where all our online and physical activities are recorded and correlated. Some companies already offer various services that gather different types of information from users. The combination and concentration of all these information provide a powerful tool to accurately profile users. For example, Google is one of the main third-party aggregators and tracks users across most web sites. In addition, it also runs the most popular search engine and, as such, stores web histories of most users (i.e. their search requests), their map searches (i.e. their

requests to the Google Map service), their images and so on [8]. Web searches have been shown to often be sensitive. Furthermore, Google is also going into the mobile and energy business, which will potentially allow it to correlate online profile with physical profiles.

The "Internet of the future" should solve these privacy problems. However, privacy is not something that occurs naturally online, it must be deliberately designed. This architecture of Privacy must be updated and reconsidered as the concept of privacy evolves and new technologies appear.

Even if our main goal is to develop general techniques with a potentially broad impact, Privatics will consider different and various concrete case studies to ensure the relevance and significance of its results. We plan to work on several case studies related to the Internet, online social networks (OSN), mobile services and smart spaces/environments (such as smart grids, smart houses,..), which correspond to challenging application domains with great impact on society.

# 3. Application Domains

## 3.1. Domain 1: Privacy in smart environments

Privacy in smart environments. One illustrative example is our latest work on privacy-preserving smart-metering [2]. Several countries throughout the world are planning to deploy smart meters in house-holds in the very near future. Traditional electrical meters only measure total consumption on a given period of time (i.e., one month or one year). As such, they do not provide accurate information of when the energy was consumed. Smart meters, instead, monitor and report consumption in intervals of few minutes. They allow the utility provider to monitor, almost in real-time, consumption and possibly adjust generation and prices according to the demand. Billing customers by how much is consumed and at what time of day will probably change consumption habits to help matching energy consumption with production. In the longer term, with the advent of smart appliances, it is expected that the smart grid will remotely control selected appliances to reduce demand. Although smart metering might help improving energy management, it creates many new privacy problems. Smart-meters provide very accurate consumption data to electricity providers. As the interval of data collected by smart meters decreases, the ability to disaggregate low-resolution data increases. Analysing high-resolution consumption data, Non-intrusive Appliance Load Monitoring (NALM) can be used to identify a remarkable number of electric appliances (e.g., water heaters, well pumps, furnace blowers, refrigerators, and air conditioners) employing exhaustive appliance signature libraries. We developed DREAM, DiffeRentially privatE smArt Metering, a scheme that is private under the differential privacy model and therefore provides strong and provable guarantees. With our scheme, an (electricity) supplier can periodically collect data from smart-meters and derive aggregated statistics while learning only limited information about the activities of individual households. For example, a supplier cannot tell from a user's trace when he watched TV or turned on heating.

## 3.2. Domain 2: Big Data and Privacy

We believe that another important problem will be related to privacy issues in big data. Public datasets are used in a variety of applications spanning from genome and web usage analysis to location-based and recommendation systems. Publishing such datasets is important since they can help us analyzing and understanding interesting patterns. For example, mobility trajectories have become widely collected in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, interact, and develop social interactions. With billion of handsets in use worldwide, the quantity of mobility data is gigantic. When aggregated, they can help understand complex processes, such as the spread of viruses, and build better transportation systems. While the benefits provided by these datasets are indisputable, they unfortunately pose a considerable threat to individual privacy. In fact, mobility trajectories might be used by a malicious attacker to discover potential sensitive information about a user, such as his habits, religion or relationships. Because privacy is so important

to people, companies and researchers are reluctant to publish datasets by fear of being held responsible for potential privacy breaches. As a result, only very few of them are actually released and available. This limits our ability to analyze such data to derive information that could benefit the general public. It is now an urgent need to develop Privacy-Preserving Data Analytics (PPDA) systems that collect and transform raw data into a version that is immunized against privacy attacks but that still preserves useful information for data analysis. This is one of the objectives of Privatics. There exists two classes of PPDA according to whether the entity that is collecting and anonymizing the data is trusted or not. In the trusted model, that we refer to as Privacy-Preserving Data Publishing (PPDP), individuals trust the publisher to which they disclose their data. In the untrusted model, that we refer to as Privacy-Preserving Data Collection (PPDC), individuals do not trust the data publisher. They may add some noise to their data to protect sensitive information from the data publisher.

Privacy-Preserving Data Publishing: In the trusted model, individuals trust the data publisher and disclose all their data to it. For example, in a medical scenario, patients give their true information to hospitals to receive proper treatment. It is then the responsibility of the data publisher to protect privacy of the individuals' personal data. To prevent potential data leakage, datasets must be sanitized before possible release. Several proposals have been recently proposed to release private data under the Differential Privacy model [25, 56, 26, 57, 50]. However most of these schemes release a "snapshot" of the datasets at a given period of time. This release often consists of histograms. They can, for example, show the distributions of some pathologies (such as cancer, flu, HIV, hepatitis, etc.) in a given population. For many analytics applications, "snapshots" of data are not enough, and sequential data are required. Furthermore, current work focusses on rather simple data structures, such as numerical data. Release of more complex data, such as graphs, are often also very useful. For example, recommendation systems need the sequences of visited websites or bought items. They also need to analyse people connection graphs to identify the best products to recommend. Network trace analytics also rely on sequences of events to detect anomalies or intrusions. Similarly, traffic analytics applications typically need sequences of visited places of each user. In fact, it is often essential for these applications to know that user A moved from position 1 to position 2, or at least to learn the probability of a move from position 1 to position 2. Histograms would typically represent the number of users in position 1 and position 2, but would not provide the number of users that moved from position 1 to position 2. Due to the inherent sequentiality and high-dimensionality of sequential data, one major challenge of applying current data sanitization solutions on sequential data comes from the uniqueness of sequences (e.g., very few sequences are identical). This fact makes existing techniques result in poor utility. Schemes to privately release data with complex data structures, such as sequential, relational and graph data, are required. This is one the goals of Privatics. In our current work, we address this challenge by employing a variable-length n-gram model, which extracts the essential information of a sequential database in terms of a set of variable-length n - grams [15]. We then intend to extend this approach to more complex data structures.

Privacy-Preserving Data Collection: In the untrusted model, individuals do not trust their data publisher. For example, websites commonly use third party web analytics services, such as Google Analytics to obtain aggregate traffic statistics such as most visited pages, visitors' countries, etc. Similarly, other applications, such as smart metering or targeted advertising applications, are also tracking users in order to derive aggregated information about a particular class of users. Unfortunately, to obtain this aggregate information, services need to track users, resulting in a violation of user privacy. One of our goals is to develop Privacy-Preserving Data Collection solutions. We propose to study whether it is possible to provide efficient collection/aggregation solutions without tracking users, i.e. without getting or learning individual contributions.

# 4. Highlights of the Year

## 4.1. Highlights of the Year

Cédric Lauradoux, Vincent Roca with the participation of Claude Castelluccia have created a MOOC on Privacy which has been followed this year by more than 20000 persons.

# 5. New Software and Platforms

## 5.1. FECFRAME

*FEC Framework following RFC 6363 specifications (https://datatracker.ietf.org/doc/rfc6363/)*
KEYWORDS: Error Correction Code - Content delivery protocol - Robust transmission
FUNCTIONAL DESCRIPTION: This sofware implements the FECFRAME IETF standard (RFC 6363) co-authored by V. Roca, and is compliant with 3GPP specifications for mobile terminals. It enables the simultaneous transmission of multimedia flows to one or several destinations, while being robust to packet erasures that happen on wireless networks (e.g., 4G or Wifi). This software relies on the OpenFEC library (the open-source http://openfec.org version or the commercial version) that provides the erasure correction codes (or FEC) and thereby offer robustness in front of packet erasures.

- Participant: Vincent Roca
- Contact: Vincent Roca

## 5.2. Mobilitcs

FUNCTIONAL DESCRIPTION: Mobilitics is a joint project, started in 2012 between Inria and CNIL, which targets privacy issues on smartphones. The goal is to analyze the behavior of smartphones applications and their operating system regarding users private data, that is, the time they are accessed or sent to third party companies usually neither with user's awareness nor consent.

In the presence of a wide range of different smartphones available in terms of operating systems and hardware architecture, Mobilitics project focuses actually its study on the two mostly used mobile platforms, IOS (Iphone) and Android. Both versions of the Mobilitics software: (1) capture any access to private data, any modification (e.g., ciphering or hashing of private data), or transmission of data to remote locations on the Internet, (2) store these events in a local database on the phone for offline analysis, and (3) provide the ability to perform an in depth database analysis in order to identify personnal information leakage.

- Authors: Jagdish Achara, James-Douglass Lefruit, Claude Castelluccia, Franck Baudot, Geoffrey Delcroix, Gwendal Le Grand, Stéphane Petitcolas and Vincent Roca
- Contact: Claude Castelluccia

## 5.3. MyTrackingChoices

KEYWORDS: Privacy - User control
FUNCTIONAL DESCRIPTION: This extension lets you control how you are being tracked on the Internet. It allows you to choose the categories (e.g., health, adult) of the websites where you don't want to be tracked on. When you browse the web, your visited webpages will be categorized on the fly and, depending on your choices, the extension will block the trackers (webpage by webpage) or not.

Existing anti-tracking (Ghostery, Disconnect etc.) and ad-blocking (AdBlock Plus etc.) tools block almost ALL trackers and as a result, ads. This has a negative impact on the Internet economy because free services/content on the Internet are fuelled by ads. As a result, websites are starting to block access to their content if they detect use of Ad-blockers or they ask users to move to a subscription-based model (where users have to pay to get access to the website).

This extension is testing another approach: It is trying to find a trade-off between privacy and economy, that would allow users to protect their privacy while still accessing to free content.

It is based on the assumption that most people are not against advertisements, but want to keep control over their data. We believe that some sites are more sensitive than others. In fact, most people don't want to be tracked on "sensitive" websites (for example related to religion, health,...), but don't see any problem to be tracked on less sensitive ones (such as news, sport,...). This extension allows you to take control and specify which on which categories of sites you don't want to be tracked on! Furthermore, the extension also gives you the option to block the trackers on specific websites.

- Contact: Claude Castelluccia
- URL: https://addons.mozilla.org/FR/firefox/addon/mytrackingchoices/

## 5.4. OMEN+

FUNCTIONAL DESCRIPTION: Omen+ is a password cracker following our previous work. It is used to guess possible passwords based on specific information about the target. It can also be used to check the strength of user password by effectively looking at the similarity of that password with both usual structures and information relative to the user, such as his name, birth date...

It is based on a Markov analysis of known passwords to build guesses. The previous work Omen needs to be cleaned in order to be scaled to real problems and to be distributed or transfered to the security community (maintainability): eventually it will become an open source software. The main challenge of Omen+ is to optimize the memory consumption.

- Participants: Claude Castelluccia and Pierre Rouveyrol
- Contact: Claude Castelluccia

## 5.5. OPENFEC

KEYWORD: Error Correction Code
FUNCTIONAL DESCRIPTION: OpenFEC is a C-language implementation of several Application-Level Forward Erasure Correction (AL-FEC) codecs, namely: Reed-Solomon (RFC 5510), LDPC-Staircase (RFC 5170) codes, and RLC (https://datatracker.ietf.org/doc/draft-ietf-tsvwg-rlc-fec-scheme/). Two versions are available: an open-source, unsupported version (http://openfec.org), and an advanced version commercialized by the Expway SME.

RELEASE FUNCTIONAL DESCRIPTION: Added support of RLC codes (Random Linear Codes), based on a sliding encoding window.

- Participants: Christophe Neumann, Belkacem Teibi, Jérome Lacan, Jonathan Detchart, Julien Laboure, Kevin Chaumont, Mathieu Cunche and Vincent Roca
- Partner: Expway
- Contact: Vincent Roca
- URL: http://openfec.org/

# 6. New Results

## 6.1. Fine-Grained Control over Tracking to Support the Ad-Based Web Economy

**Participant:** Claude Castelluccia.

The intrusiveness of Web tracking and the increasing invasiveness of digital advertising have raised serious concerns regarding user privacy and Web usability, leading a substantial chunk of the populace to adopt ad-blocking technologies in recent years. The problem with these technologies, however, is that they are extremely limited and radical in their approach, and they completely disregard the underlying economic model of the Web, in which users get content free in return for allowing advertisers to show them ads. Nowadays, with around 200 million people regularly using such tools, said economic model is in danger. In this article, we investigate an Internet technology that targets users who are not, in general, against advertising, accept the trade-off that comes with the "free" content, but—for privacy concerns—they wish to exert fine-grained control over tracking. Our working assumption is that some categories of web pages (e.g., related to health or religion) are more privacy-sensitive to users than others (e.g., about education or science). Capitalizing on this, we propose a technology that allows users to specify the categories of web pages that are privacy-sensitive to them and block the trackers present on such web pages only. As tracking is prevented by blocking network connections of third-party domains, we avoid not only tracking but also third-party ads. Since users continue receiving ads on those web pages that belong to non-sensitive categories, our approach may provide a better point of operation within the trade-off between user privacy and the Web economy. To test the appropriateness and feasibility of our solution, we implemented it as a Web-browser plug-in, which is currently available for Google Chrome and Mozilla Firefox. Experimental results from the collected data of 746 users during one year show that only 16.25% of ads are blocked by our tool, which seems to indicate that the economic impact of the ad-blocking exerted by privacy-sensitive users could be significantly reduced.

## 6.2. Differentially Private Mixture of Generative Neural Networks

**Participant:** Claude Castelluccia.

Generative models are used in a wide range of applications building on large amounts of contextually rich information. Due to possible privacy violations of the individuals whose data is used to train these models, however, publishing or sharing generative models is not always viable. In this paper, we present a novel technique for privately releasing generative models and entire high-dimensional datasets produced by these models. We model the generator distribution of the training data with a mixture of k generative neural networks. These are trained together and collectively learn the generator distribution of a dataset. Data is divided into k clusters, using a novel differentially private kernel k-means, then each cluster is given to separate generative neural networks, such as Restricted Boltzmann Machines or Variational Autoencoders, which are trained only on their own cluster using differentially private gradient descent. We evaluate our approach using the MNIST dataset, as well as call detail records and transit datasets, showing that it produces realistic synthetic samples, which can also be used to accurately compute arbitrary number of counting queries.

## 6.3. On the Cost-Effectiveness of Mass Surveillance

**Participant:** Claude Castelluccia.

In recent times, we have witnessed an increasing concern by governments and intelligence agencies to deploy mass-surveillance systems that help them fight terrorism. Although a government may be perfectly legitimate to do so, it is questionable whether a preventive-surveillance state is rational and cost-effective. In this paper, we conduct a theoretical analysis of the cost of such surveillance systems. Our analysis starts with a fairly well-known result in statistics, namely, the false-positive paradox. We propose a quantitative measure of the total cost of a monitoring program, and study a detection system that is designed to minimize it, subject to a constraint in the percentage of terrorists the agency wishes to capture. Our formulation is first illustrated by means of several simple albeit insightful examples of terrorist and innocent profiles. Then, we conduct an extensive experimental study from real-world socio-demographic data of jihadist terrorism in the U.K. and Spain, and provide insight into the rationality and cost-effectiveness of two countries with two of the biggest defense budgets in the world.

## 6.4. To Extend or not to Extend: on the Uniqueness of Browser Extensions and Web Logins

**Participants:** Claude Castelluccia, Gabor Gulyas.

Recent works showed that websites can detect browser extensions that users install and websites they are logged into. This poses significant privacy risks, since extensions and Web logins that re ect user's behavior, can be used to uniquely identify users on the Web. This paper reports on the rst large-scale behavioral uniqueness study based on 16,393 users who visited our website. We test and detect the presence of 16,743 Chrome extensions, covering 28% of all free Chrome extensions. We also detect whether the user is connected to 60 different websites. We analyze how unique users are based on their behavior, and nd out that 54.86% of users that have installed at least one detectable extension are unique; 19.53% of users are unique among those who have logged into one or more detectable websites; and 89.23% are unique among users with at least one extension and one login. We use an advanced ngerprinting algorithm and show that it is possible to identify a user in less than 625 milliseconds by selecting the most unique combinations of extensions. Because privacy extensions contribute to the uniqueness of users, we study the trade-o between the amount of trackers blocked by such extensions and how unique the users of these extensions are. We have found that privacy extensions should be considered more useful than harmful. The paper concludes with possible counter- measures.

## 6.5. Privacy-Preserving Release of Spatio-Temporal Density

**Participants:** Claude Castelluccia, Gergely Acs.

In today's digital society, increasing amounts of contextually rich spatio-temporal information are collected and used, e.g., for knowledge-based decision making, research purposes, optimizing operational phases of city management, planning infrastructure networks, or developing timetables for public transportation with an increasingly autonomous vehicle fleet. At the same time, however, publishing or sharing spatio-temporal data, even in aggregated form, is not always viable owing to the danger of violating individuals' privacy, along with the related legal and ethical repercussions. In this chapter, we review some fundamental approaches for anonymizing and releasing spatio-temporal density, i.e., the number of individuals visiting a given set of locations as a function of time. These approaches follow different privacy models providing different privacy guarantees as well as accuracy of the released anonymized data. We demonstrate some sanitization (anonymization) techniques with provable privacy guarantees by releasing the spatio-temporal density of Paris, in France. We conclude that, in order to achieve meaningful accuracy, the sanitization process has to be carefully customized to the application and public characteristics of the spatio-temporal data.

## 6.6. Algorithmic Decision Systems in the Health and Justice Sectors: Certification and Explanations for Algorithms in European and French Law

**Participant:** Daniel Le Metayer.

Algorithmic decision systems are already used in many everyday tools and services on the Internet, and they also play an increasing role in many situations in which people's lives and rights are strongly affected, such as job and loans applications, but also medical diagnosis and therapeutic choices, or legal advice and court decisions. This evolution gives rise to a whole range of questions. In this paper, we argue that certification and explanation are two complementary means of strengthening the European legal framework and enhancing trust in algorithmic decision systems. The former can be seen as the delegation of the task of checking certain criteria to an authority, while the latter allows the stakeholders themselves (for example, developers, users and decision-subjects) to understand the results or the logic of the system. We explore potential legal requirements of accountability in this sense and their effective implementation. These two aspects are tackled from the perspective of the European and French legal frameworks. We focus on two particularly sensitive application domains, namely the medical and legal sectors.

## 6.7. Capacity: an Abstract Model of Control over Personal Data

**Participant:** Daniel Le Metayer.

While the control of individuals over their personal data is increasingly seen as an essential component of their privacy, the word "control" is usually used in a very vague way, both by lawyers and by computer scientists. This lack of precision may lead to misunderstandings and makes it difficult to check compliance. To address this issue, we propose a formal framework based on capacities to specify the notion of control over personal data and to reason about control properties. We illustrate our framework with social network systems and show that it makes it possible to characterize the types of control over personal data that they provide to their users and to compare them in a rigorous way.

## 6.8. Biometric Systems Private by Design: Reasoning about privacy properties of biometric system architectures

**Participant:** Daniel Le Metayer.

In  is to show the applicability of the privacy by design approach to biometric systems and the benefit of using formal methods to this end. We build on a general framework for the definition and verification of privacy architectures introduced at STM 2014 and show how it can be adapted to biometrics. The choice of particular techniques and the role of the components (central server, secure module, biometric terminal, smart card, etc.) in the architecture have a strong impact on the privacy guarantees provided by a biometric system. Some architectures have already been analysed but on a case by case basis, which makes it difficult to draw comparisons and to provide a rationale for the choice of specific options. In this paper, we describe the application of a general privacy architecture framework to specify different design options for biometric systems and to reason about them in a formal way.

## 6.9. Privacy Risk Analysis to Enable Informed Privacy Settings

**Participant:** Daniel Le Metayer.

is a contribution to enhancing individual control over personal data which is promoted, inter alia, by the new EU General Data Protection Regulation. We propose a method to enable better informed choices of privacy preferences or privacy settings. The method relies on a privacy risk analysis framework parameterized with privacy settings. The user can express his choices, visualize their impact on the privacy risks through a user-friendly interface, and decide to revise them as necessary to reduce risks to an acceptable level.

## 6.10. Enhancing Transparency and Consent in the IoT

**Participants:** Daniel Le Metayer, Claude Castelluccia, Mathieu Cunche, Victor Morel.

The development of the IoT raises specific questions in terms of privacy, especially with respect to information to users and consent. We argue that (1) all necessary information about collected data and the collecting devices should be communicated electronically to all data subjects in their range and (2) data subjects should be able to reply also electronically and express their own privacy choices. In this position paper, we take some examples of technologies and initiatives to illustrate our position (including direct and registry-based communications) and discuss them in the light of the GDPR and the WP29 recommendations.

## 6.11. Toward privacy in IoT mobile devices for activity recognition

**Participant:** Antoine Boutet.

Recent advances in wireless sensors for personal healthcare allow to recognise human real-time activities with mobile devices. While the analysis of those datastream can have many benefits from a health point of view, it can also lead to privacy threats by exposing highly sensitive information. In this work, we propose a privacy-preserving framework for activity recognition. This framework relies on a machine learning technique to efficiently recognise the user activity pattern, useful for personal healthcare monitoring, while limiting the risk of re-identification of users from biometric patterns that characterizes each individual. To achieve that, we first deeply analysed different features extraction schemes in both temporal and frequency domain. We show that features in temporal domain are useful to discriminate user activity while features in frequency domain lead to distinguish the user identity. On the basis of this observation, we second design a novel protection mechanism that processes the raw signal on the user's smartphone and transfers to the application server only the relevant features unlinked to the identity of users. In addition, a generalisation-based approach is also applied on features in frequency domain before to be transmitted to the server in order to limit the risk of re-identification. We extensively evaluate our framework with a reference dataset: results show an accurate activity recognition (87%) while limiting the re-identification rate (33%). This represents a slightly decrease of utility (9%) against a large privacy improvement (53%) compared to state-of-the-art baselines, while reducing the computational cost on the application server.

## 6.12. The Long Road to Computational Location Privacy: A Survey

**Participant:** Antoine Boutet.

The widespread adoption of continuously connected smartphones and tablets developed the usage of mobile applications, among which many use location to provide geolocated services. These services provide new prospects for users: getting directions to work in the morning, leaving a check-in at a restaurant at noon and checking next day's weather in the evening are possible right from any mobile device embedding a GPS chip. In these location-based applications, the user's location is sent to a server, which uses them to provide contextual and personalised answers. However, nothing prevents the latter from gathering, analysing and possibly sharing the collected information, which opens the door to many privacy threats. Indeed, mobility data can reveal sensitive information about users, among which one's home, work place or even religious and political preferences. For this reason, many privacy-preserving mechanisms have been proposed these last years to enhance location privacy while using geolocated services. This work surveys and organises contributions in this area from classical building blocks to the most recent developments of privacy threats and location privacy-preserving mechanisms. We divide the protection mechanisms between online and offline use cases, and organise them into six categories depending on the nature of their algorithm. Moreover, this work surveys the evaluation metrics used to assess protection mechanisms in terms of privacy, utility and performance. Finally, open challenges and new directions to address the problem of computational location privacy are pointed out and discussed.

## 6.13. CYCLOSA: Decentralizing Private Web Search Through SGX-Based Browser Extensions

**Participant:** Antoine Boutet.

By regularly querying Web search engines, users (unconsciously) disclose large amounts of their personal data as part of their search queries, among which some might reveal sensitive information (e.g. health issues, sexual, political or religious preferences). Several solutions exist to allow users querying search engines while improving privacy protection. However, these solutions suffer from a number of limitations: some are subject to user re-identification attacks, while others lack scalability or are unable to provide accurate results. This contribution presents CYCLOSA, a secure, scalable and accurate private Web search solution. CYCLOSA improves security by relying on trusted execution environments (TEEs) as provided by Intel SGX. Further, CYCLOSA proposes a novel adaptive privacy protection solution that reduces the risk of user re-identification. CYCLOSA sends fake queries to the search engine and dynamically adapts their count according to the sensitivity of the user query. In addition, CYCLOSA meets scalability as it is fully decentralized, spreading

the load for distributing fake queries among other nodes. Finally, CYCLOSA achieves accuracy of Web search as it handles the real query and the fake queries separately, in contrast to other existing solutions that mix fake and real query results.

## 6.14. ACCIO: How to Make Location Privacy Experimentation Open and Easy

**Participant:** Antoine Boutet.

The advent of mobile applications collecting and exploiting the location of users opens a number of privacy threats. To mitigate these privacy issues, several protection mechanisms have been proposed this last decade to protect users' location privacy. However, these protection mechanisms are usually implemented and evaluated in monolithic way, with heterogeneous tools and languages. Moreover, they are evaluated using different methodologies, metrics and datasets. This lack of standard makes the task of evaluating and comparing protection mechanisms particularly hard. In this work, we present ACCIO, a unified framework to ease the design and evaluation of protection mechanisms. Thanks to its Domain Specific Language, ACCIO allows researchers and practitioners to define and deploy experiments in an intuitive way, as well as to easily collect and analyse the results. ACCIO already comes with several state-of-the-art protection mechanisms and a toolbox to manipulate mobility data. Finally, ACCIO is open and easily extensible with new evaluation metrics and protection mechanisms. This openness, combined with a description of experiments through a user-friendly DSL, makes ACCIO an appealing tool to reproduce and disseminate research results easier. In this work, we present ACCIO's motivation and architecture, and demonstrate its capabilities through several use cases involving multiples metrics, state-of-the-art protection mechanisms, and two real-life mobility datasets collected in Beijing and in the San Francisco area.

## 6.15. Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!

**Participant:** Antoine Boutet.

Recommendation systems help users identify interesting content, but they also open new privacy threats. In this contribution, we deeply analyze the effect of a Sybil attack that tries to infer information on users from a user-based collaborative-filtering recommendation systems. We discuss the impact of different similarity metrics used to identity users with similar tastes in the trade-off between recommendation quality and privacy. Finally, we propose and evaluate a novel similarity metric that combines the best of both worlds: a high recommendation quality with a low prediction accuracy for the attacker. Our results, on a state-of-the-art recommendation framework and on real datasets show that existing similarity metrics exhibit a wide range of behaviors in the presence of Sybil attacks, while our new similarity metric consistently achieves the best trade-off while outperforming state-of-the-art solutions.

## 6.16. Automatic Privacy and Utility Preservation of Mobility Data: A Nonlinear Model-Based Approach

**Participant:** Antoine Boutet.

The widespread use of mobile devices and location-based services has generated massive amounts of mobility databases. While processing these data is highly valuable, privacy issues can occur if personal information is revealed. The prior art has investigated ways to protect mobility data by providing a large range of Location Privacy Protection Mechanisms (LPPMs). However, the privacy level of the protected data significantly varies depending on the protection mechanism used, its configuration and on the characteristics of the mobility data. Meanwhile, the protected data still needs to enable some useful processing. To tackle these issues, in this work we present PULP, a framework that finds the suitable protection mechanism and automatically configures it for each user in order to achieve user-defined objectives in terms of both privacy and utility. PULP uses nonlinear models to capture the impact of each LPPM on data privacy and utility levels. Evaluation of our framework

is carried out with two protection mechanisms of the literature and four real-world mobility datasets. Results show the efficiency of PULP, its robustness and adaptability. Comparisons between LPPMs' configurator and the state of the art further illustrate that PULP better realizes users' objectives and its computations time is in orders of magnitude faster.

## 6.17. Privacy Preserving Analytics

**Participant:** Mathieu Cunche.

As communications-enabled devices are becoming more ubiquitous, it becomes easier to track the movements of individuals through the radio signals broadcasted by their devices. Thus, while there is a strong interest for physical analytics platforms to leverage this information for many purposes, this tracking also threatens the privacy of individuals. To solve this issue, we propose a privacy-preserving solution for collecting aggregate mobility patterns while satisfying the strong guarantee of $\varepsilon$-differential privacy. More precisely, we introduce a sanitization mechanism for efficient, privacy-preserving and non-interactive approximate distinct counting for physical analytics based on perturbed Bloom filters called Pan-Private BLIP. We also extend and generalize previous approaches for estimating distinct count of events and joint events (i.e., intersection and more generally t-out-of-n cardinalities). Finally, we evaluate expirementally our approach and compare it to previous ones on real datasets.

## 6.18. Detecting smartphone state changes through a Bluetooth based timing attack

**Participants:** Mathieu Cunche, Guillaume Celosia.

Bluetooth is a popular wireless communication technology that is available on most mobile devices. Although Bluetooth includes security and privacy preserving mechanisms, we show that a Bluetooth harmless inherent request-response mechanism can taint users privacy. More specifically, we introduce a timing attack that can be triggered by a remote attacker in order to infer information about a Bluetooth device state. By observing the L2CAP layer ping mechanism timing variations, it is possible to detect device state changes, for instance when the device goes in or out of the locked state. Our experimental results show that change point detection analysis of the timing allows to detect device state changes with a high accuracy. Finally, we discuss applications and countermeasures.

## 6.19. Analyzing Ultrasound-based Physical Tracking Systems

**Participant:** Mathieu Cunche.

A trending application of ultrasound communication is the implementation of ultrasound beacons to track owners of mobile phones in stores and shopping centers. We present the analysis of an Ultrasound-based tracking application. By analyzing several mobile applications along with the network communication and sample of the original audio signal, we were able to reverse engineer the ultrasonic communications and some other elements of the system. Based on those finding we show how arbitrary ultrasonic signal can be generated and how to perform jamming. Finally we analyze a real world deployment and discuss privacy implications.

# 7. Partnerships and Cooperations

## 7.1. Regional Initiatives

### *7.1.1. AMNECYS*

- Title: AMNECYS
- Duration: 2015 - .
- Coordinator: CESICE, UPMF.

- Others partners: Inria/Privatics and LIG/Moais, Gipsa-lab, LJK, Institut Fourier, TIMA, Vérimag, LISTIC (Pole MSTIC) .
- Abstract: Privatics participates to the creation of an Alpine Multidisciplinary NEtwork on CYbersecurity Studies (AMNECYS). The academic teams and laboratories participating in this project have already developed great expertise on encryption technologies, vulnerabilities analysis, software engineering, protection of privacy and personal data, international & European aspects of cybersecurity. The first project proposal (ALPEPIC ALPs-Embedded security: Protecting Iot & Critical infrastructure) focuses on the protection of the Internet of Things (IoT) and Critical Infrastructure (CI).

### *7.1.2. Data Institute*

- Title: Data Institute UGA
- Duration: 2017 - .
- Coordinator: TIMC-IMAG.
- Others partners: AGEIS, BIG, CESICE, GIN, GIPSA-lab, IAB, IGE, IPAG, LAPP, LARHRA, LIDILEM, LIG, LISTIC, LITT&ArTS, LJK, LUHCIE, LECA, OSUG, PACTE, TIMC-IMAG
- Abstract: Privatics is leading the WP5 (Data Governance, Data Protection and Privacy). This action (WP5) aims to analyze, in a multi-disciplinary perspective, why and how specific forms of data governance emerge as well as the consequences on the interaction between the state, the market and society. The focus will be on the challenges raised by the collection and use of data for privacy, on the data subjects' rights and on the obligations of data controllers and processors. A Privacy Impact/Risk assessments methodology and software will be proposed. A case study will focus on medical and health data and make recommendations on how they should be collected and processed.

### *7.1.3. CyberAlps*

- Title: CyberAlps
- Duration: 2018 - .
- Coordinator: IF.
- Others partners: CEA LETI, CERAG, CESICE, CREg, G2E lab, GIPSA-lab, GSCOP, IF, LCIS, LIG, LISTIC, LJK, PACTE, TIMC-IMAG, VERIMAG.
- Abstract: The Grenoble Alpes Cybersecurity Institute aims at undertaking ground-breaking interdisciplinary research in order to address cybersecurity and privacy challenges. Our main technical focus is on low-cost secure elements, critical infrastructures, vulnerability analysis and validation of large systems, including practical resilience across the industry and the society. Our approach to cybersecurity is holistic, encompassing technical, legal, law-enforcement, economic, social, diplomatic, military and intelligence-related aspects with strong partnerships with the private sector and robust national and international cooperation with leading institutions in France and abroad.

### *7.1.4. Antidot*

- Title: Antidot
- Type: Fédération Informatique de Lyon (inter laboratories project)
- Duration: September 2018 - 2020.
- Coordinator: Inria.
- Others partners: LIRIS.
- Abstract: The ANTIDOT project is interested in the privacy issues raised by the increasingly ubiquitous collection of mobility data and their exploitation by third-party applications. The objective of this project is to propose solutions and tools to increase the user awareness about the risks of violation of their privacy in the context of the mobile Internet. In order to achieve this objective, ANTIDOT will jointly address the study of information gathering mechanisms, the study of mobility data vulnerabilities and the protection of this personal data.

## 7.2. National Initiatives

### *7.2.1. FUI*

Title: ADAGE (Anonymous Mobile Traffic Data Generation).

Type: FUI.

Duration: July 2016 - September 2018.

Coordinator: Orange.

Others partners: Inria, CNRS LAAS.

Abstract: The project ADAGE aims at developping solutions for the anonymization of mobility traces produced by mobile operators.

### *7.2.2. ANR*

#### *7.2.2.1. CISC*

Title: Certification of IoT Secure Compilation.

Type: ANR.

Duration: April 2018 - March 2022.

Coordinator: Inria INDES project-team (France)

Others partners: Inria CELTIC project-team (France), College de France (France) (France).

See also: http://cisc.gforge.inria.fr.

Abstract: The objective of the ANR CISC project is to investigate multitier languages and compilers to build secure IoT applications with private communication. A first goal is to extend multitier platforms by a new orchestration language that we call Hiphop.js to synchronize internal and external activities of IoT applications as a whole. CISC will define the language, semantics, attacker models, and policies for the IoT and investigate automatic implementation of privacy and security policies by multitier compilation of IoT applications. To guarantee such applications are correct, and in particular that the required security and privacy properties are achieved, the project will certify them using the Coq proof assistant.

#### *7.2.2.2. SIDES 3.0*

Title: Application of privacy by design to biometric access control.

Type: ANR.

Duration: August 2017 - August 2020.

Coordinator: Uness (France).

Others partners: Inria, UGA, ENS, Theia, Viseo.

Abstract: Since 2013, faculties of medicine have used a shared national platform that enables them to carry out all of their validating exams on tablets with automatic correction. This web platform entitled SIDES allowed the preparation of the medical students to the Computerized National Classing Events (ECN) which were successfully launched in June 2016 (8000 candidates simultaneously throughout France). SIDES 3.0 proposes to upgrade the existing platform. Privatics goals in this project is to ensure that privacy is respected and correctly assessed .

#### *7.2.2.3. DAPCODS/IOTics*

Title: DAPCODS/IOTics.

Type: ANR 2016.

Duration: May 2017 - Dec. 2020.

Coordinator: Inria PRIVATICS.

Others partners: Inria DIANA, EURECOM, Univ. Paris Sud, CNIL.

Abstract:

Thanks to the exponential growth of Internet, citizens have become more and more exposed to personal information leakage in their digital lives. This trend began with web tracking when surfing the Internet with our computers. The advent of smartphones, our personal assistants always connected and equipped with many sensors, further reinforced this tendency. And today the craze for "quantified self" wearable devices, for smart home appliances or for other connected devices enable the collection of potentially highly sensitive personal information in domains that were so far out of reach. However, little is known about the actual practices in terms of security, confidentiality, or data exchanges. The enduser is therefore prisoner of a highly asymmetric system. This has important consequences in terms of regulation, sovereignty, and leads to the hegemony of the GAFAs (Google, Amazon, Facebook and Apple). Security, transparency and user control are three key properties that should be followed by all the stakeholders of the smartphone and connected devices ecosystem. Recent scandals show that the reality is sometimes at the opposite.

The DAPCODS project gathers four renowned research teams, experts in security, privacy and digital economy. They are seconded by CNIL, the French data protection agency. The project aims at contributing along several axes:

- by analyzing the inner working of a significant set of connected devices in terms of personal information leaks. This will be made possible by analyzing their data flows (and associated smartphone application if applicable) from outside (smartphone and/or Wifi network) or inside, through ondevice static and dynamic analyses. New analysis methods and tools will be needed, some of them leveraging on previous works when applicable;
- by studying the device manufacturers' privacy policies along several criteria (e.g., accessibility, precision, focus, privacy risks). In a second step, their claims will be compared to the actual device behavior, as observed during the test campaigns. This will enable an accurate and unique ranking of connected devices;
- by understanding the underlying ecosystem, from the economical viewpoint. Data collected will make it possible to define the blurred boundaries of personal information market, a key aspect to set up an efficient regulation;
- and finally, by proposing a public website that will rank those connected devices and will inform citizens. We will then test the impact of this information on the potential change of behavior of stakeholders.

By giving transparent information of hidden behaviors, by highlighting good and bad practices, this project will contribute to reduce the information asymmetry of the system, to give back some control to the endusers, and hopefully to encourage certain stakeholders to change practices.

### *7.2.3. Inria Innovation Laboratory*

Title: LEELCO (Low End-to-End Latency COmmunications).

Duration: 3 years (2015 - 2018).

Coordinator: Inria PRIVATICS.

Others partners: Expway.

Abstract:

This Inria Innovation Lab aims at strengthening Expway (http://www.expway.com/) commercial offer with technologies suited to real-time data transmissions, typically audio/video flows. In this context, the end-to-end latency must be reduced to a minimum in order to enable a high quality interaction between users, while keeping the ability to recover from packet losses that are unavoidable with wireless communications in harsh environments. In this collaboration we focus on new types of Forward Erasure Correction (FEC) codes based on a sliding encoding windows, and on the associated communication protocols, in particular an extension to FECFRAME (RFC6363)

to such FEC codes. The outcomes of this work are proposed to both IETF and 3GPP standardisation organisations, in particular in the context of 3GPP mission critical communication services activity. The idea of this 3GPP activity is to leverage on the 3GPP Evolved Multimedia Broadcast Multicast Services (eMBMS) and on the existing Long Term Evolution (LTE) infrastructure for critical communications and such services as group voice transmissions, live high-definition video streams and large data transmissions. In this context, the advanced FEC codes studied in LEELCO offer a significant improvement both from the reduced latency and increased loss recovery viewpoints compared to the Raptor codes included in the existing standard (https://hal.inria.fr/hal-01571609v1/en/).

### *7.2.4. Inria CNIL project*

Privatics is in charged of the Cnil-Inria collaboration. This collaboration was at the origin of the Mobilitics project and it is now at the source of many discussions and collaborations on data anoymisation, risk analysis, consent or IoT Privacy. Privatics and Cnil are both actively involved on the IoTics project, that is the follow-up of the Mobilitics projects. The goal of the Mobilitics project was to study information leakage in mobile phones. The goal of IoTics is to extend this work to IoT and connected devices.

Privatics is also in charged of the organization of the Cnil-Inria prize that is awarded every year to an outstanding publication in the field of data privacy.

## 7.3. European Initiatives

### *7.3.1. Collaborations in European Programs, Except FP7 & H2020*

#### *7.3.1.1. COPES*

Title: COnsumer-centric Privacy in smart Energy gridS

Programm: CHISTERA

Duration: December 2015 - december 2018

Coordinator: KTH Royal Institute of Technology

Inria contact: Cédric Lauradoux

Smart meters have the capability to measure and record consumption data at a high time resolution and communicate such data to the energy provider. This provides the opportunity to better monitor and control the power grid and to enable demand response at the residential level. This not only improves the reliability of grid operations but also constitutes a key enabler to integrate variable renewable generation, such as wind or solar. However, the communication of high resolution consumption data also poses privacy risks as such data allows the utility, or a third party, to derive detailed information about consumer behavior. Hence, the main research objective of COPES is to develop new technologies to protect consumer privacy, while not sacrificing the ”smartness”, i.e., advanced control and monitoring functionalities. The core idea is to overlay the original consumption pattern with additional physical consumption or generation, thereby hiding the consumer privacy sensitive consumption. The means to achieve this include the usage of storage, small scale distributed generation and/or elastic energy consumptions. Hence, COPES proposes and develops a radically new approach to alter the physical energy flow, instead of purely relying on encryption of meter readings, which provides protection against third party intruders but does not prevent the use of this data by the energy provider.

#### *7.3.1.2. UPRISE-IoT*

Title: User-centric PRIvacy & Security in IoT

Programm: CHISTERA

Duration: December 2016 - december 2019

Coordinator: SUPSI (Suisse)

Inria contact: Claude Castelluccia

The call states that "Traditional protection techniques are insufficient to guarantee users' security and privacy within the future unlimited interconnection": UPRISE-IoT will firstly identify the threats and model the behaviours in IoT world, and further will build new privacy mechanisms centred around the user. Further, as identified by the call "all aspects of security and privacy of the user data must be under the control of their original owner by means of as simple and efficient technical solutions as possible", UPRISE-IoT will rise the awareness of data privacy to the users. Finally, it will deeply develop transparency mechanisms to "guarantee both technically and regulatory the neutrality of the future internet." as requested by the call. The U-HIDE solution developed inn UPRISE-IoT will "empower them to understand and make their own decisions regarding their data, which is essential in gaining informed consent and in ensuring the take-up of IoT technologies", using a methodology that includes "co-design with users to address the key, fundamental, but inter-related and interdisciplinary aspects of privacy, security and trust."

## 7.4. International Initiatives

### *7.4.1. DATA*

Title: Data and Algorithmic Transparency and Accountability

International Partner (Institution - Laboratory - Researcher):

Université du Québec à Montréal (UQAM) (Canada) - Département d'informatique - Sébastien Gambs

Start year: 2018

See also: http://planete.inrialpes.fr/data-associated-team/

The accelerated growth of the Internet has outpaced our abilities as individuals to maintain control of our personal data. The recent advent of personalized services has lead to the massive collection of personal data and the construction of detailed profiles about users. However, users have no information about the data which constitute its profile and how they are exploited by the different entities (Internet companies, telecom operators, ...). This lack of transparency gives rise to ethical issues such as discrimination or unfair processing.

In this associate team, we propose to strengthen the complementary nature and the current collaborations between the Inria Privatics group and UQAM to advance research and understanding on data and the algorithmic transparency and accountability.

## 7.5. International Research Visitors

### *7.5.1. Visits of International Scientists*

- Sébastien Gambs visited the team in Lyon in April 2018 for a week to initiate the DATA collaboration. We also organized a workshop in data and algorithmic transparency during this week.
- Gergely Acs, assistant professor at Budapest University (Hungary), visited our team for 2 months, from mi-May to mid-July. He worked together with Claude Castelluccia on machine learning (in)security. In particular, he studied how adversarial examples can be used to evade monitoring, and consequently improve privacy.

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. General Chair, Scientific Chair*

Antoine Boutet: Workshop on data transparency, 23/04/2018, Lyon, France.

Claude Castelluccia: APVP 2018, 3-6/06/2018, Porquerolles, France).

Claude Castelluccia: *Intelligence Oversight : Is Human Rights-Preserving Surveillance Possible?*, Grenoble Data institute, 25/01/2018, Grenoble, France.

*8.1.1.2. Member of the Organizing Committees*

Antoine Boutet: Winter School on Distributed Systems and Networks 2018, 4-8/02/2018, Sept Laux, France.

Antoine Boutet: SRDS 2018, 02-05/10/2018, Salvador, Brasil.

Mathieu Cunche: French GNURadio Days, 03/07/2018, INSA Lyon, France.

Daniel Le Metayer: Panel *Physical tracking: nowhere to hide*, CPDP 2018, 24/01/2018, Brussels, Belgium.

Vincent Roca: IEEE WiMob 2018.

### *8.1.2. Scientific Events Selection*

*8.1.2.1. Member of the Conference Program Committees*

Antoine Boutet: Compas 2018, APVP 2018, Middleware 2018.

Mathieu Cunche: ACM WiSec 2018, Mobiquitous 2018, WCNC 2019, ICISSP 2018.

Claude Castelluccia: APF 2018.

Daniel Le Metayer: IWPE 2018, CPDP 2018, APF 2018.

Vincent Roca: SPACOMM 2018, SSCC 2018.

### *8.1.3. Invited Talks*

Antoine Boutet: *Feedback on the Shonan Meeting on Anonymization methods and inference attacks*, 4th Franco-Japanese Cybersecurity workshop, 16/05/2018, Annecy, France.

Claude Castelluccia: *Cognitive security*, 4th Franco-Japanese Cybersecurity workshop, 16/05/2018, Annecy, France.

Claude Castelluccia: *Brain Hacking*, College des Bernardins, 11/12/2018, Paris, France.

Claude Castelluccia: *Plateforme en ligne et transparence*, AFDIT, 09/11/2018, Paris, France.

Claude Castelluccia: *Internet Surveillance*, Intelligence Oversight workshop, 25/01/2018, Grenoble, France.

Cédric Lauradoux: *Cybersécurité et cybermenaces*, Club Democraties, 09/03/2018, Paris, France.

Cédric Lauradoux: *Cybersécurité et cybermenaces*, Rectorat Académie de Grenoble, 10/12/2018, Grenoble, France.

Daniel Le Metayer, *Intelligility and transparency in machine learning and AI*, Société Française de Statistique, 18/05/2018, Paris, France.

Daniel Le Metayer, *Transparency and opacity in IT systems*, INSA Lyon CITI, 23/04/2018, Lyon, France.

Vincent Roca, *Privacy and Connected Objects*, Eclipse IoT Days Grenoble, 18/01/2018, Grenoble, France.

Vincent Roca, *Archéologie de la fuite de nos données personnelles par le biais de nos téléphones*, Atelier Internet – ENSIIB, 06/04/2018, Lyon, France.

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

Master : Antoine Boutet, *Privacy*, 12h, INSA-Lyon, France.

Master : Antoine Boutet, *Security*, 12h, INSA-Lyon, France.

Undergraduate course : Mathieu Cunche, *Introduction to computer science*, 120h, L1, INSA-Lyon, France.

Master : Mathieu Cunche, *Wireless Security*, 6h, M2, INSA-Lyon, France.

Undergraduate course : Mathieu Cunche, *On Wireless Network Security*, 10h, L1, IUT-2 (UPMF - Grenoble University) , France.

Undergraduate course : Mathieu Cunche, *Security & Privacy*, 21h, L3, INSA-Lyon, France.

Master : Mathieu Cunche, *Privacy and Data protection*, 14h, M2, INSA-Lyon, France.

Master : Mathieu Cunche, *Cryptography and Communication Security*, 18h, M1, INSA-Lyon, France.

Master : Cédric Lauradoux, *Advanced Topics in Security*, 20h, M2, Ensimag/INPG, France.

Master : Cédric Lauradoux, *Systems and Network Security*, 30h, M1, Ensimag, France.

Master : Cédric Lauradoux, *Internet Security*, 12h, M2, University of Grenoble Alpes, France.

Master : Cédric Lauradoux, *Cyber Security*, 3h, M2, Laws School of University of Grenoble Alpes, France.

Master : Claude Castelluccia, *Advanced Topics in Security*, 15h, M2, Ensimag/INPG, France.

Master : Claude Castelluccia, *Cyber Security*, 6h, M2, Laws School of University of Grenoble Alpes, France.

Master : Claude Castelluccia, *Data Privacy*, 6h, M2, Laws School of University of Grenoble Alpes, France.

Master : Daniel Le Metayer, *Privacy*, 12h, M2 MASH, Université Paris Dauphine, France.

Master : Daniel Le Metayer, *Privacy*, 12h, M2, Insa Lyon, France.

Master : Vincent Roca, *On Wireless Communications*, 12h, M1, Polytech' Grenoble, France.

Undergraduate course : Vincent Roca, *On Network Communications*, 44h, L1, IUT-2 (University of Grenoble Alpes), France.

Undergraduate course : Vincent Roca, *On Security and Privacy in smartphones*, 6h, L-Pro, University of Grenoble Alpes, France.

Master : Vincent Roca, *On Security and Privacy in smartphones*, 3h, M2, France.

### *8.2.2. E-learning*

**E-learning**

Mooc: Cédric Lauradoux and Vincent Roca, *, 2 month session, FUN-MOOC, Inria, public ciblé, 23000 (13000 first session and 10000 second session).*

### *8.2.3. Supervision*

- PhD in progress : Victor Morel, *IoT privacy*, September 2016, Daniel Le Métayer and Claude Castelluccia.
- PhD in progress : Mathieu Thiery, *IoT privacy* , September 2016, Vincent Roca.
- PhD in progress : Guillaume Celosia, *Wireless Privacy in the Internet of Things* , November 2017, Mathieu Cunche and Daniel Le Métayer.
- PhD in progress : Supryia Adhatarao, *Privacy of E-learning systems* , March 2018, Cédric Lauradoux.
- PhD in progress : Coline Boniface, *Cyberweapons: from bug bounties to zero days* , March 2018, Cédric Lauradoux.
- PhD in progress : Raoul Kerkouche, *Privacy-Preserving Processing of Medical Data* , January 2018, Claude Castelluccia.

- PhD in progress : Clement Henin, *Explainable AI* , September 2018, Claude Castelluccia et Daniel Le Metayer.
- PhD in progress: Théo Jourdan, *Privacy-preserving machine learning in medical domain, October 2018, Antoine Boutet.*
- *Intern (M2): Louis Beziaud,* Privacy of national identity systems*, M2 ENS Rennes, Claude Castelluccia et Daniel Le Metayer.*
- *Intern (L3): Alexandre van Beurden, Inspect what location history reveals about an individual, Antoine Boutet.*
- *Intern (L3): Romain Fournier, Development of a cybersecurity platform, Antoine Boutet.*
- *Intern (L3): Bastien Durand, Analysis of the correlation between the mobility and the personality of an individual, Antoine Boutet.*

### *8.2.4. Juries*

PhD: David Gerault, *Security Analysis of Contactless Communication Protocols*, Université Clermont Auvergne , 27/11/2018, Cédric Lauradoux.

PhD: Jonathan Detchart, *Optimisation de codes correcteurs d'effacements par application de transformées polynomiales*, Université de Toulouse, 05/12/2018, Vincent Roca.

PhD: Elise Tourne, *Le phénome de circulation des données à caractère personnel dans le cloud: étude de droit matériel dans le contexte de l'Union Européenne*, Université Lyon 3, 11/06/2018, Daniel Le Métayer.

## 8.3. Popularization

### *8.3.1. Hearings*

- Claude Castelluccia: *Understanding Algorithmic Decision-Making Systems*, European Parliament, 10/2018, Strasbourg, France.
- Daniel Le Metayer: *Understanding Algorithmic Decision-Making Systems*, European Parliament, 10/2018, Strasbourg, France.
- Daniel Le Metayer: *Transparence et explicabilité des algorithmes d'aide à la décision*, CCNE, 02/2018, Paris, France.

### *8.3.2. Internal or external Inria responsibilities*

- Claude Castelluccia is co-leader of the Worpackage 5 (data governance and privacy) of the Grenoble Data Institute.
- Claude Castelluccia is co-leader of Grenoble CyberAlps (cybersecurity institute of Grenoble).
- Daniel Le Metayer chairs the CNIL-Inria privacy award.
- Vincent Roca is co-editor of the white book *Cybersecurity: current challenges and Inria's research directions*.

### *8.3.3. Articles and contents*

- Claude Castelluccia: *Manipulation informationnelle et psychologique*, Le blog binaire du Monde, 05/2018.
- Claude Castelluccia: *Data surveillance and manipulation*,Interview for Magazine Capital, 12/2018.
- Mathieu Cunche: *Cybersécurité et menace informatique*, Sommet des start-up sciencesetavenir.fr, 11/2018, Lyon, France.
- Mathieu Cunche: *Attaque par déni de service dans le Wi-Fi*, GNULinux Magazine HS 99, 11/2018.
- Mathieu Cunche: *Comprendre les attaques Krack*, GNULinux Magazine HS 99, 11/2018.
- Daniel Le Metayer: *Weighting the impact of the GDPR*, Communications of the ACM, 11/2018.

- Daniel Le Metayer: *Qui gouverne les algorithmes ?*, Revue THIIRD, 11/2018.
- Vincent Roca: Inria White Paper in Cyber-Security.

### *8.3.4. Education*

- Cédric Lauradoux: *Action nombres et cryptographie*, Maison pour la science, Inria, 06/02/2018, Grenoble, France.
- Cédric Lauradoux: *Action nombres et cryptographie*, Maison pour la science, 18/12/2018, Annecy, France.
- Cédric Lauradoux: *Animation du forum du MOOC Protection de la vie privée dans le monde numérique*, 02-03/2018.
- Cédric Lauradoux: : *Animation du forum du MOOC Protection de la vie privée dans le monde numérique*, 11-12/2018.

### *8.3.5. Interventions*

- Cédric Lauradoux: *Atelier cryptograpie*, Fête de la Science, 11-12/10/2018, Grenoble, France.
- Cédric Lauradoux: *Cryptologie et Vie privée*, Semaine des mathématiques, Lycée Gabriel-Faure, 15/03/2018, Tournon, France.
- Cédric Lauradoux: *Cryptologie et Vie privée*, Semaine des mathématiques, Lycée Boissy d'Anglas, 16/03/2018, Annonay, France.
- Cédric Lauradoux: *Challenge de cryptologie*, MathC2+ internship, 26/06/18, Grenoble, France.
- Cédric Lauradoux: *Challenge de cryptologie*, Cité scolaire Jean PREVOST, 02/06/2018, Villard de Lans, France.
- Cédric Lauradoux: *Challenge de cryptologie*, Collège Barnave, 18/01/2018, Saint-Égrève, France.

### *8.3.6. Internal action*

- Cédric Lauradoux: *Réglementation sur les données*, Inria, Grenoble, 19/06/2018.
- Cédric Lauradoux: *Réglementation sur les données*, IMAG, Grenoble, 09/07/2018.

# 9. Bibliography

## Publications of the year

### Articles in International Peer-Reviewed Journals

[1] J. P. ACHARA, J. PARRA-ARNAU, C. CASTELLUCCIA. *Fine-Grained Control over Tracking to Support the Ad-Based Web Economy*, in "ACM Transactions on Internet Technology", October 2018, vol. 18, n$^{o}$ 4, pp. 1-25 [*DOI :* 10.1145/3158372], https://hal.inria.fr/hal-01921868

[2] G. ACS, L. MELIS, C. CASTELLUCCIA, E. DE CRISTOFARO. *Differentially Private Mixture of Generative Neural Networks (ext)*, in "IEEE Transactions on Knowledge and Data Engineering", July 2018, pp. 1-13 [*DOI :* 10.1109/TKDE.2018.2855136], https://hal.inria.fr/hal-01921923

[3] M. ALAGGAN, M. CUNCHE, S. GAMBS. *Privacy-preserving Wi-Fi Analytics*, in "Proceedings on Privacy Enhancing Technologies", April 2018, vol. 2018, n$^{o}$ 2, pp. 4-26 [*DOI :* 10.1515/POPETS-2018-0010], https://hal.inria.fr/hal-01719211

[4] J.-S. BERGÉ, D. LE MÉTAYER. *Phénomènes de masse et droit des données*, in "Communication Commerce Électronique", 2018, vol. n° 12, https://hal.archives-ouvertes.fr/hal-01907540

[5] J. BRINGER, H. CHABANNE, D. LE MÉTAYER, R. LESCUYER. *Biometric Systems Private by Design: Reasoning about privacy properties of biometric system architectures*, in "Transactions on Data Privacy", 2018, vol. 11, n° 2, pp. 111-137, https://hal.archives-ouvertes.fr/hal-01939841

[6] S. CERF, S. BOUCHENAK, B. ROBU, N. MARCHAND, V. PRIMAULT, S. BEN MOKHTAR, A. BOUTET, L. Y. CHEN. *Automatic Privacy and Utility Preservation of Mobility Data: A Nonlinear Model-Based Approach*, in "IEEE Transactions on Dependable and Secure Computing", 2018, https://hal.archives-ouvertes.fr/hal-01910687

[7] L. DEMIR, A. KUMAR, M. CUNCHE, C. LAURADOUX. *The Pitfalls of Hashing for Privacy*, in "Communications Surveys and Tutorials, IEEE Communications Society", 2018, vol. 20, n° 1, pp. 551-565 [*DOI :* 10.1109/COMST.2017.2747598], https://hal.inria.fr/hal-01589210

[8] D. LE MÉTAYER, S. DESMOULIN-CANSELIER. *Algorithmic Decision Systems in the Health and Justice Sectors: Certification and Explanations for Algorithms in European and French Law*, in "European Journal of Law and Technology", December 2018, https://hal.archives-ouvertes.fr/hal-01972259

[9] J. PARRA-ARNAU, C. CASTELLUCCIA. *On the Cost-Effectiveness of Mass Surveillance*, in "IEEE Access", 2018, vol. 6, pp. 46538 - 46557 [*DOI :* 10.1109/ACCESS.2018.2866310], https://hal.inria.fr/hal-01921899

[10] V. PRIMAULT, A. BOUTET, S. BEN MOKHTAR, L. BRUNIE. *The Long Road to Computational Location Privacy: A Survey*, in "Communications Surveys and Tutorials, IEEE Communications Society", 2018, 1 p. [*DOI :* 10.1109/COMST.2018.2873950], https://hal.archives-ouvertes.fr/hal-01890014

### International Conferences with Proceedings

[11] A. BOUTET, F. DE MOOR, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT. *Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!*, in "DSN 2018 - the 48th International Conference on Dependable Systems and Networks", Luxembourg, Luxembourg, IEEE, June 2018, pp. 466-477 [*DOI :* 10.1109/DSN.2018.00055], https://hal.inria.fr/hal-01787060

[12] G. CELOSIA, M. CUNCHE. *Detecting smartphone state changes through a Bluetooth based timing attack*, in "WiSec '18 - 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks", Stockholm, Sweden, June 2018, pp. 154-159 [*DOI :* 10.1145/3212480.3212494], https://hal.inria.fr/hal-01870011

[13] S. J. DE, D. LE MÉTAYER. *Privacy Risk Analysis to Enable Informed Privacy Settings*, in "IWPE 2018 – 4th IEEE International Workshop on Privacy Engineering", London, United Kingdom, Proceedings of the 4th IEEE International Workshop on Privacy Engineering (IWPE 2018), April 2018, pp. 1-8, https://hal.archives-ouvertes.fr/hal-01939845

[14] G. G. GULYÁS, D. F. SOMÉ, N. BIELOVA, C. CASTELLUCCIA. *To Extend or not to Extend: On the Uniqueness of Browser Extensions and Web Logins*, in "WPES'18 - Workshop on Privacy in the Electronic Society", Toronto, Canada, ACM Press, October 2018, pp. 14-27 [*DOI :* 10.1145/3267323.3268959], https://hal.inria.fr/hal-01921863

[15] T. JOURDAN, A. BOUTET, C. FRINDEL. *Toward privacy in IoT mobile devices for activity recognition*, in "Privacy Preserving Machine Learning NeurIPS 2018 Workshop", Montréal, Canada, December 2018, pp. 1-6, https://hal.inria.fr/hal-01941453

[16] T. JOURDAN, A. BOUTET, C. FRINDEL. *Toward privacy in IoT mobile devices for activity recognition*, in "MobiQuitous 2018 - 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services", New York city, United States, November 2018, pp. 1-10, https://hal.inria.fr/hal-01882330

[17] R. KERKOUCHE, R. ALAMI, R. FÉRAUD, N. VARSIER, P. MAILLÉ. *Node-based optimization of LoRa transmissions with Multi-Armed Bandit algorithms*, in "ICT 2018 - 25th International Conference on Telecommunications", Saint Malo, France, June 2018, pp. 1-6, https://hal-imt-atlantique.archives-ouvertes.fr/hal-01946456

[18] D. LE MÉTAYER, P. RAUZY. *Capacity: an Abstract Model of Control over Personal Data*, in "CODASPY 2018 - ACM Conference on Data and Application Security and Privacy", Tempe, United States, Proceedings of the ACM Conference on Data and Application Security and Privacy (CODASPY 2018), ACM, March 2018, pp. 1-12, https://hal.archives-ouvertes.fr/hal-01939847

[19] V. MOREL, D. LE MÉTAYER, M. CUNCHE, C. CASTELLUCCIA. *Enhancing Transparency and Consent in the IoT*, in "IWPE 2018 - International Workshop on Privacy Engineering", London, United Kingdom, Proceedings of the International Workshop on Privacy Engineering (IWPE 2018), IEEE, April 2018, pp. 116-119 [*DOI :* 10.1109/EUROSPW.2018.00023], https://hal.archives-ouvertes.fr/hal-01709255

[20] R. PIRES, D. GOLTZSCHE, S. BEN MOKHTAR, S. BOUCHENAK, A. BOUTET, P. FELBER, R. KAPITZA, M. PASIN, V. SCHIAVONI. *CYCLOSA: Decentralizing Private Web Search Through SGX-Based Browser Extensions*, in "ICDCS 2018 - 38th IEEE International Conference on Distributed Computing Systems", Vienne, Austria, IEEE, July 2018, pp. 467-477 [*DOI :* 10.1109/ICDCS.2018.00053], https://hal.inria.fr/hal-01882430

[21] V. PRIMAULT, M. MAOUCHE, A. BOUTET, S. BEN MOKHTAR, S. BOUCHENAK, L. BRUNIE. *ACCIO: How to Make Location Privacy Experimentation Open and Easy*, in "ICDCS 2018 - 38th IEEE International Conference on Distributed Computing Systems", Vienna, Austria, Proceedings of the 38th IEEE International Conference on Distributed Computing Systems, IEEE, July 2018, pp. 896-906 [*DOI :* 10.1109/ICDCS.2018.00091], https://hal.archives-ouvertes.fr/hal-01784557

### Conferences without Proceedings

[22] M. CUNCHE, L. SAMPAIO CARDOSO. *Analyzing Ultrasound-based Physical Tracking Systems*, in "GreHack 2018", Grenoble, France, November 2018, https://hal.inria.fr/hal-01927513

### Scientific Books (or Scientific Book chapters)

[23] G. ACS, G. BICZÓK, C. CASTELLUCCIA. *Privacy-Preserving Release of Spatio-Temporal Density*, in "Handbook of Mobile Data Privacy", Springer, October 2018, pp. 307-335 [*DOI :* 10.1007/978-3-319-98161-1_12], https://hal.inria.fr/hal-01921891

[24] M. CUNCHE. *Syria*, in "The SAGE Encyclopedia of Surveillance, Security, andPrivacy", B. A. ARRIGO (editor), SAGE, March 2018, https://hal.inria.fr/hal-01798933

### Research Reports

[25] A. BOUTET, F. DE MOOR, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT. *Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!*, Inria Rennes - Bretagne Atlantique, April 2018, pp. 1-12, https://hal.inria.fr/hal-01767059

[26] A. KASSEM, G. ACS, C. CASTELLUCCIA. *Differential Inference Testing A Practical Approach to Evaluate Anonymized Data*, Inria, January 2018, https://hal.inria.fr/hal-01681014

[27] D. LE MÉTAYER, V. MOREL, M. CUNCHE. *A Generic Information and Consent Framework for the IoT*, Inria, December 2018, n° RR-9234, pp. 1-18, https://arxiv.org/abs/1812.06773 , https://hal.inria.fr/hal-01953052

[28] C. MATTE, M. CUNCHE. *Overview of real-world deployment of physical analytics systems*, Inria Grenoble Rhône-Alpes, January 2018, n° RR-9143, 18 p. , https://hal.inria.fr/hal-01682373

[29] C. MATTE, M. CUNCHE. *Spread of MAC address randomization studied using locally administered MAC addresses use historic*, Inria Grenoble Rhône-Alpes, January 2018, n° RR-9142, https://hal.inria.fr/hal-01682363

### Scientific Popularization

[30] M. CUNCHE. *Attaque par déni de service dans le Wi-Fi*, in "Linux Magazine France", November 2018, vol. HS 99, https://hal.inria.fr/hal-01927520

[31] M. CUNCHE. *Comprendre les attaques Krack*, in "GNU/Linux Magazine France", November 2018, vol. HS 99, https://hal.inria.fr/hal-01927522

### Other Publications

[32] B. ADAMSON, C. ADJIH, J. BILBAO, V. FIROIU, F. FITZEK, G. A. M. SAMAH, E. LOCHIN, A. MASUCCI, M.-J. MONTPETIT, M. V. PEDERSEN, G. PERALTA, V. ROCA, P. SAXENA, S. SIVAKUMAR. *Taxonomy of Coding Techniques for Efficient Network Communications*, June 2018, Internet Research Task Force, Request For Comments (RFC) 8406, https://datatracker.ietf.org/doc/rfc8406/, https://hal.inria.fr/hal-00998506

[33] A. BOUTET, M. CUNCHE. *A Privacy-Preserving Mechanism for Requesting Location Data Provider with Wi-Fi Access Points*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01949419

[34] C. BURDINAT, C. THIENOT, T. TRAN, V. ROCA, B. TEIBI. *Pseudo-CR Performance evaluation of AL-FEC and MCS dimensionning*, February 2018, 3GPP TSG-SA WG4 Meeting #97, Fukuoka, 5th – 6th February 2018, https://hal.inria.fr/hal-01731622

[35] M. CUNCHE, L. S. CARDOSO. *Analysis of an Ultrasound-Based Physical Tracking System*, May 2018, working paper or preprint, https://hal.inria.fr/hal-01798091

[36] J. PARRA-ARNAU, C. CASTELLUCCIA. *Dataveillance and the False-Positive Paradox*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01157921

[37] V. ROCA, A. BEGEN. *Forward Error Correction (FEC) Framework Extension to Sliding Window Codes*, September 2018, Working document of the TSVWG (Transport Area Working Group) group of IETF (Internet Engineering Task Force), draft-ietf-tsvwg-fecframe-ext-06 (work in progress), https://datatracker.ietf.org/doc/draft-ietf-tsvwg-fecframe-ext/, https://hal.inria.fr/hal-01345125

[38] V. ROCA, J. DETCHART, C. ADJIH, M. V. PEDERSEN. *Generic Application Programming Interface (API) for Sliding Window FEC Codes*, November 2018, pp. 1-23, Internet Research Task Force - Working document

of the Network Coding Research Group (NWCRG), draft-roca-nwcrg-generic-fec-api-04 (work in progress), https://datatracker.ietf.org/doc/draft-roca-nwcrg-generic-fec-api/, https://hal.inria.fr/hal-01630138

[39] V. ROCA, I. SWETT, M.-J. MONTPETIT. *Sliding Window Random Linear Code (RLC) Forward Erasure Correction (FEC) Schemes for QUIC*, June 2018, Internet Research Task Force, Coding for Efficient Network Communications Research Group (NWCRG) document, work in progress, https://datatracker.ietf.org/doc/draft-roca-nwcrg-rlc-fec-scheme-for-quic/, https://hal.inria.fr/hal-01919876

[40] V. ROCA, B. TEIBI. *Sliding Window Random Linear Code (RLC) Forward Erasure Correction (FEC) Schemes for FECFRAME*, September 2018, pp. 1-25, Working document of the TSVWG (Transport Area Working Group) group of IETF (Internet Engineering Task Force), draft-ietf-tsvwg-rlc-fec-scheme-09 (work in progress), https://datatracker.ietf.org/doc/draft-ietf-tsvwg-rlc-fec-scheme/, https://hal.inria.fr/hal-01630089

[41] I. SWETT, M.-J. MONTPETIT, V. ROCA. *Coding for QUIC*, June 2018, Internet Research Task Force, Coding for Efficient Network Communications Research Group (NWCRG) document, work in progress, https://datatracker.ietf.org/doc/draft-swett-nwcrg-coding-for-quic/, https://hal.inria.fr/hal-01919858