RESEARCH CENTRE

Saclay - Île-de-France

IN PARTNERSHIP WITH:

CNRS, Ecole Polytechnique

2021 ACTIVITY REPORT

Project-Team

COMETE

Privacy, Fairness and Robustness in Information Management

IN COLLABORATION WITH: Laboratoire d'informatique de l'école polytechnique (LIX)

DOMAIN

Algorithmics, Programming, Software and Architecture

THEME

Security and Confidentiality

Contents

Project-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]
- Sami Zhioua [Inria, Advanced Research Position, from Sep 2021]

Post-Doctoral Fellows

- Hafiz Asif [Inria, Feb 2021]
- Hamid Jalalzai [Inria, from Feb 2021]
- Sergio Ramirez Rico [Inria]
- Marco Romanelli [Centrale-Supélec]
- Gangsoo Zeong [Inria]

PhD Students

- Ruta Binkyte-Sadauskiene [Inria]
- Sayan Biswas [Inria]
- Ganesh Del Grosso Guzman [Inria]
- Natasha Fernandes [Université Macquarie Sydney - Australie, until Jun 2021]
- Federica Granese [Inria]
- Karima Makhlouf [Inria, from Sep 2021]
- Carlos Pinzon [Inria]
- Santiago Quintero [École polytechnique]

Administrative Assistant

- Maria Agustina Ronco [Inria]

Visiting Scientists

- Filippo Galli [ENS Pise, from Sep 2021]
- Hamid Jalalzai [Univ Paris-Saclay, Jan 2021]

External Collaborators

- Konstantinos Chatzikokolakis [CNRS]
- Juan Pablo Piantanida [Centrale-Supélec]

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 [28]. 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:

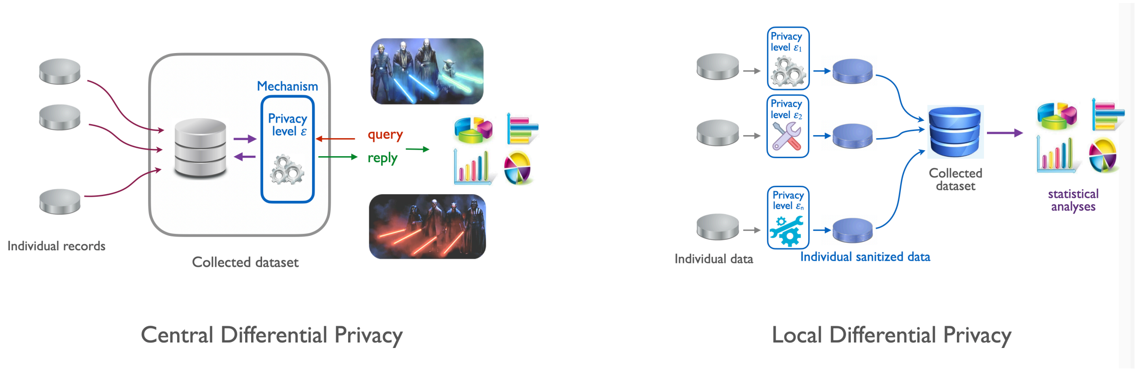
Mechanism
Privacy level ε
query
reply
Individual records
Collected dataset
Central Differential Privacy
Privacy level ε_1
Privacy level ε_2
Privacy level ε_n
Collected dataset
statistical analyses
Individual data
Individual sanitized data
Local Differential Privacy

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

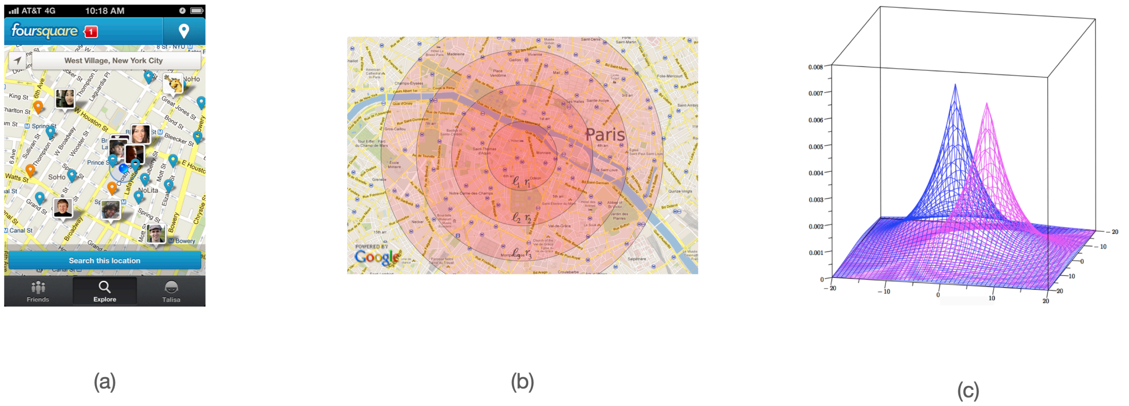

Figure 2: Geo-indistinguishability is a framework to protect the privacy of the user when dealing with location-based services (a). The framework guarrantees d-privacy, a distance-based variant of differential privacy (b). The typical implementation uses (extended) Laplace noise (c).

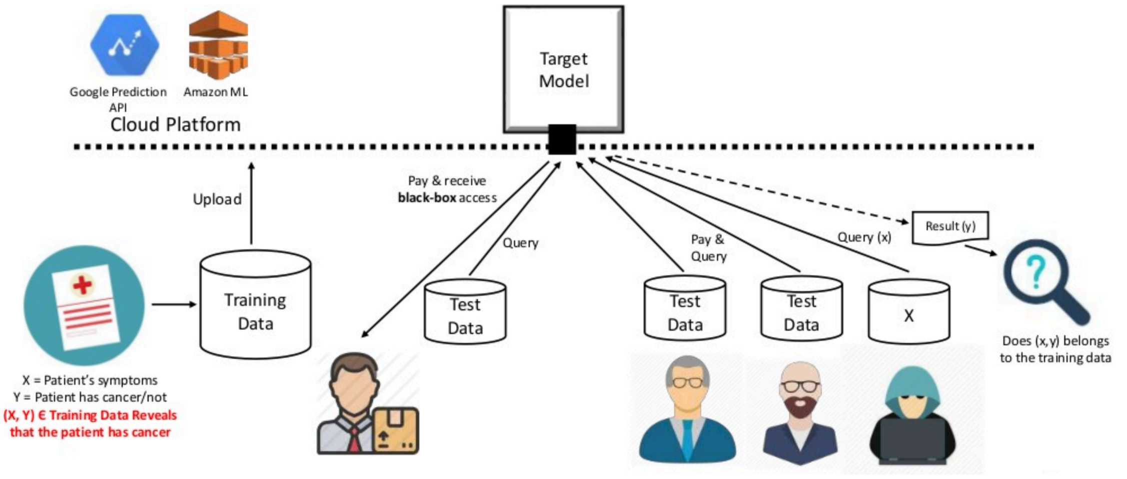

Figure 3: Privacy breach in machine learning as a service.

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. [32]). On the other hand, there are studies that show that there is a trade-off between protection from inference attacks and robustness [35]. 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* [30] and of *Equality of False Positives* and *Equality of False Negatives* [29]. 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* [36].

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 [27] 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* [26] 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 [34]. 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 [33]. 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 [11] 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 [31]. 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 Awards

- Natasha Fernandez, who was a PhD student in COMETE under the joint supervision of Catuscia Palamidessi and Annabelle McIver (University of Macquarie, Australia), has obtained the prize CORE ("John Makepeace Bennett Award") for the best PhD thesis in Australasia for the year.

- Marco Romanelli , who was a PhD student in COMETE under the joint supervision of Catuscia Palamidessi and Moreno Falaschi (University of Siena, Italy), has obtained one of the prizes of the Université Franco-Italienne (Prix de thèse en cotutelle UFI 2021).

- The paper "DOCTOR: A Simple Method for Detecting Misclassification Errors" by Federica Granese, Daniele Gorla, Catuscia Palamidessi, Pablo Piantanida and Marco Romanelli, was accepted for a presentation in "spotlight" (less than 3% of the accepted papers) at NeurIPS 2021.

5.2 New Projects

- The project "Computational Models for Social Networks Applied to Polarization" led by Frank Valencia was accepted in 2021. The project will be funded by the Ministry of Science (MinCiencias Colombia) for 1/2 million euros for 8 undergrad, 2 PhD, 1 masters students and internships and trips to Inria / LIX for the researchers and the students involved in the project. It is a multi-disciplinary project that includes economists and political scientists.

- Catuscia Palamidessi is PI in a PEPR Cybersecurity project on privacy led by Vincent Roca (PRIVATICS).
- Catuscia Palamidessi is PI in an Inria Challenge "Federated Machine Learning over the Internet", led by Aurelien Bellet (MAGNET) and Giovanni Neglia (NEO).
- Catuscia Palamidessi is PI in the HORIZON-RIA proposal "European Lighthouse for Safe and Secure AI", led by Mario Fritz (CISPA, Saarbruken).

5.3 Invitations

- Catuscia Palamidessi has been invited to be a keynote speaker conference cluster FLOC 2022, The Federated Logic Conference. FLOC brings together several top international conferences (A* and A) related to mathematical logic and computer science, including LICS, IJCAR, CAV and CSF. FLOC 2022 Invited Speakers.
- Frank Valencia gave a keynote presentation on Polarization in Social Networks and debated in a Live Forum "Redes sociales: construcción de realidades, polarizacojection, y oportunidades para la región y el país" (attendance: more than 1000 participants) about the 2021 civil unrest in Colombia. Talk by Frank Valencia on Polarization.

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

Contact: Konstantinos Chatzikokolakis

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

Contact: 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

Contact: 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

Participants: Catuscia Palamidessi, Frank Valencia, Sami Zhioua, Gangsoo Zeong, Sergio Ramírez, Marco Romanelli, Sayan Biswas, Federica Granese, Santiago Quintero, Karima Makhlouf, Natasha Fernandes.

7.1 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 [11], we 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 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 showed 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 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 showed how to adapt the basic techniques to these enhanced policies. We concluded by giving a new approach to the maximization/minimization problem by finding a trade-off between the risk and the gain function through biobjective optimization.

7.2 Fairness Notions for Machine Learning: Bridging the Gap with Real-World Applications

Fairness has emerged as an important requirement to guarantee that Machine Learning (ML) predictive systems do not discriminate against specific individuals or entire sub-populations, in particular, minorities. Given the inherent subjectivity of viewing the concept of fairness, several notions of fairness have been introduced in the literature.

In [13, 14] we surveyed and illustrated the subtleties between fairness notions through a large number of examples and scenarios. In addition, unlike other surveys in the literature, we addressed the question “which notion of fairness is most suited to a given real-world scenario and why?”. Our attempt to answer this question consisted in (1) identifying the set of fairness-related characteristics of the real-world scenario at hand, (2) analyzing the behavior of each fairness notion, and then (3) fitting these two elements to recommend the most suitable fairness notion in every specific setup. The results were summarized in a decision diagram that can be used by practitioners and policy makers to navigate the relatively large catalog of ML fairness notions.

7.3 DOCTOR: A Simple Method for Detecting Misclassification Errors

Deep neural networks (DNNs) have shown to perform very well on large scale object recognition problems and lead to widespread use for real-world applications, including situations where DNN are implemented as “black boxes”. A promising approach to secure their use is to accept decisions that are likely to be correct while discarding the others.

In [19], we proposed DOCTOR, a simple method that aims to identify whether the prediction of a DNN classifier should (or should not) be trusted so that, consequently, it would be possible to accept it or to reject it. Two scenarios were investigated: Totally Black Box (TBB) where only the soft-predictions are available and Partially Black Box (PBB) where gradient-propagation to perform input pre-processing is allowed. Empirically, we showed that DOCTOR outperforms all state-of-the-art methods on various well-known images and sentiment analysis datasets. In particular, we observed a reduction of up to 4rate (FRR) in the PBB scenario. DOCTOR can be applied to any pre-trained model, it does not require prior information about the underlying dataset and is as simple as the simplest available methods in the literature.

7.4 Locality Sensitive Hashing with Extended Differential Privacy

Extended differential privacy (aka d-privacy), a generalization of standard differential privacy (DP) using a general metric, has been widely studied to provide rigorous privacy guarantees while keeping high utility. However, existing works on extended DP are limited to few metrics, such as the Euclidean metric. Consequently, they have only a small number of applications, such as location-based services and document processing.

In [18], we proposed a couple of mechanisms providing extended DP with a different metric: angular distance (or cosine distance). Our mechanisms are based on locality sensitive hashing (LSH), which can be applied to the angular distance and work well for personal data in a high-dimensional space. We theoretically analyzed the privacy properties of our mechanisms, and proved extended DP for input data by taking into account that LSH preserves the original metric only approximately. We applied our mechanisms to friend matching based on high-dimensional personal data with angular distance in the local model, and evaluated our mechanisms using two real datasets. We showed that LDP requires a very large privacy budget and that RAPPOR does not work in this application. Then we showed that our mechanisms enable friend matching with high utility and rigorous privacy guarantees based on extended DP.

7.5 An Incentive Mechanism for Trading Personal Data in Data Markets

Personal data is becoming one of the most essential resources in today's information-based society. Accordingly, there is a growing interest in data markets, which operate data trading services between data providers and data consumers. One issue the data markets have to address is that of the potential threats to privacy. Usually some kind of protection must be provided, which generally comes to the detriment of utility. A correct pricing mechanism for private data should therefore depend on the level of privacy.

In [17], we introduced a pricing mechanism that takes into account the trade-off between privacy and accuracy. We proposed a method to induce the data provider to accurately report her privacy price and, we optimized it in order to maximize the data consumer's profit within budget constraints. We showed formally that the proposed mechanism achieves these properties, and also, validated them experimentally.

In [23], we proposed a model of data federation in which data providers, who are, generally, less influential on the market than data consumers, form a coalition for trading their data, simultaneously shielding against privacy threats by means of differential privacy. Additionally, we proposed a technique to price private data, and an revenue-distribution mechanism to distribute the revenue fairly in such federation data trading environments. Our model also motivates the data providers to cooperate with their respective federations, facilitating a fair and swift private data trading process. We validated our result through various experiments, showing that the proposed methods provide benefits to both data providers and consumers.

7.6 A Multi-Agent Model for Polarization under Confirmation Bias in Social Networks

In social networks a group may shape their beliefs by attributing more value to the opinions of outside 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 preexisting beliefs. This is 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*

In [15] we described a model for polarization in multi-agent systems based on Esteban and Ray's standard measure of polarization from economics. In our model agents evolve by updating their beliefs (opinions) based on an underlying influence graph, as in the standard DeGroot model for social learning, but under a *confirmation bias*; i.e., a discounting of opinions of agents with dissimilar views. We showed that even under this bias polarization eventually vanishes (converges to zero) if the influence graph is strongly-connected. If the influence graph is a regular symmetric circulation, we determined the unique belief value to which all agents converge. Our more insightful result establishes that, under some natural assumptions, if polarization does not eventually vanish then either there is a disconnected subgroup of agents, or some agent influences others more than she is influenced. We also showed that polarization does not necessarily vanish in weakly-connected graphs under confirmation bias. Finally, we illustrated our model with a series of case studies and simulations, and showed how it relates to the classic DeGroot model for social learning.

7.7 Reasoning about distributed information with infinitely many agents

Distributed information of a group of agents is the information that results from combining the individual information of the members of the group. Spatial constraint systems (scs) are semantic structures for reasoning about spatial and epistemic information in concurrent systems.

In [12] we developed a theory of scs to reason about the *distributed information* of potentially *infinite groups*. We first characterized the notion of distributed information of a group of agents as the infimum of the set of join-preserving functions that represent the spaces of the agents in the group. We then provided an alternative characterization of this notion as the greatest family of join-preserving functions that satisfy certain basic properties. For completely distributive lattices, we established that the distributed information of c amongst a group is the greatest lower bound of all possible combinations of information in the spaces of the agents in the group that derive c. We also showed compositionality results for these characterizations and conditions under which information that can be obtained by an infinite group can also be obtained by a finite group. Finally, we provided an application on mathematical morphology where dilations, one of its fundamental operations, induce an scs on a powerset lattice. In particular, we showed that distributed information represents a dilation in such scs.

7.8 Computing Distributed Knowledge as the Greatest Lower Bound of Knowledge

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 [22] we characterized the standard notion of distributed knowledge of a group as the greatest lower bound of the join-endomorphisms representing the knowledge of each member of the group. We also studied the problem of computing the greatest lower bound of two given join-endomorphisms over lattice orders. We proved that this problem can be solved, in the worst-case, in linear time for distributed lattices. The complexity is expressed in terms of the size of the lattice and the basic binary lattice operations performed by the algorithm. Furthermore, we showed that deciding whether an agent has the distributed knowledge of two other agents can be computed in quadratic time in the size of the underlying set of states. For the special case of S5 knowledge, we show that it can be decided in pseudo-linear time.

8 Partnerships and cooperations

Participants: Catuscia Palamidessi, Frank Valencia.

8.1 International initiatives

8.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

LOGIS

Title: Logical and Formal Methods for Information Security and privacy

Duration: Jan 1, 2019 – Dec 31, 2022

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 Participation in other International Programs

FACTS

Title: Foundational Approach to Cognition in Today's Society.

Program: ECOS NORD

Partner Institution(s):

- Inria (France)
- LIP6, Sorbonne University (France)
- Universidad Javeriana de Cali (Colombia)

Date/Duration: Jan 1 2019 - Dec 31, 2022.

Description: This project 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.

PROMUEVA

Title: A Polarization Unified Model for Colombia.

Program: Research Grant from the Ministery of Science and Technology, Colombia.

Partner Institution(s):

- LIX, Ecole Polytechnique (France)
- Universidad del Valle (Colombia)
- Universidad Javeriana de Cali (Colombia)

Date/Duration: 2022-2026 (exact start date to be defined).

Description: This project aims developing models for polarization in social networks applied to the Colombian social conflict.

Note: The project proposal successfully passed the evaluation of the scientific committee. It is expected to be finally approved in 2022 by the administration committee.

8.2 International research visitors

8.2.1 Visits of international scientists

Daniele Gorla

Status: Associate Professor

Institution of origin: Università di Roma “La Sapienza”

Country: Italy

Dates: Sept 8 – Sept 11, 2021

Context of the visit: Daniele Gorla is collaborating with Catuscia Palamidessi, Pablo Piantanida, and their common PhD student Federica Granese. The visit was aimed at defined the lines of research in the context of the PhD thesis of Federica

Mobility program/type of mobility: Research stay

Filippo Galli

Status: PhD student

Institution of origin: Scuola Normale Superiore di Pisa

Country: Italy

Dates: Sept 2021 – March 2022

Context of the visit: Filippo Galli visited in order to collaborate with Catuscia Palamidessi on the topic of d-privacy for privacy protection in machine learning

Mobility program/type of mobility: Internship

8.3 European initiatives

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

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.

8.3.1 Other european programs/initiatives

CRYPTECS

Title: Cloud-Ready Privacy-Preserving Technologies

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

Duration: June 1, 2021 - May 31, 2024

Coordinator: 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

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 initiatives

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 initiatives

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

Participants: Catuscia Palamidessi, Frank Valencia.

9.1 Promoting scientific activities

9.1.1 Scientific events: organisation

Catuscia Palamidessi is member of the steering committee of the following associations:

- (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.

9.1.2 Scientific events: selection

- Catuscia Palamidessi is / has been a member of the Program Committee of the following conferences and workshops:

  Conferences
  - Senior PC member of PETS 2022. The 22nd Privacy Enhancing Technologies Symposium. Sydney, Australia. July 11–15, 2022.
  - FORTE 2022. The 42nd International Conference on Formal Techniques for Distributed Objects, Components, and Systems. Lucca, Italy, June 13-17, 2022
  - EuroS&P 2022. The 7th IEEE European Symposium on Security and Privacy. Genoa, Italy. June 6-10, 2022.
  - CSF 2022. The 35th IEEE Computer Security Foundations Symposium. Co-located with FLoC 2022. Haifa, Israel, August 2022.
  - SEFM 2021. The 19th International Conference on Software Engineering and Formal Methods. Virtual conference, 6-10 December 2021.
  - VECoS 2021. The 15th International Conference on Verification and Evaluation of Computer and Communication Systems. Beijing, China. 22-23 November 2021.
  - ICTAC 2021. The 18th International Colloquium on Theoretical Aspects of Computing. Nur-Sultan, Kazakhstan, September 6-10, 2021
  - 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.

  Workshops
  - PPAI 2022. The 3rd AAAI Workshop on Privacy-Preserving Artificial Intelligence. Online. February 28, 2022.
  - CIRM Logic and Interaction thematic month: Logical Foundations of Probabilistic Programming. Lineal Logic International Research Network. Luminy, France. Spring 2022.
  - WPES 2021. 20th Workshop on Privacy in the Electronic Society. Virtual. November 15, 2021.
  - TPDP 2021. Theory and Practice of Differential Privacy. Workshop affiliated with ICML 2021. Virtual. 23 July 2021.
  - PPML 2021. Privacy Preserving Machine Learning - Virtual ACM CCS Workshop November 19, 2021.
  - 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.

- Frank Valencia has been a member of the Program Committee of ICLP DC 2021, the Doctoral Consortium of the 37th International Conference on Logic Programming 2021 and CLEI 2021, the 47th The Latin American Computing Conference.

9.1.3 Journal

Catuscia Palamidessi is member of the editorial board of the following journals:

- (2021-) Member of the Editorial Board of Proceedings on Privacy Enhancing Technologies (PoPETs), De Gruyter.

- (2021-) Member of the Editorial Board of TheoretiCS, a diamond Open Access journal published by Episciences .

- (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.
- (2015-) Member of the Editorial Board of Acta Informatica, Springer.
- (2014-) Member of the Editorial 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 Invited talks

- Catuscia Palamidessi is / has been invited speaker at the following events:
  - FLOC 2022. Keynote speaker at the Federated Logic Conference. Haifa, Israel, July-August 2022.
  - Talk on Differential Privacy at the Collège de France. 24 March 2022.
  - Quantitative Information Flow. Virtual round table organized by the Florida International University. September 2021.
- Frank Valencia was an invited Key speaker and Panelist at the live event *construcción de realidades, polarización, y oportunidades para la región y el país* about Polarization in Social Networks and its role in the 2021 protests in Colombia.

9.1.5 Leadership within the scientific community

Catuscia Palamidessi is a member of the following boards:

- (2021-) Member of the Board of Trustees of the IMDEA Software Institute.
- (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.

9.1.6 Scientific expertise

Catuscia Palamidessi is / has been in the following evaluation committees

- (2021) Panel member for the evaluation of projects submitted to the call HORIZON-CL3-2021-CS-01 “Increased Cybersecurity.
- (2021) Evaluator of projects for the calls PRIN 2020 and FARE 2020 of the Italian Ministry of University and Research.
- (2021) Member of the jury for the CONCUR Test-of-Time Award.
- (2021) Vice-president of the jury for the CNIL-Inria Privacy Award.
- (2021-) President of the Commission Scientifique of INRIA Saclay.
- (2021) Member of the CORE conference ranking committee for the area "Cybersecurity and Privacy" in the 2021 round of CORE conference rankings.

- (2021) Evaluator for the applications to the Austrian Science Fund (FWF) Special Research Programmes.
- (2021) Evaluator for the applications to the AXA Chair program: Foundations of Experimental CyberSecurity and Digital Risks.
- (2020-21) Vice-president of the Commission Scientifique of INRIA Saclay.
- (2020-) Member of the EAPLS PhD Award Committee.

9.2 Teaching - Supervision - Juries

9.2.1 Teaching

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.

9.2.2 Supervision

Supervision of PhD students:

- (2021-) Karima Makhlouf. IPP. Supervised by Catuscia Palamidessi. Thesis subject: Combination of Fairness and Privacy in training data.
- (2020-) Ruta Binkite-Saudaskiene. IPP. Co-supervised by Catuscia Palamidessi and Frank Valencia. Thesis subject: Fairness and Privacy in machine learning: interdisciplinary approach.
- (2020-) Sayan Biswas. IPP. Supervised by Catuscia Palamidessi. Thesis subject: On the tradeoff between Local Differential Privacy and Statistical Utility.
- (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.
- (2018-21) Santiago Quintero Pabón. IPP. Co-supervised by Catuscia palamidessi and Frank Valencia. Thesis subject: Algebraic Structures for Distributed Information and Polarization in Multi-Agent Systems. (The thesis was submitted in December 2021 and successfully defended in January 2022.)
- (2018-21) Natasha Fernandez. IPP and University of Maquaire. Co-supervised by Catuscia Palamidessi and Annabelle McIver. Thesis subject: Privacy Protection Methods for Textual Documents. (The thesis was submitted in March 2021 and successfully defended in April 2022.)

Supervision of Postdocs:

- (Jan 2021-) Hamid Jalalzai. Supervised by Catuscia Palamidessi.
- (Nov 2020-Jan 2022) Sergio Ramirez, supervised by Frank Valencia.
- (Nov 2020-) Gangsoo Zeong (aka Kangsoo Jung). Supervised by Catuscia Palamidessi.

Supervision of interns:

- Filippo Galli, PhD student, ENS Pisa, Italy. From Sept 2021 until March 2022. Supervised by Catuscia Palamidessi.

9.2.3 Juries

Catuscia Palamidessi has been member of the jury for the following thesis defenses:

- Habilitation thesis
  - Aurélien Bellet (Inria Lille, France). Title of the thesis: *Contributions to Decentralized and Privacy-Preserving Machine Learning*. Defended on November 30, 2021.
  - David Baelde (LSV, ENS Saclay, France). Tite of the thesis: *Contributions to the Verification of Cryptographic Protocols*. Defended on February 10, 2021.
- PhD thesis
  - Changmin Wu (IPP, palaiseau). Member of the committee board at the PhD defense. Title of the thesis: *Graph Representation Learning: from Kernels to Neural Networks*. Advised by Michalis Vazirgiannis. Defended in November 2021.
  - Itsaka Rakotonirina (LORIA, Nancy). Member of the committee board at the PhD defense. Title of the thesis: *Efficient verification of observational equivalences of cryptographic processes: theory and practice*. Co-advised by Steve Kremer and Vincent Cheval. Defended in February 2021.

Catuscia Palamidessi is member of advisory boards for PhD programs and thesis:

- (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.

9.3 Popularization

9.3.1 Internal or external Inria responsibilities

Catuscia Palamidessi is:

- (2021-) President of the Commission Scientifique of INRIA Saclay.
- (2019-23) Deputy Member of the Commission Consultatives Paritaire de l'Inria.
- (2017-) Member of the committee for the assignment of the INRIA International Chairs.

9.3.2 Interventions

Catuscia Palamidessi has been:

- Invited speaker at the First Webinar of the Hi! PARIS: Center on AI and Data Analytics for Science, Business and Society. The topic of this webinar is AI Bias and Privacy issues. 15 January 2021.
- Ambassador for The Logic Day and speaker at the round table La logica: pensiero, scienza e società. 14 January 2021.

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] 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.

[12] M. Guzmán, S. Knight, S. Quintero, S. Ramírez, 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.

[13] K. Makhlouf, S. Zhioua and C. Palamidessi. 'Machine learning fairness notions: Bridging the gap with real-world applications'. In: *Information processing & management* 58.5 (2021). DOI: 10.1016/j.ipm.2021.102642. URL: https://hal.archives-ouvertes.fr/hal-03624025.

[14] K. Makhlouf, S. Zhioua and C. Palamidessi. 'On the Applicability of ML Fairness Notions'. In: *SIGKDD Explorations Newsletter* 23.1 (2021), pp. 14–23. DOI: 10.1145/3468507.3468511. URL: https://hal.archives-ouvertes.fr/hal-03091436.

International peer-reviewed conferences

[15] M. S. Alvim, B. Amorim, S. Knight, S. Quintero and F. Valencia. 'A Multi-agent Model for Polarization Under Confirmation Bias in Social Networks'. In: FORTE 2021 - 41st International Conference on Formal Techniques for Distributed Objects, Components, and Systems. Valletta, Malta, 14th June 2021. URL: https://hal.archives-ouvertes.fr/hal-03095987.

[16] N. Bertrand, L. de Alfaro, R. J. Van Glabbeek, C. Palamidessi and N. Yoshida. 'CONCUR Test-Of-Time Award 2021'. In: Concur 2021 - International Conference on Concurrency Theory. Paris, France, 23rd Aug. 2021, pp. 1–3. URL: https://hal.inria.fr/hal-03480255.

[17] S. Biswas, K. Jung and C. Palamidessi. 'An Incentive Mechanism for Trading Personal Data in Data Markets'. In: International Colloquium on Theoretical Aspects of Computing 2021. Vol. 12819. Lecture Notes in Computer Science. Nur-Sultan, Kazakhstan: Springer International Publishing, 20th Aug. 2021, pp. 197–213. DOI: 10.1007/978-3-030-85315-0_12. URL: https://hal.inria.fr/hal-03589835.

[18] N. Fernandes, Y. Kawamoto and T. Murakami. 'Locality Sensitive Hashing with Extended Differential Privacy'. In: ESORICS 2021 - 26th European Symposium on Research in Computer Security. Vol. 12973. Lecture Notes in Computer Science. Darmstadt / Virtual, Germany: Springer International Publishing, 4th Oct. 2021, pp. 563–583. DOI: 10.1007/978-3-030-88428-4_28. URL: https://hal.archives-ouvertes.fr/hal-03319774.

[19] F. Granese, M. Romanelli, D. Gorla, C. Palamidessi and P. Piantanida. 'DOCTOR: A Simple Method for Detecting Misclassification Errors'. In: Advances in Neural Information Processing Systems (NeurIPS). Proceedings. Virtual event, United States, 2021. URL: https://hal.archives-ouvertes.fr/hal-03624023.

[20] A. Kumar Mishra, A. Carneiro Viana, N. Achir and C. Palamidessi. 'Public Wireless Packets Anonymously Hurt You'. In: *2021 IEEE 46th Conference on Local Computer Networks (LCN)*. IEEE LCN 2021 (Doctoral-track - Promising ideas). Edmonton / Virtual, Canada, 4th Oct. 2021. DOI: 10.1109/LCN52139.2021.9524956. URL: https://hal.archives-ouvertes.fr/hal-03298339.

[21] C. Pinzón, C. Palamidessi, P. Piantanida and F. Valencia. 'On the Impossibility of non-Trivial Accuracy in Presence of Fairness Constraints'. In: 36th AAAI Conference on Artificial Intelligence. Proceedings. Vancouver / Virtual, Canada, 22nd Feb. 2022. URL: https://hal.archives-ouvertes.fr/hal-03452324.

[22] S. Quintero, S. Ramírez, C. Rueda and F. D. Valencia. 'Computing Distributed Knowledge as the Greatest Lower Bound of Knowledge'. In: RAMICS 2021 - Relational and Algebraic Methods in Computer Science. Vol. 13027. Lecture Notes in Computer Science. Marseille, France, 19th July 2021, pp. 413–432. URL: https://hal.archives-ouvertes.fr/hal-02422624.

Scientific book chapters

[23] K. Jung, S. Biswas and C. Palamidessi. 'Establishing the Price of Privacy in Federated Data Trading'. In: *Protocols, Strands, and Logic*. Vol. 13066. Lecture Notes in Computer Science. Springer International Publishing, 19th Nov. 2021, pp. 232–250. DOI: 10.1007/978-3-030-91631-2_13. URL: https://hal.inria.fr/hal-03589837.

Reports & preprints

[24] G. Alves, F. Bernier, M. Couceiro, K. Makhlouf, C. Palamidessi and S. Zhioua. *Survey on Fairness Notions and Related Tensions*. 16th Dec. 2021. URL: https://hal.archives-ouvertes.fr/hal-03484009.

[25] N. Díaz-Rodríguez, R. Binkytė-Sadauskienė, W. Bakkali, S. Bookseller, P. Tubaro, A. Bacevicius and R. Chatila. *Questioning causality on sex, gender and COVID-19, and identifying bias in large-scale data-driven analyses: the Bias Priority Recommendations and Bias Catalog for Pandemics*. 18th May 2021. URL: https://hal.archives-ouvertes.fr/hal-03228983.

10.3 Cited publications

[26] 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].

[27] 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: 10.1109/CSF.2012.26. URL: http://hal.inria.fr/hal-00734044/en.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] M. Romanelli, K. Chatzikokolakis and C. Palamidessi. ‘Optimal Obfuscation Mechanisms via Machine Learning’. In: *CSF 2020 - 33rd IEEE Computer Security Foundations Symposium*. Preprint version of a paper that appeared on the Proceedings of the IEEE 33rd Computer Security Foundations Symposium, CSF 2020. Online, United States: IEEE, June 2020, pp. 153–168. URL: https://hal.inria.fr/hal-03091514.

[34] 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.

[35] 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.

[36] 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.