Profile Cominlabs : Final report

Introduction

The Profile cominlabs project has been running for 3 years. It is a truly interdisciplinary where computer scientists, lawyers and sociologists interact to address the problem of online profiling.

This project has produced various main contributions which are currently synthesized in a book entitled “Online profiling: between liberalism and regulation” (written in french). This book will be published in November and accompanied by a conference which will be held on Friday 27th of september 2019 in Rennes. This conference will synthesized the works made by the project and confront our ideas with other internationally recognized researcher working on the domain of private data and profiling.

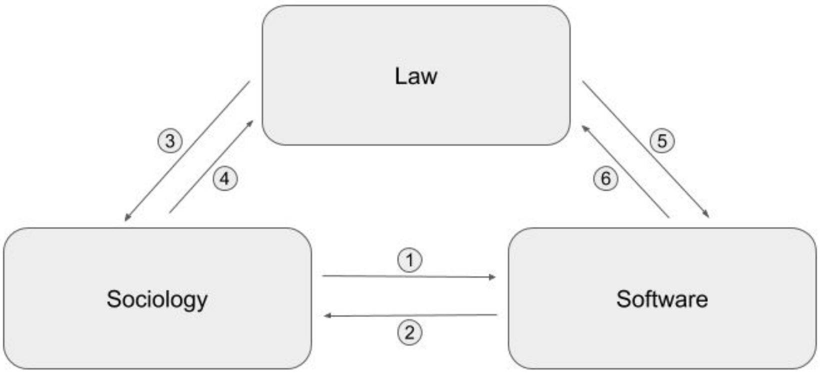
One of the main result of the Profile project comes from the PhD Thesis of Pierre Laperdrix (his PhD was not funded by the project but he worked closely with us) which won the **2018 INRIA/CNIL price “Privacy protection”** for his article “Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints” and the **accesit** to the **Gilles Kahn 2018 Phd thesis price for his thesis "Browser Fingerprinting Exploring Device Diversity to Augment Authentication and Build Client-Side countermeasures"**

Context

The profile cominlabs project has been studying online profiling for the last three years. The project is truly interdisciplinary with software computer scientists, lawyers and sociologists. We have been studying the envision impact and technical abilities of online profiling together with the tools to protect against this type of profiling both from a socio-technical and law point of view.

To carry out this research, the project includes three disciplinary fields that interacted together as shown in the diagram below. Each arrow corresponds to a type of exchange between two disciplinary fields. The nature of these exchanges is considered as a lesson learned from this project and is detailed below:

1. Sociologists have interacted with computer scientists by studying the reactions of the general public to the phenomenon of online profiling and by highlighting the privacy paradox.
2. The computer scientists interacted with the sociologists to define the specifications and co-develop a system to study the traces left by the use of a smartphone.
3. Jurists have interacted with sociologists on the conditions for possible free and informed consent to profiling
4. In-house lawyers interacted with computer scientists to identify and qualify the data involved in profiling.
5. Sociologists and jurists worked on framework for the CGU of a future application aimed to collect data from users’ smartphones.
6. Sociologists have presented their results to explicit the concept of “resigned” consent to jurists



Contributions

Our contributions are organized into two themes :

- **The privacy and profiling paradox**: which state that most people really care about their privacy, but on the other hand they agree to give all their private data to access various services.
- **Profiling regulation**: the computer and legal control instruments enabling users to understand what the operator does with their data.

Theme 1 : The privacy and profiling paradox

1st Contribution : Risks perception analysis from a sociologue point of view.

Relationships between users and smartphone collected-datas are our main focus point in this study. In that way, we developed a methodology based on users dataselves auto-confrontation.

While studies explaining those relationships can be analyzed through what is called *Privacy Paradox* (which balanced the users' willingness to be protected against data gathering, and their practices generating collected data), we suggest that framing the question in this way is problematic because users are far from being enlightened as to what they have agreed to. When they noticed the variety and the quantity of collected-datas on them and their practices, the reactions we observed show most of the time: surprise and fears.

Observation and lessons learned : We noted that most consent were resigned consent and most users were trying to implement a protection scheme against data profiling but they had the impression they lack the tools and knowledge to be efficiently protected. These observation justify the need and drive the development of tools to protect user privacy.

2nd Contribution "Risk assessment of recidivism. An occurrence of criminal profiling".

What are the occurrences of criminal profiling? Rather than an overview, the choice was made to analyse an emblematic case of profiling and the evolution of criminal law at the same time: the assessment of the risk of recidivism. Increasingly, the old fear of new crimes being committed by offenders is leading contemporary legislators to reorient the functions of sentencing and criminal procedure. From the punishment of the past offence, the system shifts, at least in part, towards managing the risk of a future offence. Carried out at low noise, against the background of the challenges of this paradigmatic break in terms of fundamental freedoms, this evolution is promoted by new technologies, especially the application of algorithms for the purpose of profiling individuals. However, to date, it has not attracted much interest from legal researchers, especially in France.

Observation and lessons learned : In addition to a contribution relating to profiling in an interdisciplinary context, this study contribute to the development of the analysis of this neglected legal-political phenomenon.

Theme 2 : Profiling regulation

1st contribution : The DGPR's legal safeguards against discriminatory profiling practices

Profiling is an essential concept aimed by the data protection regulation. In many country, it was because of the profiling that a data protection law was adopted. Naturally, the GDPR deal with this notion. Article 4 § 4 of the European regulation define it, and article 22 prohibit the use of profiling. However, the prohibition is not as clear as it seems and the goal of this contribution is to show why.

First of all, article 22 prohibit every automated individual decision-making, not only profiling. Yet, the text of this article doesn't implement a general prohibition of these techniques. It implement a right for the person. That is to say that "the data subject shall have not right not to be subject to a decision solely based on an automated processing". Many information can be extract from this wording.

On one hand, the data subject have, indeed, a concrete right to oppose to an automated processing. But, on the other hand, the sentence contain, in itself, the exception to the rule. If a human being can act on the processing, or on the decision, it means that the processing is not solely based on an automated decision. So the processing, is lawful.

The other main right related to the profiling is the right to object to the processing. Once again, this right is not absolute because many exception are practical.

An other exception to the profiling is the consent of the data subject. The consent, perceive as a protective right for the data subject, can mainly be a breach in data protection regulation. Many protective principle of the GDPR, like the prohibition of processing sensitive data, can be sidelined by the consent of the data subject.

Observation and lessons learned : All of these exceptions to the prohibition of profiling, and every concept approach in this contribution reveal the complexity of the profiling regulation.

2nd contribution : Framing profiling in the General Data Protection Regulation (GDPR) in the light of European instruments for the protection of fundamental rights

This contribution analyses the framework of profiling carried out by the General Data Protection Regulation (GDPR) of 2016 in the light of European instruments for the protection of fundamental rights, namely, on the European Union (EU) side, the EU Charter of Fundamental Rights 2000 and the Council of Europe, the European Convention on Human Rights (ECHR) of 1950 and the Convention for the Protection of Persons in Respect of Automated Processing of Personal Data of 1981. It should be noted that the latter Convention, known as Convention 108, will be examined in parallel with its modernized version, the so-called “108 + Convention”, even though the latter is not yet in force.

This contribution assesses the degree of alignment of the GDPR profiling framework with the European body of fundamental rights protection. The examination shall be based on the content of the relevant GDPR pivot article, namely Article 22 entitled "Automated individual decision, including profiling", even if, on the one hand, this article does not concern only the profiling hypothesis and, on the other hand, not all profiling hypotheses are limited to the content of this article. If Article 22 specifies in its § 1 that the person concerned has the right not to be the subject of a decision based exclusively on automated processing, including profiling, producing legal effects on it or significantly affecting it in a similar way”, it then provides for exceptions to this right whose conformity with fundamental rights must be assessed (first part of the contribution) and it accompanies them with a number of guarantees whose articulation with fundamental rights must be examined (second part of the contribution).

Observation and lessons learned : Thanks to the convergence of the GDPR and the European instruments for the protection of fundamental rights, Internet users benefit from protection against profiling by web giants.

3rd contribution : Instruments for regulating connected objects in health insurance

This contribution has two components. On the one hand, we have carried out a legal inventory of the situation on connected object links and health insurance and, on the other hand, we have considered what possible profiling practices based, in particular, on the data produced by connected objects may change in the field of health insurance. This raises the question of the legal regulations to be built for the protection of the insured person and the ethical use of data.

Observation and lessons learned : If the proposed law tabled on 23 January 2019, aimed at prohibiting the use of personal data collected by connected objects in the insurance field, does not result, the processing of personal data could become an instrument of differentiation in the market of health and/or provident insurance.

4th contribution : Mitigating browser Fingerprinting

Browser fingerprinting is a technique that collects information about the browser configuration and the environment in which it is running. This information is so diverse that it can partially or totally identify users online. Over time, several countermeasures have emerged to mitigate tracking through browser fingerprinting. However, these measures do not offer full coverage in terms of privacy protection, as some of them may introduce inconsistencies or unusual behaviors, making these users stand out from the rest.

Observation and lessons learned : We address these limitations by proposing a novel approach that minimizes both the identifiability of users and the required changes to browser configuration. To this end, we exploit clustering protocols to identify the devices that are prone to share the same or similar fingerprints and to provide them with a new non-unique fingerprint. We then use this fingerprint to automatically "reconfigure" the devices by running a browser within a docker container. Thus all the devices in the same cluster will end up running a dockerized browser with the same indistinguishable and consistent fingerprint.

5th contribution : Open the black box of customization algorithms

The ever-increasing amount of personal data collected by profiling systems, online or not, is fueling the real-life implementation of highly personalized online services based on recent successful machine learning techniques such as deep neural networks. In a nutshell, these techniques input a detailed personal profile (ex : browsing history, or socio-demographic information with possible criminal background) and typically output a prediction (ex : a list of suggested products, or a score that quantifies the risk of recidivism). Despite the fact that these systems are used widely and intensively, their inner working is often opaque, both about the exact information they use and about the operations performed. Given that these systems may suffer from various biases (sometimes involuntarily) while they may impact strongly some individuals (ex : the result of judgment) it is crucial to be able to put them under scrutiny.

Observation and lessons learned : We advocate for a two-step approach that consists (1) in gathering pairs of (input, output) to/from these systems (ex : by constructing profiles and observing the resulting suggestions or prices) in order to (2) construct a human-understandable view of the way the system under study maps the inputs to the outputs. To the best of our knowledge, the existing methodologies for collecting data from these systems disclose information in an uncontrolled manner which may lead to a biased output for a given input. Proposing a robust methodology that limits the side leaks is thus our objective with respect to the first step. We have designed the methodology and are currently implementing it. The second step is related to explaining machine learning algorithms. Related works are numerous. As a preliminary study we have built a decision tree over the COMPAS dataset that contains profiles of criminals together with various predicted risks (ex : violent recidivism).

Publications

Profile book

Books on Profiling (profilage in French), éditeur Mare et Martin (direction S. Turgis, A. Bensamoun et M. Boizard)

1. Marine Gout, Florian Hemont, “Consentement résigné : en finir avec le Privacy Paradox” Editor Mare et Martin and was presented in several invited talk at different venues.
2. Johann Bourcier “Profiling on mobile phones. Computer tool for identifying traces and reconstructing pieces of life.” Editor Mare et Martin
3. Laurent Rousvoal, “L’évaluation du risque de récidive, Une occurrence du profilage en matière pénale”, Editor Mare et Martin
4. Maryline Boizard, Erwan Picart, “Les garanties juridiques du RGPD contre les pratiques discriminatoires de profilage”, editor Mare et Martin;
5. Sandrine Turgis : “L’encadrement du profilage dans le règlement général sur la protection des données (RGPD) à l’aune des instruments européens de protection des droits fondamentaux” Editor Mare et Martin
6. Marion Del Sol, “Les instruments de régulation des objets connectés en matière d’assurance santé”, Editor Mare et Martin;
7. Benoît Baudry, Davide Frey, Alejandro Gomez Boix, “ La régulation par le « contrôle informatique » : le Browser fingerprinting”, Editor Mare et Martin;
8. Tristan Allard, Sébastien Gambs, Julien Lolive, “Ouvrir la boîte noire des algorithmes de personnalisation”, editor Mare et Martin;
9. Margaux Redon, Les incertitudes juridiques entourant les données issues des objets connectés en santé, Editor Mare et Martin

Books on Le règlement général sur la protection des données, aspects institutionnels et matériels, (in French) Mare et Martin, 2019, (direction A. Bensamoun)

International Conferences

1. M. Del Sol, Enjeux juridiques des objets connectés en matière d’assurance santé. Réflexions à partir et au-delà du cadre français – 23ème colloque de l’Association Information & Management (Montréal, Canada, mai 2018);
2. Alejandro Gómez-Boix, Pierre Laperdrix, Benoit Baudry: Hiding in the Crowd: an Analysis of the Effectiveness of Browser Fingerprinting at Large Scale WWW 2018: The 2018 Web Conference : 2018;

International workshop

1. Maryline Boizard: Protection des algorithmes et sécurité juridique des personnes Workshop international sur la sécurité globale : 2017;

2. M. Del Sol, Les enjeux juridiques et éthiques des objets connectés en matière d'assurance santé – Université d'été pluridisciplinaire et internationale « Travail et innovations technologiques » (Bordeaux, juill. 2018);
3. Marion Del Sol, Margaux Redon: "M-health: the legal issues of connected objects", CYBER SECURITY AND HEALTHCARE, 2017
4. T. Allard : Towards Using Differential Privacy as a Building Block for Privacy-Preserving Algorithms. In Shonan Meeting n°116 : Anonymization methods and inference attacks: theory and practice. Shonan Center, Japan, 2018;
5. Tristan Allard, Davide Frey, George Giakkoupis, Julien Lepiller: Lightweight Privacy-Preserving Averaging for the Internet of Things : 2016
6. Alejandro Gómez-Boix, Davide Frey, Yérom-David Bromberg, Benoit Baudry, "A Collaborative Strategy for mitigating Tracking through Browser Fingerprinting", to appear in Moving Target Defense (MTD) workshop at the 26 ACM Conference on Computer Security (ACM CCS 2019)
7. A. Bensamoun, « AI and data privacy », UNESCO International Symposium, Mobile Learning Week, plénière Safeguarding transparent and auditable use of education data, UNESCO (Paris), 7 mars 2019;

French workshop

1. Maryline Boizard: Les garanties juridiques du règlement européen sur la protection des données personnelles contre les pratiques discriminatoires des algorithmes et des plateformes numériques Journée d'étude : Les discriminations à l'ère de la société numérique : 2017;
2. Maryline Boizard: La protection des droits fondamentaux des migrants et des réfugiés : durée de conservation des données collectées aux frontières Journée d'étude Les données numériques des migrants et des réfugiés : 2017
3. Alejandro Gómez-Boix, Pierre Laperdrix, Benoit Baudry: Fingerprinting mobile devices: A short analysis CIEL 2017 - 6ème Conférence en IngénieriE du Logiciel : 2017;
4. A. Bensamoun « IA : éthique, confiance et responsabilité », Table ronde « Le Droit comme outil de sécurisation », colloque Cyberlex La sécurité dans un monde de transformation numérique, Palais du Luxembourg, 15 janv. 2018.
5. Propos introductifs : l'évolution du régime juridique de la protection des données personnelles, Table ronde : Les données personnes face aux libertés fondamentales, organisée par le M2 Droit privé général promotion 2015-2016, le 30 septembre 2016, Faculté de droit de Rennes I, Erwann PICART.
6. « La consécration du modèle français de l'action de groupe » , Séminaire doctoral, organisé par le CRJO, 14 juin 2017, Erwann PICART

Related work published during Profile by people from the Profile project

1. A. Boutet, F. De Moor, D. Frey, R. Guerraoui, A. Kermarrec, A. Rault: Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter! Dependable Systems and Networks (DSN) : 2018;
2. Pierre Laperdrix, Benoit Baudry, Vikas Mishra: FPRandom: Randomizing core browser objects to break advanced device fingerprinting techniques ESSoS 2017 - 9th International Symposium on Engineering Secure Software and Systems : 2017;
3. Nicolas Harrand and Benoit Baudry: Software Diversification as an Obfuscation Technique, International Workshop on obfuscation : 2017;
4. Benoit Baudry: How can we reconcile diversity and privacy? Dagstuhl seminar on Online Privacy and Web Transparency : 2017;
5. A. Boutet, F. De Moor, D. Frey, R. Guerraoui, A. Kermarrec and A. Rault, "Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!," 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Luxembourg City, 2018, pp. 466-477.
6. Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints P Laperdrix, W Rudametkin, B Baudry Security and Privacy (SP), 2016 IEEE Symposium on, 878-894
7. libmask: Protecting browser JIT engines from the devil in the constants M Mishra, B Baudry Privacy, Security and Trust (PST), 2016 14th Annual Conference on, 301-308

Dissemination

The PROFILE project has strongly contributed to the RadX seminar (“Responsabilité, Algorithmes, Données personnelles : regards croisés”) by making the computer scientists, the lawyers, and the sociologists of the project be aware of each others’ works. Two PROFILE participants have thus given a talk at RadX during the academic year 2018/2019. Florian Hémont (sociology, *maître de conférences*) has presented at the Irisa laboratory his research results, obtained along the PROFILE project, about the privacy paradox. Erwan Picart (laws, PhD student funded by PROFILE) has presented his analysis of the European GDPR.

The Profile

Perspectives and follow-up

The PROFILE project will be directly followed by the international extension PROFILE-INT in the years 2019 and 2020. PROFILE-INT will focus on data privacy, algorithms transparency and algorithms bias, especially in the specific context of AI and laws. It will contribute to the collaboration between the IRISA laboratory and the Université du Québec à Montréal

(UQAM) by funding partially a joint PhD thesis, participations to conferences, as well as research stays between Rennes and Montréal.

With an extended team (LEGOS, IRISA, PREFICS, CRIStAL, and the COSTECH research laboratories) we are trying to follow this work through ANR funding. As the PROFILE project has shown us that people know little about the data collected from them, we would like to co-develop with users the application that PROFILE has helped us to implement. This application will have to be integrated into a set of meditation tools. This new project, "MyDataBuse", mainly concerns data privacy awareness. The "MyDataBuse" project passed the first selection step (selection rate 30%) at the ANR, but was unfortunately rejected at the second step (selection rate 30% of the projects selected in the first phase). The team plans to resubmit this project.