RESEARCH CENTRE

Saclay - Île-de-France

2020 ACTIVITY REPORT

Team COMETE

Concurrency, Mobility and Transactions

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions)

DOMAIN

Algorithmics, Programming, Software and Architecture

THEME

Security and Confidentiality

Contents

Team COMETE

Creation of the Project-Team: 2008 January 01

Keywords

Computer sciences and digital sciences

A2.1.1. – Semantics of programming languages

A2.1.5. – Constraint programming

A2.1.6. – Concurrent programming

A2.1.9. – Synchronous languages

A2.4.1. – Analysis

A3.4. – Machine learning and statistics

A3.5. – Social networks

A4.1. – Threat analysis

A4.5. – Formal methods for security

A4.8. – Privacy-enhancing technologies

A8.6. – Information theory

A8.11. – Game Theory

A9.1. – Knowledge

A9.2. – Machine learning

A9.7. – AI algorithmics

A9.9. – Distributed AI, Multi-agent

Other research topics and application domains

B6.1. – Software industry

B6.6. – Embedded systems

B9.5.1. – Computer science

B9.6.10. – Digital humanities

B9.9. – Ethics

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Catuscia Palamidessi [Team leader, Inria, Senior Researcher]
- Frank Valencia [CNRS, Researcher]

Post-Doctoral Fellows

- Sergio Ramirez Rico [Inria, from Nov 2020]
- Marco Romanelli [Centrale-Supélec, from Nov 2020]
- Gangsoo Zeong [Inria, from Oct 2020]

PhD Students

- Ruta Binkyte-Sadauskiene [Inria, from Dec 2020]
- Sayan Biswas [Inria, from Sep 2020]
- Ganesh Del Grosso Guzman [Inria]
- Natasha Fernandes [Université Macquarie Sydney - Australie]
- Federica Granese [Inria]
- Anna Pazii [École polytechnique, until Sep 2020]
- Carlos Pinzon [Inria, from Jul 2020]
- Santiago Quintero [École polytechnique]
- Marco Romanelli [Inria, until Oct 2020]

Technical Staff

- Ruta Binkyte-Sadauskiene [Inria, Engineer, from Sep 2020 until Nov 2020]
- Ehab Elsalamouny [Inria, Engineer, until Oct 2020]

Interns and Apprentices

- Lucas Bouju [Inria, from Apr 2020 until Oct 2020]

Administrative Assistant

- Maria Agustina Ronco [Inria]

External Collaborators

- Konstantinos Chatzikokolakis [CNRS]
- Mario Ferreira Alvim Junior [Universidade Federal de Minas Gerais - Brésil, until Feb 2020]

2 Overall objectives

The leading objective of COMETE is to develop a principled approach to privacy protection to guide the design of sanitization mechanisms in realistic scenarios. We aim to provide solid mathematical foundations were we can formally analyze the properties of the proposed mechanisms, considered as leading evaluation criteria to be complemented with experimental validation. In particular, we focus on privacy models that:

- allow the sanitization to be *applied and controlled directly by the user*, thus avoiding the need of a trusted party as well as the risk of security breaches on the collected data,
- are *robust with respect to combined attacks*, and
- provide an *optimal trade-off between privacy and utility.*

Two major lines of research are related to machine learning and social networks. These are prominent presences in nowadays social and economical fabric, and constitute a major source of potential problems. In this context, we explore topics related to the propagation of information, like *group polarization*, and other issues arising from the deep learning area, like *fairness* and *robustness with respect to adversarial inputs*, that have also a critical relation with privacy.

3 Research program

The objective of COMETE is to develop principled approaches to some of the concerns in today's technological and interconnected society: privacy, machine-learning-related security and fairness issues, and propagation of information in social networks.

3.1 Privacy

The research on privacy will be articulated in several lines of research.

3.1.1 Three way optimization between privacy and utility

One of the main problems in the design of privacy mechanisms is the preservation of the utility. In the case of local privacy, namely when the data are sanitized by the user before they are collected, the notion of utility is twofold:

Utility as quality of service (QoS): The user usually gives his data in exchange of some service, and in general the quality of the service depends on the precision of such data. For instance, consider a scenario in which Alice wants to use a LBS (Location-Based Service) to find some restaurant near her location x. The LBS needs of course to know Alice's location, at least approximately, in order to provide the service. If Alice is worried about her privacy, she may send to the LBS an approximate location x' instead of x. Clearly, the LBS will send a list of restaurants near x', so if x' is too far from x the service will degrade, while if it is too close Alice's privacy would be at stake.

Utility as statistical quality of the data (Stat): Bob, the service provider, is motivated to offer his service because in this way he can collect Alice's data, and quality data are very valuable for the big-data industry. We will consider in particular the use of the data collections for statistical purposes, namely for extracting general information about the population (and not about Alice as an individual). Of course, the more Alice's data are obfuscated, the less statistical value they have.

We intend to consider both kinds of utility, and study the "three way" optimization problem in the context of d-privacy, our approach to local differential privacy [31]. Namely, we want to develop methods for producing mechanisms that offer the best trade-off between d-privacy, QoS and Stat, at the same time. In order to achieve this goal, we will need to investigate various issues. In particular:

Figure 1: The central and the local models of differential privacy

- how to best estimate the original distribution from a collection of noisy data, in order to perform the intended statistical analysis,
- what metrics to use for assessing the statistical value of a distributions (for a given application), in order to reason about Stat, and
- how to compute in an efficient way the best noise from the point of view of the trade-off between d-privacy, QoS and Stat.

Estimation of the original distribution The only methods for the estimation of the original distribution from perturbed data that have been proposed so far in the literature are the iterative Bayesian update (IBU) and the matrix inversion (INV). The IBU is more general and based on solid statistical principles, but it is not ye well known in the in the privacy community, and it has not been studied much in this context. We are motivated to investigate this method because from preliminary experiments it seems more efficient on date obfuscated by geo-indistinguishability mechanisms (cfr. next section). Furthermore, we believe that the IBU is compositional, namely it can deal naturally and efficiently with the combination of data generated by different noisy functions, which is important since in the local model of privacy every user can, in principle, use a different mechanisms or a different level of noise. We intend to establish the foundations of the IBU in the context of privacy, and study its properties like the compositionality mentioned above, and investigate its performance in the state-of-the-art locally differentially private mechanisms.

Hybrid model An interesting line of research will be to consider an intermediate model between the local and the central models of differential privacy (cfr. Figure 1). The idea is to define a privacy mechanism based on perturbing the data locally, and then collecting them into a dataset organized as an histogram. We call this model "hibrid" because the collector is trusted like in central differential privacy, but the data are sanitized according to the local model. The resulting dataset would satisfy differential privacy from the point of view of an external observer, while the statistical utility would be as high as in the local model. One further advantage is that the IBU is compositional, hence the datasets sanitized in this way could be combined without any loss of precision in the application of the IBU. In other words, the statistical utility of the union of sanitized datasets is the same as the statistical utility of the sanitized union of datasets, which is of course an improvement (for the law of large numbers) wrt each separate dataset. One important application would be the cooperative sharing of sanitized data owned by different different companies or institution, to the purpose of improving statistical utility while preserving the privacy of their respective datasets.

3.1.2 Geo-indistinguishability

We plan to further develop our line of research on location privacy, and in particular, enhance our framework of geo-indistinguishability [4] (cfr. Figure 2) with mechanisms that allow to take into

Figure 2: Geo-indistinguishability is a framework to protect the privacy of the user when dealing with location-based services (a). The framework guarrantees d-privacy, a distance-based variant of differential privacy (b). The typical implementation uses (extended) Laplace noise (c).

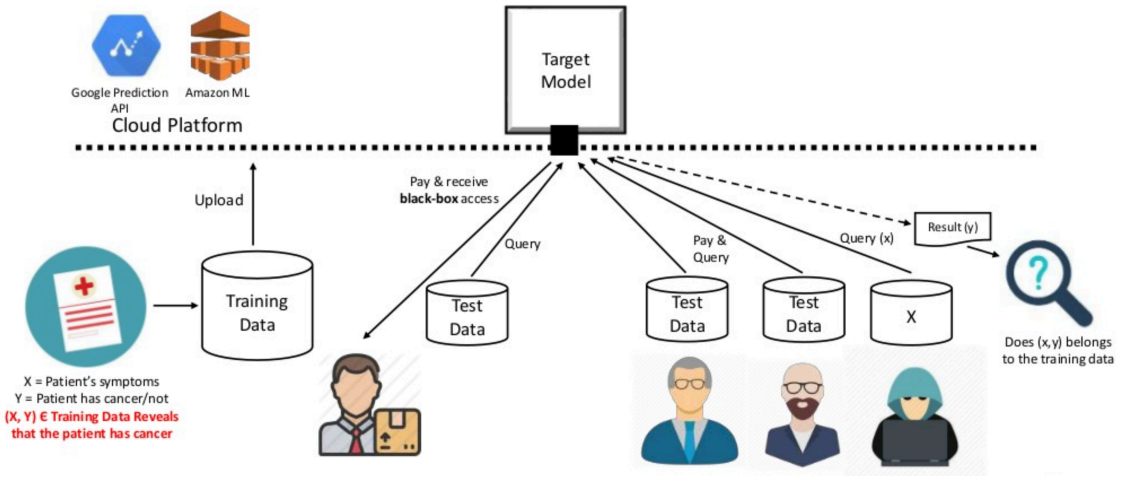


Figure 3: Privacy breach in machine learning as a service.

account sanitize high-dimensional traces without destroying utility (or privacy). One problem with the geo-indistinguishable mechanisms developed so far (the planar Laplace an the planar geometric) is that they add the same noise function uniformly on the map. This is sometimes undesirable: for instance, a user located in a small island in the middle of a lake should generate much more noise to conceal his location, so to report also other locations on the ground, because the adversary knows that it is unlikely that the user is in the water. Furthermore, for the same reason, it does not offer a good protection with respect to re-identification attacks: a user who lives in an isolated place, for instance, can be easily singled out because he reports locations far away from all others. Finally, and this is a common problem with all methods based on DP, the repeated use of the mechanism degrades the privacy, and even when the degradation is linear, as in the case of all DP-based methods, it becomes quickly unacceptable when dealing with highly structured data such as spatio-temporal traces.

3.1.3 Threats for privacy in machine learning

In recent years several researchers have observed that machine learning models leak information about the training data. In particular, in certain cases an attacker can infer with relatively high probability whether a certain individual participated in the dataset (*membership inference attack*)

od the value of his data (*model inversion attack*). This can happen even if the attacker has nop access to the internals of the model, i.e., under the *black box assumption*, which is the typical scenario when machine learning is used as a service (cfr. Figure 3). We plan to develop methods to reason about the information-leakage of training data from deep learning systems, by identifying appropriate measures of leakage and their properties, and use this theoretical framework as a basis for the analysis of attacks and for the development of robust mitigation techniques. More specifically, we aim at:

- Developing compelling case studies based on state-of-the-art algorithms to perform attacks, showcasing the feasibility of uncovering specified sensitive information from a trained software (model) on real data.

- Quantifying information leakage. Based on the uncovered attacks, the amount of sensitive information present in trained software will be quantified and measured. We will study suitable notions of leakage, possibly based on information-theoretical concepts, and establish firm foundations for these.

- Mitigating information leakage. Strategies will be explored to avoid the uncovered attacks and minimize the potential information leakage of a trained model.

3.1.4 Relation between privacy and robustness in machine learning

The relation between privacy and robustness, namely resilience to adversarial attacks, is rather complicated. Indeed the literature on the topic seems contradictory: on the one hand, there are works that show that differential privacy can help to mitigate both the risk of inference attacks and of misclassification (cfr. [35]). On the other hand, there are studies that show that there is a trade-off between protection from inference attacks and robustness [36]. We intend to shed light on this confusing situation. We believe that the different variations of differential privacy play a role in this apparent contradiction. In particular, *preprocessing* the training data with d-privacy seems to go along with the concept of robustness, because it guarantees that small variations in the input cannot result in large variations in the output, which is exactly the principle of robustness. On the other hand, the addition of random noise on the output result (*postprocessing*), which is the typical method in central DP, should reduce the precision and therefore increase the possibility of misclassification. We intend to make a taxonomy of the differential privacy variants, in relation to their effect on robustness, and develop a principled approach to protect both privacy and security in an optimal way.

One promising research direction for the deployment of d-privacy in this context is to consider Bayesian neural networks (BNNs). These are neural networks with distributions over their weights, which can capture the uncertainty within the learning model, and which provide a natural notion of distance (between distributions) on which we can define a meaningful notion of d-privacy. Such neural networks allow to compute an uncertainty estimate along with the output, which is important for safety-critical applications.

3.1.5 Relation between privacy and fairness

Both fairness and privacy are multi-faces notions, assuming different meaning depending on the application domain, on the situation, and on what exactly we want to protect. Fairness, in particular, has received many different definitions, some even in contrast with each other. One of the definitions of fairness is the property that similar “similar” input data produce "similar" outputs. Such notion corresponds closely to d-privacy. Other notions of fairness, however, are in opposition to standard differential privacy. This is the case, notably, of *Equalized Odds* [33] and of *Equality of False Positives* and *Equality of False Negatives* [32]. We intend to study a tassonomy of the relation between the main notions of fairness an the various variants of differential privacy. In particular, we intend to study the relation between the recently-introduced notions of *causal fairness* and *causal differential privacy* [37].

Another line of research related to privacy and fairness, that we intend to explore, is the design of to pre-process the training set so to obtain machine learning models that are both privacy-friendly and fair.

3.2 Quantitative information flow

In the area of quantitive information flow (QIF), we intend to pursue two lines of research: the study of non-0-sum games, and the estimation of g-leakage [30] under the black-box assumption.

3.2.1 Non-0-sum games

The framework of g-leakage does not take into account two important factors: (a) the loss of the user, and (b) the cost of the attack for the adversary. Regarding (a), we observe that in general the goal of the adversary may not necessarily coincide with causing maximal damage to the user, i.e., there may be a mismatch between the aims of the attacker and what the user tries to protect the most. To model this more general scenario, we had started investigating the interplay between defender and attacker in a game-theoretic setting, starting with the simple case of 0-sum games which corresponds to g-leakage. The idea was that, once the simple 0-sum case would be well understood, we would extend the study to the non-0-sum case, that is needed to represent (a) and (b) above. However, we had first to invent and lay the foundations of a new kind of games, the *information leakage games* [29] because the notion of leakage cannot be expressed in terms of payoff in standard game theory. Now that the theory of these new games is well established, we intend to go ahead with our plan, namely study costs and damages of attacks in terms of non-0-sum information leakage games.

3.2.2 Black-box estimation of leakage via machine learning

Most of the works in QIF rely on the so-called white-box assumption, namely, they assume that it is possible to compute exactly the (probabilistic) input-output relation of the system, seen as an information-theoretic channel. This is necessary in order to apply the formula that expresses the leakage. In practical situations, however, it may not be possible to compute the input-output relation, either because the system is too complicated, or simply because it is not accessible. Such scenario is called black-box. The only assumption we make is that the adversary can interact with the system, by feeding to it inputs of his choice and observing the corresponding outputs.

Given the practical interest of the black-box model, we intend to study methods to estimate its leakage. Clearly the standard QIF methods are not applicable. We plan to use, instead, a machine learning approach, continuing the work we started in [20]. In particular, we plan to investigate whether we can improve the efficiency of the method proposed by leveraging on the experience that we have acquired with the GANs [19]. The idea is to construct a training set and a testing set from the input-output samples collected by interacting with the system, and then build a classifier that learns from the training set to classify the input from the output so to maximize its gain. The measure of its performance on the testing set should then give an estimation of the posterior g-vulnerability.

3.3 Information leakage, bias and polarization in social networks

One of the core activities of the team will be the study of how information propagate in the highly interconnected scenarios made possible by modern technologies. We will consider the issue of privacy protection as well as the social impact of privacy leaks. Indeed, recent events have shown that social networks are exposed to actors malicious agents that can collect *private information* of millions of users with or without their consent. This information can be used to build psychological profiles for microtargeting, typically aimed at discovering users preconceived beliefs and at reinforcing them. This may result in polarization of opinions as people with opposing views would tend to interpret new information in a biased way causing their views to move further apart. Similarly, a group with uniform views often tends to make more extreme decisions than its individual. As a result, users

may become more radical and isolated in their own ideological circle causing dangerous splits in society.

3.3.1 Privacy protection

In [14] we have investigated potential leakage in social networks, namely, the unintended propagation and collection of confidential information. We intend to enrich this model with epistemic aspects, in order to take into account the belief of the users and how it influences the behavior of agents with respect the transmission of information.

Furthermore, we plan to investigate attack models used to reveal a user's private information, and explore the framework of g-leakage to formalize the privacy threats. This will provide the basis to study suitable protection mechanisms.

3.3.2 Polarization and Belief in influence graphs

In social scenarios, a group may shape their beliefs by attributing more value to the opinions of influential figures. This cognitive bias is known as *authority bias*. Furthermore, in a group with uniform views, users may become extreme by reinforcing one another's opinions, giving more value to opinions that confirm their own beliefs; another common cognitive bias known as *confirmation bias*. As a result, social networks can cause their users to become radical and isolated in their own ideological circle causing dangerous splits in society (polarization). We intend to study these dynamics in a model called *influence graph*, which is a weighted directed graph describing connectivity and influence of each agent over the others. We will consider two kinds of belief updates: the authority belief update, which gives more value to the opinion of agents with higher influence, and the confirmation bias update, which gives more value to the opinion of agents with similar views.

We plan to study the evolution of polarization in these graphs. In particular, we aim at defining a suitable measure of polarization, characterizing graph structures and conditions under which polarization eventually converges to 0 (vanishes), and methods to compute the change in the polarization value over time.

Another purpose of this line of research is how the bias of the agents whose data are being collected impacts the *fairness* of learning algorithms based on these data.

3.3.3 Concurrency models for the propagation of information

Due to their popularity and computational nature, social networks have exacerbated group polarization. Existing models of group polarization from economics and social psychology state its basic principles and measures [34]. Nevertheless, unlike our computational ccp models, they are not suitable for describing the dynamics of agents in distributed systems. Our challenge is to coherently combine our ccp models for epistemic behavior with principles and techniques from economics and social psychology for GP. We plan to develop a ccp-based process calculus which incorporates structures from social networks, such as communication, influence, individual opinions and beliefs, and privacy policies. The expected outcome is a *computational model* that will allow us to specify the interaction of groups of agents exchanging *epistemic information* among them and to predict and measure the *leakage of private information*, as well as the *degree of polarization* that such group may reach.

4 Application domains

The application domains of our research include the following:

Protection of sensitive personal data Our lives are growingly entangled with internet-based technologies and the limitless digital services they provide access to. The ways we communicate, work, shop, travel, or entertain ourselves are increasingly depending on these services. In turn, most such services heavily rely on the collection and analysis of our personal data, which are often

generated and provided by ourselves: tweeting about an event, searching for friends around our location, shopping online, or using a car navigation system, are all examples of situations in which we produce and expose data about ourselves. Service providers can then gather substantial amounts of such data at unprecedented speed and at low cost.

While data-driven technologies provide undeniable benefits to individuals and society, the collection and manipulation of personal data has reached a point where it raises alarming privacy issues. Not only the experts, but also the population at large are becoming increasingly aware of the risks, due to the repeated cases of violations and leaks that keep hitting the headlines. Examples abound, from iPhones storing and uploading device location data to Apple without users' knowledge to the popular Angry Birds mobile game being exploited by NSA and GCHQ to gather users' private information such as age, gender and location.

If privacy risks connected to personal data collection and analysis are not addressed in a fully convincing way, users may eventually grow distrustful and refuse to provide their data. On the other hand, misguided regulations on privacy protection may impose excessive restrictions that are neither necessary nor sufficient. In both cases, the risk is to hinder the development of many high-societal-impact services, and dramatically affect the competitiveness of the European industry, in the context of a global economy which is more and more relying on Big Data technologies.

The EU General Data Protection Regulation (GDPR) imposes that strong measures are adopted by-design and by-default to guarantee privacy in the collection, storage, circulation and analysis of personal data. However, while regulations set the high-level goals in terms of privacy, it remains an open research challenge to map such high-level goals into concrete requirements and to develop privacy-preserving solutions that satisfy the legally-driven requirements. The current de-facto standard in personal data sanitization used in the industry is anonymization (i.e., personal identifier removal or substitution by a pseudonym). Anonymity however does not offer any actual protection because of potential *linking attacks* (which have actually been known since a long time). Recital 26 of the GDPR states indeed that anonymization may be insufficient and that anonymized data must still be treated as personal data. However the regulation provide no guidance on how or what constitutes an effective data re-identification scheme, leaving a grey area on what could be considered as adequate sanitization.

In COMETE, we pursue the vision of a world where pervasive, data-driven services are inalienable life enhancers, and at the same time individuals are fully guaranteed that the privacy of their sensitive personal data is protected. Our objective is to develop a principled approach to the design of sanitization mechanisms providing an optimal trade-off between privacy and utility, and robust with respect to composition attacks. We aim at establishing solid mathematical foundations were we can formally analyze the properties of the proposed mechanisms, which will be regarded as leading evaluation criteria, to be complemented with experimental validation.

We focus on privacy models where the sanitization can be applied and controlled directly by the user, thus avoiding the need of a trusted party as well as the risk of security breaches on the collected data.

Ethical machine learning Machine learning algorithms have more and more impact on and in our day-to-day lives. They are already used to take decisions in many social and economical domains, such as recruitment, bail resolutions, mortgage approvals, and insurance premiums, among many others. Unfortunately, there are many ethical challenges:

- Lack of transparency of machine learning models: decisions taken by these machines are not always intelligible to humans, especially in the case of neural networks.
- Machine learning models are not neutral: their decisions are susceptible to inaccuracies, discriminatory outcomes, embedded or inserted bias.
- Machine learning models are subject to privacy and security attacks, such as data poisoning and membership and attribiute inference attacks.

The time has therefore arrived that the most important area in machine learning is the implementation of algorithms that adhere to ethical and legal requirements. For example, the

United States' Fair Credit Reporting Act and European Union's General Data Protection Regulation (GDPR) prescribe that data must be processed in a way that is fair/unbiased. GDPR also alludes to the right of an individual to receive an explanation about decisions made by an automated system.

One of the goals of COMETE's research is to contribute to make the machine learning technology evolve towards compliance with the human principles and rights, such as fairness and privacy, while continuing to improve accuracy and robustness.

Polarization in Social Networks *Distributed systems* have changed substantially with the advent of social networks. In the previous incarnation of distributed computing the emphasis was on consistency, fault tolerance, resource management and other related topics. What marks the new era of distributed systems is an emphasis on the flow of *epistemic* information (knowledge, facts, opinions,beliefs and lies) and its impact on democracy and on society at large.

Indeed in social networks a group may shape their beliefs by attributing more value to the opinions of influential figures. This cognitive bias is known as *authority bias*. Furthermore, in a group with uniform views, users may become extreme by reinforcing one another's opinions, giving more value to opinions that confirm their own beliefs; another common cognitive bias known as *confirmation bias*. As a result, social networks can cause their users to become radical and isolated in their own ideological circle causing dangerous splits in society in a phenomenon known as *polarization*.

One of our goals in COMETE is to study the flow of epistemic information in social networks and its impact on opinion shaping and social polarization. We study models for reasoning about distributed systems whose agents interact with each other like in social networks; by exchanging epistemic information and interpreting it under different biases and network topologies. We are interested in predicting and measuring the degree of polarization that such agents may reach. We focus on polarization with strong influence in politics such as affective polarization; the dislike and distrust those from the other political party. We expect the model to provide social networks with guidance as to how to distribute newsfeed to mitigate polarization.

5 Highlights of the year

5.1 Project CRYPTECS: Cloud-Ready Privacy-Preserving Technologies

Our project CRYPTECS has been accepted !

Program: ANR-BMBF French-German Joint Call on Cybersecurity.

Goals: The project aims at building an open source cloud platform promoting the adoption of privacy-preserving computing (PPC) technology by offering a broad spectrum of business-ready PPC techniques (Secure Multiparty Computation, Homomorphic Encryption, Trusted Execution Environments, and methods for Statistical Disclosure Control, in particular Differential Privacy) as reusable and composable services.

Consortium: Besides Inria (COMETE), the other partners are: the Orange Group (France), the Bosh Group (Germany), the university of Stuttgart (Germany), and three SME's: Zama (a spin-off of CryptoExperts, France), Aircloak (Germany), and Edgeless Systems (Germany)

5.2 QIF book

Our book on Quantitative Information Flow has appeared !

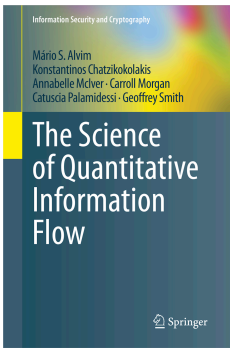
Authors: Mário S. Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, Geoffrey Smith

Title: The Science of Quantitative Information Flow

Series: Information Security and Cryptography

Publisher: Springer International Publishing

Number of pages: XXVIII, 478
Year: 2020
Reference: [21]



6 New software and platforms

6.1 New software

6.1.1 libqif - A Quantitative Information Flow C++ Toolkit Library

Keywords: Information leakage, Privacy, C++, Linear optimization

Functional Description: The goal of libqif is to provide an efficient C++ toolkit implementing a variety of techniques and algorithms from the area of quantitative information flow and differential privacy. We plan to implement all techniques produced by Com\'ete in recent years, as well as several ones produced outside the group, giving the ability to privacy researchers to reproduce our results and compare different techniques in a uniform and efficient framework.

Some of these techniques were previously implemented in an ad-hoc fashion, in small, incompatible with each-other, non-maintained and usually inefficient tools, used only for the purposes of a single paper and then abandoned. We aim at reimplementing those – as well as adding several new ones not previously implemented – in a structured, efficient and maintainable manner, providing a tool of great value for future research. Of particular interest is the ability to easily re-run evaluations, experiments, and case-studies from QIF papers, which will be of great value for comparing new research results in the future.

The library's development continued in 2020 with several new added features. 68 new commits were pushed to the project's git repository during this year. The new functionality was directly applied to the experimental results of several publications of COMETE.

URL: https://github.com/chatziko/libqif

Contact: Konstantinos Chatzikokolakis

6.1.2 Location Guard

Keywords: Privacy, Geolocation, Browser Extensions

Scientific Description: The purpose of Location Guard is to implement obfuscation techniques for achieving location privacy, in a an easy and intuitive way that makes them available to the general public. Various modern applications, running either on smartphones or on the web, allow third parties to obtain the user's location. A smartphone application can obtain this information from the operating system using a system call, while web application obtain it from the browser using a JavaScript call.

Functional Description: Websites can ask the browser for your location (via JavaScript). When they do so, the browser first asks for your permission, and if you accept, it detects your location (typically by transmitting a list of available wifi access points to a geolocation provider such as Google Location Services, or via GPS if available) and gives it to the website.

Location Guard is a browser extension that intercepts this procedure. The permission dialog appears as usual, and you can still choose to deny it. If you give permission, then Location Guard obtains your location and adds "random noise" to it, creating a fake location. Only the fake location is then given to the website.

Location Guard is by now a stable tool with a large user base. No new features were added in 2020, however, the tool is still actively maintained.

URL: https://github.com/chatziko/location-guard

Contact: Konstantinos Chatzikokolakis

Participants: Catuscia Palamidessi, Konstantinos Chatzikokolakis, Marco Stronati, Miguel Andrés, Nicolas Bordenabe

6.1.3 IBU: A java library for estimating distributions

Keywords: Privacy, Statistic analysis, Bayesian estimation

Functional Description: The main objective of this library is to provide an experimental framework for evaluating statistical properties on data that have been sanitized by obfuscation mechanisms, and for measuring the quality of the estimation. More precisely, it allows modeling the sensitive data, obfuscating these data using a variety of privacy mechanisms, estimating the probability distribution on the original data using different estimation methods, and measuring the statistical distance and the Kantorovich distance between the original and estimated distributions. This is one of the main software projects of Palamidessi's ERC Project HYPATIA.

We intend to extend the software with functionalities that will allow estimating statistical properties of multi-dimensional (locally sanitized) data and using collections of data locally sanitized with different mechanisms.

URL: https://gitlab.com/locpriv/ibu

Contact: Ehab ElSalamouny

6.1.4 F-BLEAU

Name: F-BLEAU

Keywords: Information leakage, Machine learning, Privacy

Functional Description: F-BLEAU is a tool for estimating the leakage of a system about its secrets in a black-box manner (i.e., by only looking at examples of secret inputs and respective outputs). It considers a generic system as a black-box, taking secret inputs and returning outputs accordingly, and it measures how much the outputs "leak" about the inputs.

F-BLEAU is based on the equivalence between estimating the error of a Machine Learning model of a specific class and the estimation of information leakage.

This code was also used for the experiments of a COMETE publication that appeared in S&P 2019 , on the following evaluations: Gowalla, e-passport, and side-channel attack to finite field exponentiation.

The software is maintained and some new features were added in 2020.

Release Contributions: First F-BLEAU release. Supports frequentist and k-NN estimates with several parameters, and it allows stopping according to delta-convergence criteria.

URL: https://github.com/gchers/fbleau

Contacts: Giovanni Cherubin, Konstantinos Chatzikokolakis, Catuscia Palamidessi

6.1.5 MILES

Name: ML Leakage Estimation

Keywords: Information leakage, Machine learning

Functional Description: This software provides a tool for estimating the g-leakage of a system in the black-box setting, i.e., when the true posterior distributions of the outputs given the inputs are unknown, and the only available knowledge comes from observing input-output examples.

The tool is based on two methods: The first one relies on the Artificial Neural Networks' ability to output probability distributions, which can be used to directly estimate the g-leakage. The second method is based on a preprocessing of the data to be used for the training phase. In practice, the pre-processing reduces the problem of g-leakage estimation to that of estimating the Bayes risk, a task that can be achieved by generating an approximation of the Bayes classifier, using any universally consistent learning rule.

This package is a software project of Palamidessi's ERC Project HYPATIA.

URL: https://gitlab.com/marcoromane.gitlab.public/miles_server_version

Contacts: Catuscia Palamidessi, Marco Romanelli

6.1.6 MIPAN

Name: Mutual Information Privacy Adversarial Networks

Keywords: Privacy, Neural networks

Functional Description: This package provides a GAN (Generative Adversarial Network) to produce an optimal mechanism for privacy protection.

The system consists of two nets: the generator, which tries to produce an optimal obfuscation mechanism to protect the data, and the classifier, which tries to de-obfuscate the data. By letting the two nets compete against each other, the mechanism improves its degree of protection, until an equilibrium is reached.

The package contains an application to the case of location privacy, and experiments performed on synthetic data and on real data from the Gowalla dataset (https://snap.stanford.edu/data/loc-gowalla_totalCheckins.txt.gz).

This package is a software project of Palamidessi's ERC Project HYPATIA.

URL: https://gitlab.com/MIPAN/mipan

Contacts: Catuscia Palamidessi, Marco Romanelli

6.1.7 MinEntropyFeatureSelection

Name: Feature Selection for Machine Learning

Keywords: Machine learning, Entropy

Functional Description: This is a library for feature selection, namely a tool to reduce the number of features to be considered by an algorithm for machine learning. It can be used to make the learning phase more efficient and more accurate.

The idea is to try to find a minimal set of features that contain the maximal information about the labeling task. The tool works iteratively in a greedy fashion, and at each step, it

adds to the set a feature that is as independent as possible from those that have been selected. The metric for the independence test is mutual information, and the user can choose between Shannon or Rényi mutual information.

URL: https://gitlab.com/marcoromane.gitlab.public/minentropyfeatureselection

Contact: Marco Romanelli

6.1.8 dspacenet

Name: Distributed-Spaces Network.

Keywords: Social networks, Distributed programming

Functional Description: DSpaceNet is a tool for social networking based on multi-agent spatial and timed concurrent constraint language.

I - The fundamental structure of DSPaceNet is that of *space*: A space may contain

(1) spatial-mobile-reactive tcc programs, and (2) other spaces.

Furthermore, (3) each space belongs to a given agent. Thus, a space of an agent j within the space of agent i means that agent i allows agent j to use a computation sub-space within its space.

II - The fundamental operation of DSPaceNet is that of *program posting*: In each time unit, agents can post spatial-mobile-reactive tcc programs in the spaces they are allowed to do so (ordinary message posting corresponds to the posting of tell processes). Thus, an agent can for example post a watchdog tcc process to react to messages in their space, e.g. whenever (*happy b*frank*) do tell("thank you!"). More complex mobile programs are also allowed (see below).

The language of programs is a spatial mobile extension of tcc programs:

$P,Q... : tell(c)|whencdoP||nextP|P||Q|unlesscnextP|[P]_i|\uparrow_i P|recX.P$

Computation of timed processes proceeds as in tcc. The spatial construct $[P]_i$ runs P in the space of agent i and the mobile process $\uparrow_i P$, extrudes P from the space of i. By combining space and mobility, arbitrary processes can be moved from one a space into another. For example, one could send a trojan watchdog to another space for spying for a given message and report back to one's space.

III- Constraint systems can be used to specify advance text message deduction, arithmetic deductions, scheduling, etc.

IV - Epistemic Interpretation of spaces can be used to derive whether they are users with conflicting/inconsistent information, or whether a group of agents may be able to deduce certain message.

V - The scheduling of agent requests for program posts, privacy settings, friendship lists are handled by an external interface. For example, one could use type systems to check whether a program complies with privacy settings (for example checking that the a program does not move other program into a space it is not allowed into).

URL: http://dspacenet.javerianacali.edu.co/

Contact: Frank Valencia

Partner: Pontificia Universidad Javeriana Cali

7 New results

7.1 A Logical Characterization of Differential Privacy

Differential privacy is a formal definition of privacy ensuring that sensitive information relative to individuals cannot be inferred by querying a database. In [11], we have exploited a modeling of this framework via labeled Markov Chains (LMCs) to provide a logical characterization of differential privacy: we have considered a probabilistic variant of the Hennessy-Milner logic and we have defined a syntactical distance on for-mulae in it measuring their syntactic disparities. Then, we have defined a trace distance on LMCs in terms of the syntactic distance between the sets of formulae satisfied by them. We have proved that such distance corresponds to the level of privacy of the LMCs. Moreover, we have used the distance on formulae to define a real-valued semantics for them, from which we have obtained a logical characterization of weak anonymity where the level of anonymity is measured in terms of the smallest formula distinguishing the considered LMCs. Then, we have focused on bisimulation semantics on nondeterministic probabilistic processes and we have provided a logical characterization of the generalized bisimulation metrics, namely those defined via the generalized Kantorovich lifting. Our characterization is based on the notion of mimicking formula of a process and the syntactic distance on formulae, where the former captures the observable behavior of the corresponding process and allows us to characterize bisimilarity. We have shown that the generalized bisimulation distance on processes is equal to the syntactic distance on their mimicking formulae. Moreover, we have used the distance on mimicking formulae to obtain bounds on differential privacy.

7.2 Refinement Orders for Quantitative Information Flow and Differential Privacy

Quantitative Information Flow (QIF) and Differential Privacy (DP) are both concerned with the protection of sensitive information, but they are rather different approaches. In particular, QIF considers the expected probability of a successful attack, while DP (in both its standard and local versions) is a max-case measure, in the sense that it is compromised by the existence of a possible attack, regardless of its probability. Comparing systems is a fundamental task in these areas: one wishes to guarantee that replacing a system A by a system B is a safe operation that is the privacy of B is no worse than that of A. In QIF, a refinement order provides strong such guarantees, while, in DP, mechanisms are typically compared w.r.t. the privacy parameter ε in their definition. In [12], we have explored a variety of refinement orders, inspired by the one of QIF, providing precise guarantees for max-case leakage. We have studied simple structural ways of characterising the relation between them, and efficient methods for verifying them and their lattice properties. Moreover, we have applied these orders to the task of comparing DP mechanisms, raising the question of whether the order based on ε provides strong privacy guarantees. We have shown that, while it is often the case for mechanisms of the same "family" (geometric, randomised response, etc.), it rarely holds across different families.

7.3 Full Convergence of the Iterative Bayesian Update and Applications to Mechanisms for Privacy Protection

The iterative Bayesian update (IBU) and the matrix inversion (INV) are the main methods to retrieve the original distribution from noisy data resulting from the application of privacy protection mechanisms. In [16] we have shown that the theoretical foundations of the IBU established in the literature are flawed, as they rely on an assumption that in general is not satisfied in typical real datasets. We then have fixed the theory of the IBU, by providing a general convergence result for the underlying Expectation-Maximization method. Our framework does not rely on the above assumption, and also covers a more general local privacy model. Finally we have evaluated the precision of the IBU on data sanitized with the Geometric, k-RR, and RAPPOR mechanisms, and we have shown that it outperforms INV in the first case, while it is comparable to INV in the other two cases.

7.4 Estimating g-Leakage via Machine Learning

In [20] we have considered the problem of estimating the information leakage of a system in the black-box scenario. It is assumed that the system's internals are unknown to the learner, or anyway too complicated to analyze, and the only available information are pairs of input-output data samples, possibly obtained by submitting queries to the system or provided by a third party. Previous research has mainly focused on counting the frequencies to estimate the input-output conditional probabilities (referred to as frequentist approach), however this method is not accurate when the domain of possible outputs is large. To overcome this difficulty, the estimation of the Bayes error of the ideal classifier was recently investigated using Machine Learning (ML) models and it has been shown to be more accurate thanks to the ability of those models to learn the input-output correspondence. However, the Bayes vulnerability is only suitable to describe one-try attacks. A more general and flexible measure of leakage is the g-vulnerability, which encompasses several different types of adversaries, with different goals and capabilities. In [20], we have proposed a novel approach to perform black-box estimation of the g-vulnerability using ML. A feature of our approach is that it does not require to estimate the conditional probabilities, and that it is suitable for a large class of ML algorithms. First, we have formally shown the learnability for all data distributions. Then, we have evaluated the performance via various experiments using k-Nearest Neighbors and Neural Networks. Our results outperform the frequentist approach when the observables domain is large.

7.5 Optimal Obfuscation Mechanisms via Machine Learning

In [19] we have considered the problem of obfuscating sensitive information while preserving utility, and we propose a machine learning approach inspired by the generative adversarial networks paradigm. The idea is to set up two nets: the generator, that tries to produce an optimal obfuscation mechanism to protect the data, and the classifier, that tries to de-obfuscate the data. By letting the two nets compete against each other, the mechanism improves its degree of protection, until an equilibrium is reached. We have applied our method to the case of location privacy, and we have performed experiments on synthetic data and on real data from the Gowalla dataset. We have evaluated the privacy of the mechanism not only by its capacity to defeat the classifier, but also in terms of the Bayes error, which represents the strongest possible adversary. We have compared the privacy-utility tradeoff of our method to that of the planar Laplace mechanism used in geo-indistinguishability, showing favorable results. Like the Laplace mechanism, our system can be deployed at the user end for protecting his location.

7.6 Applications of Game-Theoretic Principles

Game theory is the study of the strategic behavior of rational decision makers who are aware that their decisions affect one another. Its simple but universal principles have found applications in the most diverse disciplines, including economics, social sciences, evolutionary biology, as well as logic, system science and computer science. Despite its long-standing tradition and its many advances, game theory is still a young and developing science. In [17], we have described some recent and intriguing applications in the fields of machine learning and privacy.

7.7 Derivation of Constraints from Machine Learning Models and Applications to Security and Privacy

In [22] we have shown how to combine the power of machine learning with the flexibility of constraints. More specifically, we have shown how machine learning models can be represented by first-order logic theories, and how to derive these theories. The advantage of this representation is that it can be augmented with additional formulae, representing constraints of some kind on the data domain. For instance, new knowledge, or potential attackers, or fairness desiderata. We have considered various kinds of learning algorithms (neural networks, k-nearest-neighbours, decision trees, support vector machines) and for each of them we have shown how to infer the FOL formulae.

We have focused on one particular application domain, namely the field of security and privacy. The idea is to represent the potentialities and goals of the attacker as a set of constraints, then use a constraint solver (more precisely, a solver modulo theories) to verify the satisfiability. If a solution exists, then it means that an attack is possible, otherwise, the system is safe. We have shown various examples from different areas of security and privacy; specifically, a side-channel attack on a password checker, a malware attack on smart health systems, and a model-inversion attack on a neural network.

7.8 Enhanced Models for Privacy and Utility in Continuous-Time Diffusion Networks

Controlling the propagation of information in social networks is a problem of growing importance. On one hand, users wish to freely communicate and interact with their peers. On the other hand, the information they spread can bring to harmful consequences if it falls in the wrong hands. There is therefore a trade-off between utility, i.e., reaching as many intended nodes as possible, and privacy, i.e., avoiding the unintended ones. The problem has attracted the interest of the research community: some models have already been proposed to study how information propagates and to devise policies satisfying the intended privacy and utility requirements. In [14] we have adapted the basic framework of Backes et al. to include more realistic features, that in practice influence the way in which information is passed around. More specifically, we have considered: (a) the topic of the shared information, (b) the time spent by users to forward information among them and (c) the user social behaviour. For all features, we have shown a way to reduce our model to the basic one, thus allowing the methods provided in the original paper to cope with our enhanced scenarios. Furthermore, we have proposed an enhanced formulation of the utility/privacy policies, to maximize the expected number of reached users among the intended ones, while minimizing this number among the unintended ones, and we have shown how to adapt the basic techniques to these enhanced policies. Finally, we have provided a new approach to the maximization/minimization problem by finding a trade-off between the risk and the gain function through bi-objctive optimization.

7.9 Dynamic slicing for Concurrent Constraint Languages

Concurrent Constraint Programming (CCP) is a declarative model for concurrency where agents interact by telling and asking constraints (pieces of information) in a shared store. Some previous works have developed (approximated) declarative debuggers for CCP languages. However, the task of debugging concurrent programs remains difficult. In [13] we have defined a dynamic slicer for CCP (and other language variants) and we have showed that it is a useful companion tool for the existing debugging techniques. The debugger operates on partial computations (traces) that show the presence of bugs. Often, the quantity of information in such a trace is overwhelming, and the user gets easily lost, since she cannot focus on the sources of the bugs. Our slicer allows for marking part of the state of the computation and assists the user to eliminate most of the redundant information in order to highlight the errors. We have shown that this technique can be tailored to several variants of CCP, such as the timed language ntcc, linear CCP (an extension of CCP-based on linear logic where constraints can be consumed) and some extensions of CCP dealing with epistemic and spatial information. We have also developed a prototypical implementation freely available for making experiments.

7.10 Counting and Computing Join-Endomorphisms in Lattices

Structures involving a lattice order and the join-endomorphisms on it are ubiquitous in computer science. In particular they can be used to represent agents' knowledge and beliefs in distributed systems and economy. In [18] we have studied and given closed formulas for the cardinality of the set of all join-endomorphisms of for linear, discrete and powerset lattices. We have also studied the problem of computing the greatest lower bound of two given join-endomorphisms over lattice orders. We have proven that this problem can be solved, in the worst-case, in linear time for for powerset lattices, quadratic time for lattices of sets, and cubic for arbitrary lattices. The complexity

is expressed in terms of the size of the lattice and the basic binary lattice operations performed by the algorithm.

8 Partnerships and cooperations

8.1 International initiatives

8.1.1 Equipe Associée

Achronym: LOGIS

Title: Logical and Formal Methods for Information Security

Program: Inria Associate Teams

Duration: Jan 1 2019 - Dec 31, 2021

Coordinator: Catuscia Palamidessi

Partners:

- Keio University (Japan), Mitsuhiro Okada
- AIST (Japan), Yusuke Kawamoto
- JAIST (Japan), Tachio Terauchi
- University of Tokyo (Japan), Masami Hagiya
- Inria (France), Catuscia Palamidessi
- LSV, UPS (France), Hubert Comon

Inria contact: Catuscia Palamidessi

Description: With the ever-increasing use of internet-connected devices, such as computers, IoT appliances and GPS-enabled equipments, personal data are collected in larger and larger amounts, and then stored and manipulated for the most diverse purposes. Although privacy is of fundamental importance for the users of these systems, the protection of personal data is challenging for a variety of reasons.

First, personal data can be leaked due to numerous attacks on cryptographic protocols, often affecting those that were long thought to be secure. Second, partially releasing personal data is often desirable, either to access a desired service (e.g. Location-Based Services), or to collect statistics, which provides enormous benefits to individuals and society.

To address these challenges, our project aims at advancing the state of the art of (A) protocol verification and (B) privacy control. The two approaches are complementary, addressing different types of information leaks: those caused by flaws in the protocol (A) and those caused by the partial (voluntary or not) release of information (B).

8.1.2 International partners

COMETE has several international partners. In the following, we list these collaborations according to the various topics.

g-leakage On the topic of quantitative information flow, the collaboration with the following scientists has been extremely important for COMETE. In particular, our collaboration has produced the framework of g-leakage that allows to formalize the leakage of information for a large class of adversaries. In 2020 we have written a book on quantitative information flow [21], which, as far as we know, is the first book on information flow that considers the probabilistic dimension.

- Mario Ferreira Alvim Junior, Assistant Professor, Federal University of Minas Gerais, Brazil

- Kostas Chatzikokolakis, Associate Professor at the Univ. of Athens, on leave from CNRS where is is researcher.
- Carroll Morgan, Senior Scientist at NICTA , Australia, and Full Professor at The University of New South Wales, Australia
- Annabelle McIver, Full Professor at Maquarie University, Australia
- Geoffrey Smith, Full Professor, Florida International University, USA

Our collaboration continues on the topics of quantitative information flow and privacy. Annabelle McIver is co-supervising with Catuscia Palamidessi a PhD student of COMETE (Natasha Fernandez).

***d*-privacy and geo-indistinguishability** Konstantinos Chatzikokolakis, University of Athens, Greece, in leave from CNRS, has been member of COMETE until 2018. He has played a crucial role in the development and of the frameworks of *d*-privacy and geo-indistinguishability, and related implementations. Our collaboration continues on the topics of privacy and of quantitative information flow.

Information leakage and machine learning On the topics involving the estimation of information leakage via machine learning, the collaboration with the following scientists has been extremely beneficial.

- Giovanni Cherubin, Researcher, Turing Institute, UK.
- Pablo Piantanida, Associate Professor at Centrale Supélec, researcher at CNRS, on leave to MILA (Canada), the Quebec AI Institute.

The collaboration with Giovanni Cherubin continues on the topics of quantitative information flow applied to inference attacks in machine learning. The collaboration with Pablo Piantanida continues on the topics of information leakage in deep learning, robustness of machine learning models, and fairness in machine learning. Pablo Piantanida is also co-supervising with Catuscia Palamidessi and Frank Valencia various COMETE PhD students (Ganesh Del Grosso, Federica Granese, and Carlos Pinzón).

Robustness of machine learning models COMETE is collaborating with Daniele Gorla (Full professor at the university of Roma "La Sapienza") on the topic of resilience to adversarial attacks to machine learning. Daniele Gorla and Pablo Piantanida are co-supervising with Catuscia Palamidessi a COMETE PhD students (Federica Granese).

Privacy in social networks COMETE is also collaborating with Daniele Gorla on the topic of propagation of information and protection of private data in social networks.

Concurrency epistemic models for social networks The collaboration with the following people has been useful for the development of an epistemic version of *concurrent constraint programming*, a calculus for concurrent process that has been used to represent and implement a model for propagation of information in social networks, as well as to study phenomena like polarization.

- Carlos Olarte, Universidade Federal do Rio Grande do Norte, Brazil
- Camilo Rocha, Associate Professor, Universidad Javeriana de Cali, Colombia
- Camilo Rueda, Professor, Universidad Javeriana de Cali, Colombia

Our collaboration continues on the topics of polarization and evolution of bias.

8.1.3 Participation in other international programs

Achronym: FACTS

Title: Foundational Approach to Cognition in Today's Society

Program: ECOS NORD

Duration: Jan 1 2019 - Dec 31, 2021

Coordinator: Frank Valencia, CNRS, Ecole Polytechnique.

Partners:

- Inria (France), Frank Valencia
- LIP6, Sorbonne University (France), Jean-Gabriel Ganascia
- Universidad Javeriana de Cali (Colombia), Camilo Rueda

Inria contact: Frank Valencia

Description: This projects aims at studying the phenomenon of "Group Polarization"; the tendency for a group to learn or acquire beliefs or to make decisions that are more extreme than the initial inclinations of its members.

8.2 International research visitors

Frank Valencia visited the research group AVISPA from Pontificia Universidad Javeriana in the context of our ECOSNORD project FACTS in February 2020.

Usually COMETE hosts several visitors every year, and COMETE researchers pay visits to other institutions, but in 2020, due to the pandemic, most of our visitors and visiting plans had to be postponed.

8.3 European initiatives

8.3.1 ERC grant

Achronym: HYPATIA

Title: Privacy and Utility Allied

Program: European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme.

Duration: October 2019 – September 2024

Principal Investigator: Catuscia Palamidessi

URL: https://project.inria.fr/hypatia/

Description: The objective of this project is to develop the theoretical foundations, methods and tools to protect the privacy of the individuals while letting their data to be collected and used for statistical purposes. We aim in particular at developing mechanisms that can be applied and controlled directly by the user thus avoiding the need of a trusted party, are robust with respect to combination of information from different sources, and provide an optimal trade-off between privacy and utility.

8.3.2 Collaboration with major European organizations

Achronym: CRYPTECS

Title: Cloud-Ready Privacy-Preserving Technologies

Program: ANR-BMBF French-German Joint Call on Cybersecurity

Duration: June 1, 2021 - May 31, 2024

Coordinators: Baptiste Olivier and Sven Trieflinger

Partners:

- Orange (France), Baptiste Olivier
- The Bosch Group (Germany) Sven Trieflinger
- Inria (France), Catuscia Palamidessi
- University of Stuttgart (Germany), Ralf Kuesters
- Zama (SME spin-off of CryptoExperts, France), Pascal Paillier and Matthieu Rivain
- Edgeless Systems (SME, Germany), Felix Schuster
- Aircloak (SME, Germany), Felix Bauer

Inria contact: Catuscia Palamidessi

Description: The project aims at building an open source cloud platform promoting the adoption of privacy-preserving computing (PPC) technology by offering a broad spectrum of business-ready PPC techniques (Secure Multiparty Computation, Homomorphic Encryption, Trusted Execution Environments, and methods for Statistical Disclosure Control, in particular Differential Privacy) as reusable and composable services.

8.4 National initiative

Achronym: REPAS

Title: Reliable and Privacy-Aware Software Systems via Bisimulation Metrics

Program: ANR Blanc

Duration: October 1, 2016 - September 30, 2021

Coordinator: Catuscia Palamidessi

Partners:

- Inria Saclay (EPI COMETE), Catuscia Palamidessi
- Inria Sophia Antipolis (EPI FOCUS), Ugo Dal Lago and Davide Sangiorgi
- ENS Lyon, Matteo Mio
- ENS Paris, Vincent Danos

Inria contact: Catuscia Palamidessi

Description: In this project we investigate quantitative notions and tools for proving program correctness and protecting privacy. In particular, we focus on bisimulation metrics, which are the natural extension of bisimulation on quantitative systems. As a key application, we will develop a mechanism to protect the privacy of users when their location traces are collected.

8.5 Regional initiative

Achronym: LOST2DNN

Title: Leakage of Sensitive Training Data from Deep Neural Networks

Program: DATAIA Call for Research Projects

Duration: October 1, 2019 - September 30, 2022

Coordinators: Catuscia Palamidessi and Pablo Piantanida

Partners:

- Inria, Catuscia Palamidessi
- Centrale Supélec, Pablo Piantanida
- TU Wien, Austria (Associate). Georg Pichler

Inria contact: Catuscia Palamidessi

Description: The overall project goal is to develop a fundamental understanding with experimental validation of the information-leakage of training data from deep learning systems. We plan to establish the foundations for a suitable measure of leakage which will serve as a basis for the analysis of attacks and for the development of robust mitigation techniques.

9 Dissemination

9.1 Scientific Activities

9.1.1 Executive and Steering Committees

Catuscia Palamidessi has been / is involved in the following committees:

- (2019-) Member of the Scientific Advisory Board of ANSSI, the French National Cybersecurity Agency.
- (2019-) Member of the Scientific Advisory Board of CISPA, the Helmholtz Center for Information Security.
- (2016-) Member of the Steering Committee of CONCUR, the International Conference in Concurrency Theory.
- (2015-) Member of the Steering Committee of EACSL, the European Association for Computer Science Logics.
- (2014-) Member of the Executive Committee of SIGLOG, the ACM Special Interest Group on Logic and Computation.
- (2005-20) Member of the Steering Committee of ETAPS, the European Joint Conferences on Theory and Practice of Software.
- (1997-) Member of the Steering Committee of EXPRESS, the international workshop on Expressiveness in Concurrency.

Frank Valencia has been a member of the Steering Committee of EXPRESS, the international workshop on Expressiveness in Concurrency (2011-).

9.1.2 Scientific Associations

Catuscia Palamidessi is member of the following associations:

- (2010-) Member of the IFIP Working Group 1.7 – Theoretical Foundations of Security Analysis and Design.
- (2007-20) Member of the IFIP Technical Committee 1. Foundations of Computer Science.
- (2005-) Member of the IFIP Working Group 1.8 – Concurrency Theory.

9.1.3 Editorial boards

Catuscia Palamidessi is member of the following boards:

- (2020-) Member of the Editorial Board of the IEEE Transactions on Dependable and Secure Computing. IEEE Computer Society.
- (2020-) Member of the Editorial Board of the Journal of Logical and Algebraic Methods in Programming, Elsevier.
- (2019-) Member of the Editorial Board of the Journal of Computer Security. IOS Press.
- (2017-20) Member of the Editorial Board of Proceedings on Privacy Enhancing Technologies (PoPETs), De Gruyter.
- (2015-) Member of the Editorial Board of Acta Informatica, Springer.
- (2014-) Member of the Advising Board of LIPIcs: Leibniz International Proceedings in Informatics, Schloss Dagstuhl–Leibniz Center for Informatics.
- (2006-) Member of the Editorial Board of Mathematical Structures in Computer Science, Cambridge University Press.

9.1.4 Program committee member for conferences and workshops

Catuscia Palamidessi has been / is member of the following program committees:

Conferences

- CSF 2021. The 34th IEEE Computer Security Foundations Symposium. Dubrovnik, Croatia, 21-25 June 2021.
- FORTE 2021. The 41st IFIP International Conference on Formal Techniques. La Valletta, Malta, 14-18 June 2021.
- AAAI 2021. The 35th AAAI Conference on Artificial Intelligence. Virtual conference, 2-9 February 2021.
- SECURWARE 2020. The 13th International Conference on Emerging Security Information, Systems and Technologies. Valencia, Spain. November 15-20, 2020.
- CCS 2020. The ACM Conference on Computer and Communications Security. Orlando, USA, November 9-13 2020.
- VECoS 2020. The 14th International Conference on Verification and Evaluation of Computer and Communication Systems. Xi'an, China. 26-27 October 2020.
- SEFM 2020. The 18th International Conference on Software Engineering and Formal Methods. Virtual conference, 14-17 September 2020.

- CSF 2020. The 33rd IEEE Computer Security Foundations Symposium. Boston, MA, USA, June 22-26, 2020.
- PETS 2020. The 20th Privacy Enhancing Technologies Symposium. Montréal, Canada, July 14 – 18, 2020.
- FORTE 2020. The 40th IFIP International Conference on Formal Techniques for Distributed Objects, Components, and Systems. University of Malta, Valletta, June 15-19, 2020.

Workshops

- PPAI 2021. The 2nd AAAI Workshop on Privacy-Preserving Artificial Intelligence. Online. February 8-9, 2021.
- DS3, The fourth Data Science Summer School. Online. January 4-9, 2021.
- PPML 2020. Privacy Preserving Machine Learning - PriML and PPML Joint Edition. Online. December 11th, 2020.
- DIP 2020. Recent Developments of the Design and Implementation of Programming Languages. Online. November 27th, 2020.
- HotSpot 2020. Hot Issues in Security Principles and Trust. Online. September 7th, 2020.
- EXPRESS/SOS 2020. Combined 27th International Workshop on Expressiveness in Concurrency and 17th Workshop on Structural Operational Semantics. Vienna, Austria, August 31, 2020.
- APVP 2020. 11ème édition de l'Atelier sur la Protection de la Vie Privée. Saint-Martin-de-Londres. June 24-26, 2020.
- TML 2020. Towards Trustworthy ML: Rethinking Security and Privacy for ML. Addis Ababa, Ethiopia, April 26, 2020.
- PPAI 2020. The AAAI Workshop on Privacy-Preserving Artificial Intelligence. New York, USA, February 7, 2020.

Frank Valencia has been a PC member for ICLP 2020 The International Conference on Logic Programming 2020.

9.1.5 Invited speaker at international conferences and workshops

Catuscia Palamidessi has been invited speaker at the following events:

- CONCUR 2020. The 31st International Conference on Concurrency Theory Vienna, Austria, September 1-4, 2020.
- PPAI 2020. The AAAI Workshop on Privacy-Preserving Artificial Intelligence. New York, USA, February 7, 2020.
- Federated Learning and Privacy-Preserving Machine Learning. Société Française de Statistique. Virtual workshop. 24 November 2020.
- Journées Nationales du GDR Sécurité Informatique. Paris, France, June 2019.

9.2 Teaching - Supervision - Juries

9.2.1 Supervision of PhD students

- (2020-) Sayan Biswas. IPP. Thesis subject: On the tradeoff between Local Differential Privacy and Statistical Utility.
- (2020-) Ruta Binkite-Saudaskiene. IPP. Co-supervised by Catuscia Palamidessi and Frank Valencia. Thesis subject: Fairness issues in Social Networks.
- (2020-) Carlos Pinzon. IPP. Co-supervised by Catuscia Palamidessi, Pablo Piantanida and Frank Valencia. Thesis subject: On the tradeoff between Privacy and Fairness in Machine Learning.
- (2019-) Federica Granese. IPP and Università di Roma "La Sapienza". Co-supervised by Catuscia Palamidessi, Daniele Gorla and Pablo Piantanida. Thesis subject: Security in Machine Learning.
- (2019-) Ganesh Del Grosso Guzman. IPP. Co-supervised by Catuscia Palamidessi and Pablo Piantanida. Thesis subject: Privacy in Machine Learning.
- (2018-) Santiago Quintero Pabón. IPP. Co-supervised by Frank Valencia and Catuscia Palamidessi. Thesis subject: Foundational Models for Group Polarization, Lies and Privacy in Social Networks.
- (2018-) Natasha Fernandez. IPP and University of Maquaire. Co-supervised by Catuscia Palamidessi and Annabelle McIver. Thesis subject: Privacy Protection Methods for Textual Documents.
- (2017-) Anna Pazii. Co-supervised by Catuscia palamidessi and Konstantinos Chatzikokolakis. Thesis subject: Local Differential Privacy.
- (2017-20) Marco Romanelli. IPP and Università di Siena. Co-supervised by catuscia Palamidessi and Moreno Falaschi, University of Siena, Italy. Thesis subject: Machine Learning and Private Information Flow. Marco defended his thesis in October 2020 and after the thesis defense he took a postdoc position at Centrale Supélec, France.
- (2017-2020) Sergio Ramirez. Co-supervised by Frank Valencia and Camilo Rueda, Universidad Javeriana Cali. Thesis subject: Constraint Systems to Reason About Distributed Knowledge. Thesis has been submitted and it is expected to be defended in April, 2021.

9.2.2 Supervision of Master students

COMETE has hosted as an intern Lucas Bouju, Master student at ENSAI (France), from Mar 2020 until Sep 2020.

9.2.3 Supervision of postdocs

- (Nov 2020-) Gangsoo Zeong (aka Kangsoo Jung), supervised by Catuscia palamidessi.
- (Nov 2020-) Marco Romanelli, co-supervised by Catuscia Palamidessi and Pablo Piantanida.
- (Nov 2020-) Sergio Ramirez, supervised by Frank Valencia.

9.2.4 Advisory boards for PhD programs and thesis

Catuscia Palamidessi is:

- (2012-) External member of the scientific council for the PhD in Computer Science at the University of Pisa, Italy.
- (2020-) Member of the advising committee of Abhishek Sharma, PhD student supervised by Maks Ovsjanikov, IPP, France.

9.2.5 Examination of PhD thesis

Catuscia Palamidessi has been president of the committee board at the PhD defense of Sammy Khalife (IPP, Palaiseau). Title of the thesis: *Graphs, geometry and representations for language and networks of entities*. Supervised by Michalis Vazirgiannis. Defended in June 2020.

9.2.6 Teaching

Catuscia Palamidessi has designed and been responsible for the MPRI course 2.3.2: *Foundations of privacy* on the protection of privacy and control of information leakage. The course consists of 24 hours of lectures, and is worth 3 ECTS. She has also been teaching the course during the academic years 2015-16, 2016-17, 2017-18 and 2019-20. Frank Valencia has been teaching the following courses at Pontificia Universidad Javeriana: Two masters courses, one on Foundations of Computation and the other on Concurrency Theory, and two undergraduate courses, one on Discrete Math and the other on Computability and Complexity. Each course consists 42 hours of lectures.

9.3 Popularization

9.3.1 Promotion of scientific activities

Since 2020, Catuscia Palamidessi is animating the DATAIA Expert Group on Digital Trust. The objective of this EG is to set up projects and initiatives of various kind aimed at facilitating the development of practical solutions for the application of the GDPR, with particular focus on differential privacy and homomorphic encryption.

9.4 Service

9.4.1 Service to the department and college

Catuscia Palamidessi is:

- (2020-) Vice-president of the Commission Scientifique du Centre de Recherche Inria Saclay.
- (2019-) Deputy Member of the Commission Consultatives Paritaire de l'Inria.
- (2018-20) Member of the Commission Scientifique du Centre de Recherche Inria Saclay.
- (2017-) Member of the committee for the assignment of the INRIA International Chairs.

9.4.2 Evaluation committees

Catuscia Palamidessi has been involved in the following evaluation committees:

- (2020-) Member of the EAPLS PhD Award Committee.
- (2020) Member of the KON-NT panel for natural and engineering sciences, Sweden.
- (2020) Member of the committee for a professor position at the Scuola Normale Superiore, Pisa.
- (2020) Member of the committee for a professor position at the CS department of the University of Pisa.
- (2020) Member of the evaluation commitee for Assistant professor positions at the Department of Computer Science (DIX), Ecole Polytechnique.
- (2020) External Member of the committee for the promotion to full professor of Prof. Kévin Huguenin. HEC Lausanne, Switzerland.

- (2005-) Reviewer for the project proposals for the program PRIN, sponsored by the Italian MIUR ("Ministero dell'Istruzione, dell'Università e della Ricerca").

Frank Valencia has been a member of the committee for for the 2020 full professor and assistant professor positions at the Department of Computer Science (DIX), Ecole Polytechnique.

10 Scientific production

10.1 Major publications

[1] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi and G. Smith. 'Additive and multiplicative notions of leakage, and their capacities'. In: *27th Computer Security Foundations Symposium (CSF 2014)*. Vienna, Austria: IEEE, July 2014, pp. 308–322. DOI: 10.1109/CSF.2014.29. URL: https://hal.inria.fr/hal-00989462.

[2] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi and G. Smith. 'An Axiomatization of Information Flow Measures'. In: *Theoretical Computer Science* 777 (2019), pp. 32–54. DOI: 10.1016/j.tcs.2018.10.016. URL: https://hal.archives-ouvertes.fr/hal-01995712.

[3] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, P. Degano and C. Palamidessi. 'On the information leakage of differentially-private mechanisms'. In: *Journal of Computer Security* 23.4 (2015), pp. 427–469. DOI: 10.3233/JCS-150528. URL: https://hal.inria.fr/hal-00940425.

[4] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis and C. Palamidessi. 'Geo-Indistinguishability: Differential Privacy for Location-Based Systems'. Anglais. In: *20th ACM Conference on Computer and Communications Security*. DGA, Inria large scale initiative CAPPRIS. ACM. Berlin, Allemagne: ACM Press, 2013, pp. 901–914. DOI: 10.1145/2508859.2516735. URL: http://hal.inria.fr/hal-00766821.

[5] N. E. Bordenabe, K. Chatzikokolakis and C. Palamidessi. 'Optimal Geo-Indistinguishable Mechanisms for Location Privacy'. In: *CCS - 21st ACM Conference on Computer and Communications Security*. Ed. by G.-J. Ahn, M. Yung and N. Li. Proceedings of the 21st ACM Conference on Computer and Communications Security. Gail-Joon Ahn. Scottsdale, Arizona, United States: ACM, Nov. 2014, pp. 251–262. DOI: 10.1145/2660267.2660345. URL: https://hal.inria.fr/hal-00950479.

[6] G. Cherubin, K. Chatzikokolakis and C. Palamidessi. 'F-BLEAU: Fast Black-Box Leakage Estimation'. In: *Proceedings of the 40th IEEE Symposium on Security and Privacy (SP)*. San Francisco, United States: IEEE, May 2019, pp. 835–852. DOI: 10.1109/SP.2019.00073. URL: https://hal.archives-ouvertes.fr/hal-02422945.

[7] M. Guzmán, S. Haar, S. Perchy, C. Rueda and F. D. Valencia. 'Belief, Knowledge, Lies and Other Utterances in an Algebra for Space and Extrusion'. In: *Journal of Logical and Algebraic Methods in Programming* (Sept. 2016). DOI: 10.1016/j.jlamp.2016.09.001. URL: https://hal.inria.fr/hal-01257113.

[8] M. Guzmán, S. Knight, S. Quintero, S. Ramírez, C. Rueda and F. D. Valencia. 'Reasoning about Distributed Knowledge of Groups with Infinitely Many Agents'. In: *CONCUR 2019 - 30th International Conference on Concurrency Theory*. Ed. by W. Fokkink and R. van Glabbeek. Vol. 140. Amsterdam, Netherlands, Aug. 2019, 29:1–29:15. DOI: 10.4230/LIPIcs.CONCUR.2019.29. URL: https://hal.archives-ouvertes.fr/hal-02172415.

[9] S. Knight, C. Palamidessi, P. Panangaden and F. D. Valencia. 'Spatial and Epistemic Modalities in Constraint-Based Process Calculi'. In: *CONCUR 2012 - Concurrency Theory - 23rd International Conference, CONCUR 2012*. Vol. 7454. Newcastle upon Tyne, United Kingdom, Sept. 2012, pp. 317–332. DOI: 10.1007/978-3-642-32940-1. URL: http://hal.inria.fr/hal-00761116.

[10] M. Romanelli, K. Chatzikokolakis, C. Palamidessi and P. Piantanida. 'Estimating g-Leakage via Machine Learning'. In: *CCS '20 - 2020 ACM SIGSAC Conference on Computer and Communications Security*. This is the extended version of the paper which appeared in the Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS), November 9-13, 2020, Virtual Event, USA. Online, United States: ACM, Nov. 2020, pp. 697–716. URL: https://hal.archives-ouvertes.fr/hal-03091469.

10.2 Publications of the year

International journals

[11] V. Castiglioni, K. Chatzikokolakis and C. Palamidessi. 'A Logical Characterization of Differential Privacy'. In: *Science of Computer Programming* 188 (2020), p. 102388. DOI: 10.1016/j.scico.2019.102388. URL: https://hal.archives-ouvertes.fr/hal-02423048.

[12] K. Chatzikokolakis, N. Fernandes and C. Palamidessi. 'Refinement Orders for Quantitative Information Flow and Differential Privacy'. In: *Journal of Cybersecurity and Privacy* 1 (12th Dec. 2020), pp. 40–77. DOI: 10.3390/jcp1010004. URL: https://hal.inria.fr/hal-03091754.

[13] M. Falaschi, M. Gabbrielli, C. Olarte and C. Palamidessi. 'Dynamic slicing for Concurrent Constraint Languages'. In: *Fundamenta Informaticae* 177.3-4 (10th Dec. 2020), pp. 331–357. DOI: 10.3233/FI-2020-1992. URL: https://hal.archives-ouvertes.fr/hal-02423973.

[14] D. Gorla, F. Granese and C. Palamidessi. 'Enhanced Models for Privacy and Utility in Continuous-Time Diffusion Networks'. In: *International Journal of Information Security* (2021). DOI: 10.1007/s10207-020-00530-7. URL: https://hal.inria.fr/hal-03094843.

[15] M. Guzmán, S. Knight, S. Quintero, C. Rueda and F. Valencia. 'Algebraic Structures from Concurrent Constraint Programming Calculi for Distributed Information in Multi-Agent Systems'. In: *Journal of Logical and Algebraic Methods in Programming* (2021). URL: https://hal.archives-ouvertes.fr/hal-03098441.

International peer-reviewed conferences

[16] E. Elsalamouny and C. Palamidessi. 'Full Convergence of the Iterative Bayesian Update and Applications to Mechanisms for Privacy Protection'. In: EuroS&P 2020 - 5th IEEE European Symposium on Security and Privacy. Genova, Italy, 7th Sept. 2020, pp. 490–507. URL: https://hal.inria.fr/hal-03091504.

[17] C. Palamidessi and M. Romanelli. 'Modern Applications of Game-Theoretic Principles'. In: CONCUR 2020 - 31st International Conference on Concurrency Theory. Vol. 171. Leibniz International Proceedings in Informatics (LIPIcs). Vienne / Virtual, Austria, 1st Sept. 2020, 4:1–4:9. DOI: 10.4230/LIPIcs.CONCUR.2020.4. URL: https://hal.inria.fr/hal-03091743.

[18] S. Quintero, S. Ramírez, C. Rueda and F. D. Valencia. 'Counting and Computing Join-Endomorphisms in Lattices'. In: RAMICS 2020 - 18th International Conference on Relational and Algebraic Methods in Computer Science. Lecture Notes in Computer Science. Paris, France, 1st Apr. 2020. URL: https://hal.archives-ouvertes.fr/hal-02422624.

[19] M. Romanelli, K. Chatzikokolakis and C. Palamidessi. 'Optimal Obfuscation Mechanisms via Machine Learning'. In: CSF 2020 - 33rd IEEE Computer Security Foundations Symposium. Online, United States, 22nd June 2020, pp. 153–168. URL: https://hal.inria.fr/hal-03091514.

[20] M. Romanelli, K. Chatzikokolakis, C. Palamidessi and P. Piantanida. 'Estimating g-Leakage via Machine Learning'. In: CCS '20 - 2020 ACM SIGSAC Conference on Computer and Communications Security. Online, United States, 9th Nov. 2020, pp. 697–716. URL: https://hal.archives-ouvertes.fr/hal-03091469.

Scientific books

[21] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi and G. Smith. *The Science of Quantitative Information Flow*. 2020, pp. XXVIII, 478. DOI: 10.1007/978-3-319-96131-6. URL: https://hal.inria.fr/hal-01971490.

Scientific book chapters

[22] M. Falaschi, C. Palamidessi and M. Romanelli. 'Derivation of Constraints from Machine Learning Models and Applications to Security and Privacy'. In: *Recent Developments in the Design and Implementation of Programming Languages*. Vol. 86. OASICS. 2020, 11:1–11:20. DOI: 10.4230/OASIcs.Gabbrielli.2020.11. URL: https://hal.archives-ouvertes.fr/hal-03091740.

Reports & preprints

[23] M. S. Alvim, B. Amorim, S. Knight, S. Quintero and F. Valencia. *Polarization and Belief Convergence of Agents in Strongly-Connected Influence Graphs*. 1st Aug. 2020. URL: https://hal.archives-ouvertes.fr/hal-03095987.

[24] M. S. Alvim, K. Chatzikokolakis, Y. Kawamoto and C. Palamidessi. *Information Leakage Games: Exploring Information as a Utility Function*. 31st Dec. 2020. URL: https://hal.archives-ouvertes.fr/hal-03091413.

[25] K. Chatzikokolakis, G. Cherubin, C. Palamidessi and C. Troncoso. *The Bayes Security Measure*. 31st Dec. 2020. URL: https://hal.inria.fr/hal-03091416.

[26] S. Knight, P. Panangaden and F. Valencia. *Computing with Epistemic and Spatial Modalities*. 12th Feb. 2020. URL: https://hal.archives-ouvertes.fr/hal-03148149.

[27] K. Makhlouf, S. Zhioua and C. Palamidessi. *On the Applicability of ML Fairness Notions*. 31st Dec. 2020. URL: https://hal.archives-ouvertes.fr/hal-03091436.

[28] K. Makhlouf, S. Zhioua and C. Palamidessi. *Survey on Causal-based Machine Learning Fairness Notions*. 31st Dec. 2020. URL: https://hal.archives-ouvertes.fr/hal-03091428.

10.3 Cited publications

[29] M. S. Alvim, K. Chatzikokolakis, Y. Kawamoto and C. Palamidessi. *Information Leakage Games: Exploring Information as a Utility Function*. 2020. arXiv: 2012.12060 [cs.CR].

[30] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi and G. Smith. 'Measuring Information Leakage Using Generalized Gain Functions'. In: *Proceedings of the 25th IEEE Computer Security Foundations Symposium (CSF)*. 2012, pp. 265–279. DOI: http://doi.ieeecomputersociety.org/10.1109/CSF.2012.26. URL: http://hal.inria.fr/hal-00734044/en.

[31] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe and C. Palamidessi. 'Broadening the scope of Differential Privacy using metrics'. In: *Proceedings of the 13th International Symposium on Privacy Enhancing Technologies (PETS 2013)*. Ed. by E. De Cristofaro and M. Wright. Vol. 7981. Lecture Notes in Computer Science. Springer, 2013, pp. 82–102.

[32] R. Cummings, V. Gupta, D. Kimpara and J. Morgenstern. 'On the Compatibility of Privacy and Fairness'. In: *Proceedings of the 27th Conference on User Modeling, Adaptation and Personalization*. UMAP'19 Adjunct. Larnaca, Cyprus: Association for Computing Machinery, 2019, pp. 309–315. DOI: 10.1145/3314183.3323847. URL: https://doi.org/10.1145/3314183.3323847.

[33] M. D. Ekstrand, R. Joshaghani and H. Mehrpouyan. 'Privacy for All: Ensuring Fair and Equitable Privacy Protections'. In: *Proceedings of the First ACM Conference on Fairness, Accountability and Transparency (FAT)*. Ed. by S. A. Friedler and C. Wilson. Vol. 81. Proceedings of Machine Learning Research. PMLR, 2018, pp. 35–47. URL: http://proceedings.mlr.press/v81/ekstrand18a.html.

[34] J.-M. Esteban and D. Ray. 'On the Measurement of Polarization'. In: *Econometrica* 62.4 (1994), pp. 819–851. URL: http://www.jstor.org/stable/2951734.

[35] J. Jia, A. Salem, M. Backes, Y. Zhang and N. Z. Gong. 'MemGuard: Defending against Black-Box Membership Inference Attacks via Adversarial Examples'. In: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*. CCS '19. London, United Kingdom: Association for Computing Machinery, 2019, pp. 259–274. DOI: 10.1145/3319535.3363201. URL: https://doi.org/10.1145/3319535.3363201.

[36] L. Song, R. Shokri and P. Mittal. 'Privacy Risks of Securing Machine Learning Models against Adversarial Examples'. In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019*. Ed. by L. Cavallaro, J. Kinder, X. Wang and J. Katz. ACM, 2019, pp. 241–257. DOI: 10.1145/3319535.3354211. URL: https://doi.org/10.1145/3319535.3354211.

[37] M. C. Tschantz, S. Sen and A. Datta. 'SoK: Differential Privacy as a Causal Property'. In: *2020 IEEE Symposium on Security and Privacy, SP 2020, San Francisco, CA, USA, May 18-21, 2020*. IEEE, 2020, pp. 354–371. DOI: 10.1109/SP40000.2020.00012. URL: https://doi.org/10.1109/SP40000.2020.00012.