RESEARCH CENTRE

Inria Saclay Center
at Institut Polytechnique
de Paris

IN PARTNERSHIP WITH:

CNRS, Institut Polytechnique de Paris

2022
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]

Post-Doctoral Fellows

- Héber Hwang Arcolezi [INRIA, from Feb 2022]
- Hamid Jalalzai [INRIA]

PhD Students

- Andreas Athanasiou [INRIA]
- Ruta Binkyte-Sadauskiene [INRIA]
- Sayan Biswas [INRIA]
- Ganesh Del Grosso [INRIA]
- Federica Granese [INRIA]
- Karima Makhlouf [INRIA]
- Carlos Pinzon Henao [INRIA]

Technical Staff

- Gangsoo Zeong [INRIA, Engineer, until Sep 2022]
- Majid Zolfaghari [INRIA, Engineer]

Administrative Assistant

- Maria Ronco [INRIA]

Visiting Scientist

- Filippo Galli [ENS PISA, until Feb 2022]

External Collaborators

- Konstantinos Chatzikokolakis [CNRS]
- Pablo Piantanida [CentraleSupélec]

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 [34]. 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:

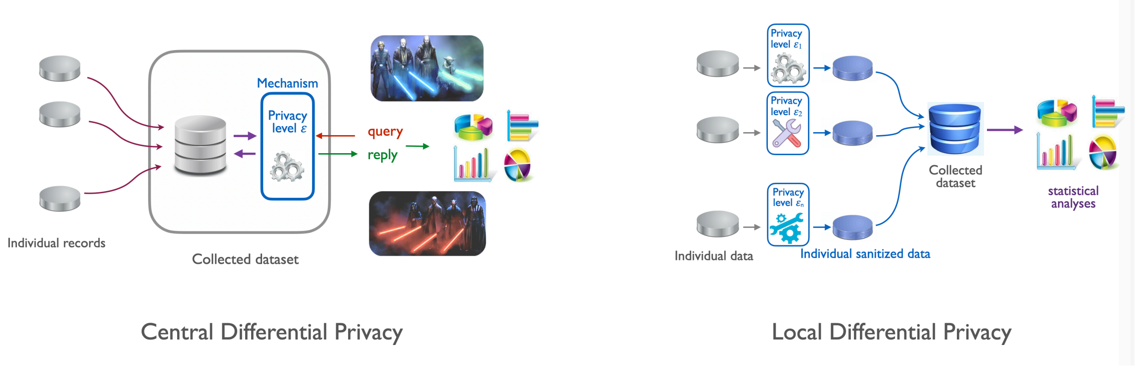

Figure 1: The central and the local models of differential privacy

- 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

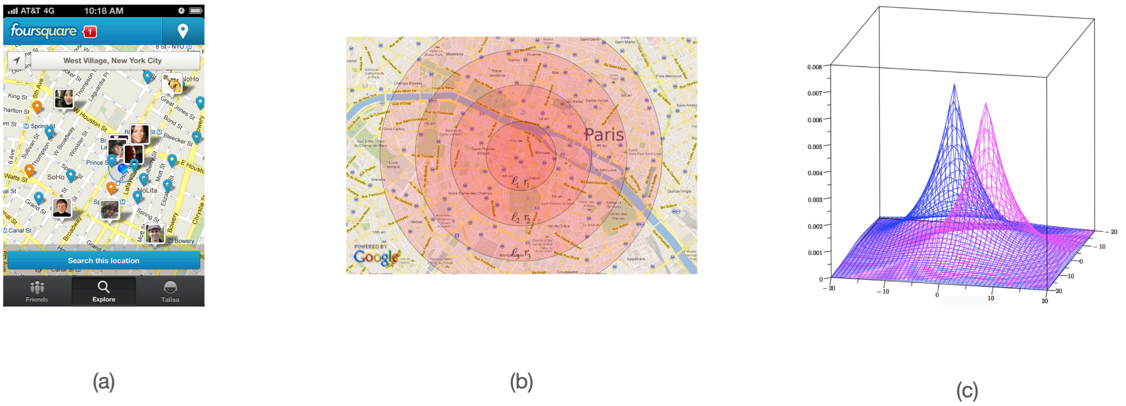

Figure 2: Geo-indistinguishability is a framework to protect the privacy of the user when dealing with location-based services (a). The framework guarrantees d-privacy, a distance-based variant of differential privacy (b). The typical implementation uses (extended) Laplace noise (c).

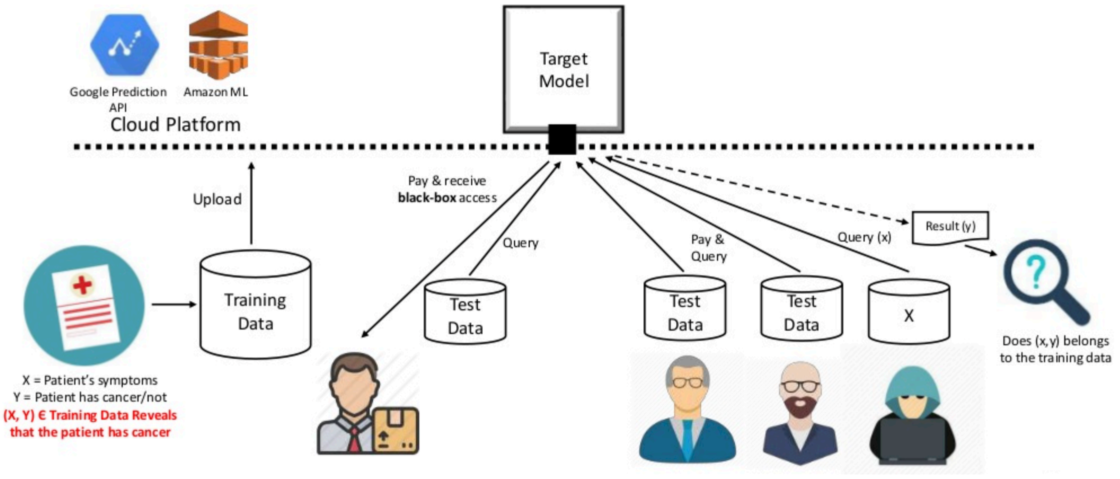

Figure 3: Privacy breach in machine learning as a service.

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. [40]). On the other hand, there are studies that show that there is a trade-off between protection from inference attacks and robustness [43]. 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* [36] and of *Equality of False Positives* and *Equality of False Negatives* [35]. 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* [44].

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 [33] 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* [32] 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 [42]. 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 [41]. 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 [38] 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 [37]. 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 Social and environmental responsibility

5.1 Footprint of research activities

Whenever possible, the members of COMETE have privileged attendance of conferences and workshops on line, to reduce the environmental impact of traveling.

6 Highlights of the year

6.1 Awards

Catuscia Palamidessi has received the Gran Prix Inria 2022 of the French Académie de Science.

6.2 Contracts

Frank Valencia obtained a 500K-euro grant from the Colombian Ministry of Science to work on Polarization in Social Networks. The grant will be used to pay doctoral studies, internships and visits of Colombian students and senior researchers at COMETE.

7 New software and platforms

7.1 New software

7.1.1 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

7.1.2 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

7.1.3 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

7.1.4 Multi-Freq-LDPy

Name: Multiple Frequency Estimation Under Local Differential Privacy in Python

Keywords: Privacy, Python, Benchmarking

Scientific Description: The purpose of Multi-Freq-LDPy is to allow the scientific community to benchmark and experiment with Locally Differentially Private (LDP) frequency (or histogram) estimation mechanisms. Indeed, estimating histograms is a fundamental task in data analysis and data mining that requires collecting and processing data in a continuous manner. In addition to the standard single frequency estimation task, Multi-Freq-LDPy features separate and combined multidimensional and longitudinal data collections, i.e., the frequency estimation of multiple attributes, of a single attribute throughout time, and of multiple attributes throughout time.

Functional Description: Local Differential Privacy (LDP) is a gold standard for achieving local privacy with several real-world implementations by big tech companies such as Google, Apple, and Microsoft. The primary application of LDP is frequency (or histogram) estimation, in which the aggregator estimates the number of times each value has been reported.

Multi-Freq-LDPy provides an easy-to-use and fast implementation of state-of-the-art LDP mechanisms for frequency estimation of: single attribute (i.e., the building blocks), multiple attributes (i.e., multidimensional data), multiple collections (i.e., longitudinal data), and both multiple attributes/collections.

Multi-Freq-LDPy is now a stable package, which is built on the well-established Numpy package - a de facto standard for scientific computing in Python - and the Numba package for fast execution.

URL: https://github.com/hharcolezi/multi-freq-ldpy

Publication: hal-03816212

Contact: Heber Hwang Arcolezi

Participants: Heber Hwang Arcolezi, Jean-François Couchot, Sébastien Gambs, Catuscia Palamidessi, Majid Zolfaghari

8 New results

Participants: Catuscia Palamidessi, Frank Valencia, Sami Zhioua, Héber Arcolezi, Hamid Jalalzai, Gangsoo Zeong, Majid Zolfaghari, Sayan Biswas, Ruta Binkyte-Sadauskiene, Ganesh Del Grosso, Natasha Fernandes, Federica Granese, Karima Makhlouf, Carlos Pinzon Henao, Santiago Quintero.

8.1 Information Leakage Games

A common goal in the areas of secure information flow and privacy is to build effective defenses against unwanted leakage of information. To this end, one must be able to reason about potential attacks and their interplay with possible defenses. In this paper we propose a game-theoretic framework to formalize strategies of attacker and defender in the context of information leakage, and provide a basis for developing optimal defense methods. A crucial novelty of our games is that their utility is given by information leakage, which in some cases may behave in a non-linear way. This causes a significant deviation from classic game theory, in which utility functions are linear with respect to players' strategies.

In [12] we have established the foundations of information leakage games. We considered two main categories of games, depending on the particular notion of information leakage being captured. The first category, which we call QIF-games, is tailored for the theory of quantitative information flow (QIF). The second one, which we call DP-games, corresponds to differential privacy (DP).

8.2 Longitudinal and multidimensional frequency estimates

In [13] we have investigated the problem of collecting multidimensional data throughout time (i.e., longitudinal studies) for the fundamental task of frequency estimation under Local Differential Privacy (LDP) guarantees. LDP is a gold standard for achieving local privacy with several real-world implementations by big tech companies such as Google, Apple, and Microsoft. Contrary to frequency estimation of a single attribute, the multidimensional aspect demands particular attention to the privacy budget. Besides, when collecting user statistics longitudinally, privacy progressively degrades. Indeed, the "multiple" settings in combination (i.e., many attributes and several collections throughout time) impose several challenges, for which our paper has proposed the first solution for frequency estimates under LDP. To tackle these issues, we have extended the analysis of three state-of-the-art LDP protocols (Generalized Randomized Response-GRR, Optimized Unary Encoding-OUE, and Symmetric Unary Encoding-SUE) for both longitudinal and multidimensional data collections. While the known literature uses OUE and SUE for two rounds of sanitization (a.k.a. memoization), i.e., L-OUE and L-SUE, respectively, we have shown analytically and experimentally that starting with OUE and then with SUE provides higher data utility (i.e., L-OSUE). Also, for attributes with small domain sizes, we have proposed Longitudinal GRR (L-GRR), which provides higher utility than the other protocols based on unary encoding. Last, we have also proposed a new solution named Adaptive LDP for LOngitudinal and Multidimensional FREquency Estimates (ALLOMFREE), which randomly samples a single attribute to be sent with the whole privacy budget and adaptively selects the optimal protocol, i.e., either L-GRR or L-OSUE. As shown in the results, ALLOMFREE consistently and considerably outperforms the state-of-the-art L-SUE and L-OUE protocols in the quality of the frequency estimates.

In [15] we have introduced the multi-freq-ldpy Python package for multiple frequency estimation under Local Differential Privacy (LDP) guarantees. The primary application of LDP is frequency (or histogram) estimation, in which the aggregator estimates the number of times each value has been reported. The presented package provides an easy-to-use and fast implementation of state-of-the-art solutions and LDP protocols for frequency estimation of: single attribute (i.e., the building blocks), multiple attributes (i.e., multidimensional data), multiple collections (i.e., longitudinal data), and both multiple attributes/collections. Multi-freq-ldpy is built on the well-established Numpy package-a de facto standard for scientific computing in Python-and the Numba package for fast execution. These features were described and illustrated in our paper with four worked examples. This package is open-source and publicly available under an MIT license via GitHub () and can be installed via PyPi ().

8.3 Human mobility

In [14] we have investigated the problem of forecasting multivariate aggregated human mobility while preserving the privacy of the individuals concerned. Differential privacy, a state-of-the-art formal notion, has been used as the privacy guarantee in two different and independent steps when training deep learning models. On one hand, we have considered gradient perturbation, which uses the differentially private stochastic gradient descent algorithm to guarantee the privacy of each time series sample in the learning stage. On the other hand, we have considered input perturbation, which adds differential privacy guarantees in each sample of the series before applying any learning. We have compared four state-of-the-art recurrent neural networks: Long ShortTerm Memory, Gated Recurrent Unit, and their Bidirectional architectures, i.e., Bidirectional-LSTM and Bidirectional-GRU. Extensive experiments were conducted with a real-world multivariate mobility dataset, which we have published openly along with this paper. As shown in the results, differentially private deep learning models trained under gradient or input perturbation achieve nearly the same performance as non-private deep learning models, with loss in performance varying between 0.57% to 2.8%. The contribution of this paper is significant for those involved in urban planning and decision-making, providing a solution to the human mobility multivariate forecast problem through differentially private deep learning models.

8.4 Metric differential privacy

In [20] we have studied the privacy-utility trade-off in the context of metric differential privacy. Ghosh et al. introduced the idea of universal optimality to characterize the “best” mechanism for a certain query that simultaneously satisfies (a fixed) differential privacy constraint whilst at the same time providing better utility compared to any other differentially private mechanism for the same query. They showed that the Geometric mechanism is universally optimal for the class of counting queries. On the other hand, Brenner and Nissim showed that outside the space of counting queries, and for the Bayes risk loss function, no such universally optimal mechanisms exist. Except for universal optimality of the Laplace mechanism, there have been no generalizations of these universally optimal results to other classes of differentially-private mechanisms. In this paper we have used metric differential privacy and quantitative information flow as the fundamental principle for studying universal optimality. Metric differential privacy is a generalization of both standard (i.e., central) differential privacy and local differential privacy, and it is increasingly being used in various application domains, for instance in location privacy and in privacy preserving machine learning. As Ghosh et al. and Brenner and Nissim did, we have measure utility in terms of loss functions, and we have interpreted the notion of a privacy mechanism as an information-theoretic channel satisfying constraints defined by ε-differential privacy and a metric meaningful to the underlying state space. Using this framework we were able to clarify Nissim and Brenner’s negative results by (a) that in fact all privacy types contain optimal mechanisms relative to certain kinds of non-trivial loss functions, and (b) extending and generalizing their negative results beyond Bayes risk specifically to a wide class of non-trivial loss functions. Our exploration suggests that universally optimal mechanisms are indeed rare within privacy types. We therefore have proposed weaker universal benchmarks of utility called privacy type capacities. We have shown that such capacities always exist and can be computed using a convex optimization algorithm. Finally, we have illustrated these ideas on a selection of examples with several different underlying metrics.

8.5 The shuffle model

The shuffle model is an intermediate paradigm between the central and the local models of differential privacy (DP), and it has recently gained popularity. As an initial step, the shuffle model uses a local mechanism to perturb the data individually like the local model of DP. After this local sanitization, a shuffler uniformly permutes the noisy data to dissolve their links with the corresponding data providers. This allows the shuffle model to achieve a certain level of DP guarantee using less noise than the local model, thus providing a better utility for the same level of privacy.

The privacy guarantees provided by the shuffle model have been rigorously studied by community of late and various results have been derived, both analytical and numerical. Obviously, analytical bounds have the advantage that they provide a concrete basis for reasoning and mathematically analyzing properties such as privacy-utility trade-off. However, in the case of the shuffle model, most analytical bounds found in the literature are far from being tight. In [18], we have covered this gap and derive tight necessary and sufficient condition for having the tightest (ϵ, δ)-bounds for the DP guarantee provided by the shuffle model with the k-randomized response local mechanism.

8.6 Membership inference attacks

The use of personal data for training machine learning systems comes with a privacy threat and measuring the level of privacy of a model is one of the major challenges in machine learning today. Identifying training data based on a trained model is a standard way of measuring the privacy risks induced by the model. In [19], we have developed a novel approach to address the problem of membership inference in pattern recognition models, relying on information provided by adversarial examples. The strategy we have proposed consists of measuring the magnitude of a perturbation necessary to build an adversarial example. Indeed, we argue that this quantity reflects the likelihood of belonging to the training data. Extensive numerical experiments on multivariate data and an array of state-of-the-art target models have shown that our method performs comparable or even outperforms stateof-the-art strategies, but without requiring any additional training samples.

8.7 Adversarial examples

Detection of adversarial examples in machine learning has been a hot topic in the last years due to its importance for safely deploying algorithms in critical applications. However, the detection methods are generally validated by assuming a single implicitly known attack strategy, which does not necessarily account for real-life threats. Indeed, this can lead to an overoptimistic assessment of the detectors' performance and may induce some bias in the comparison between competing detection schemes. To overcome this limitation, in [21] we have proposed a novel multi-armed framework, called MEAD, for evaluating detectors based on several attack strategies. We have made use of three new objectives to generate attacks. The proposed performance metric is based on the worst-case scenario: detection is successful if and only if all different attacks are correctly recognized. We have shown empirically the effectiveness of our approach. Moreover, the poor performance obtained for state-of-the-art detectors opens a new exciting line of research.

8.8 The cost of fairness

One of the main concerns about fairness in machine learning (ML) is that, in order to achieve it, one may have to trade off some accuracy. To overcome this issue, Hardt et al. [39] proposed the notion of equality of opportunity (EO), which is compatible with maximal accuracy when the target label is deterministic with respect to the input features. In the probabilistic case, however, the issue is more complicated: It has been shown that under differential privacy constraints, there are data sources for which EO can only be achieved at the total detriment of accuracy, in the sense that a classifier that satisfies EO cannot be more accurate than a trivial (i.e., constant) classifier. In [22] we have strengthened this result by removing the privacy constraint. Namely, we have shown that for certain data sources, the most accurate classifier that satisfies EO is a trivial classifier. Furthermore, we have studied the trade-off between accuracy and EO loss (opportunity difference), and have provided a sufficient condition on the data source under which EO and non-trivial accuracy are compatible.

8.9 Causal discovery

It is crucial to consider the social and ethical consequences of AI and ML based decisions for the safe and acceptable use of these emerging technologies. Fairness, in particular, guarantees that the ML decisions do not result in discrimination against individuals or minorities. Identifying and measuring reliably fairness/discrimination is better achieved using causality which considers the causal relation, beyond mere association, between the sensitive attribute (e.g. gender, race, religion, etc.) and the decision (e.g. job hiring, loan granting, etc.). The big impediment to the use of causality to address fairness, however, is the unavailability of the causal model (typically represented as a causal graph). Existing causal approaches to fairness in the literature do not address this problem and rely on the availability of the causal model. In [16], we reviewed the major algorithms to discover causal relations from observable data. Our study was focused on causal discovery and its impact on fairness. In particular, we have shown how different causal discovery approaches may result in different causal models and, most importantly, how even slight differences between causal models can have significant impact on fairness/discrimination conclusions. These results were consolidated by empirical analysis using synthetic and standard fairness benchmark datasets.

8.10 Polarization under Confirmation Bias

In our team we have developed models for polarization in multi-agent systems based on Esteban and Ray's standard family of polarization measures from economics. 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. In [11] 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 determine the unique belief value to which all agents converge. Our more insightful result in [11] 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 proved that polarization does not necessarily vanish in weakly-connected graphs under confirmation bias. Furthermore, we showed how our model relates to the classic DeGroot model for social learning. We illustrated our model with several simulations of a running example about polarization over vaccines and of other case studies. The theoretical results and simulations in [11] provided insight into the phenomenon of polarization.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Inria associate team not involved in an IIL or an international program

LOGIS

Participants: Ruta Binkite-Sadauskiene, Sayan Biswas, Catuscia Palamidessi, Gangsoo Zeong.

Web Page: Link

Title: Logical and Formal Methods for Information Security and privacy

Duration: 2019 - 2022

Coordinator: Mitsuhiro Okada (mitsu@abelard.flet.keio.ac.jp)

Partners: Keio University, AIST, and Waseda University (Japan)

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 this 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).

9.1.2 Participation in other International Programs

FACTS

Participants: Frank Valencia, Catuscia Palamidessi, Carlos Pinzón.

Title: Foundational Approach to Cognition in Today's Society

Program: ECOS Nord

Partner Institution: Universidad Javeriana Cali, Colombia

Duration: June 2019 - May 2023

Description: The goal of this project is to develop a process calculus to study the dynamics of agents in a distributed system, the communication flow, and the formation of individual opinions and beliefs. The expected outcome is a computational model that will allow us to describe the interaction of groups of agents exchanging epistemic information among them.

PROMUEVA

Participants: Frank Valencia, Catuscia Palamidessi.

Title: Polarization Research On Social Networks

Program: Colombian Grants from royalties generated from hydrocarbons for technology and scientific research.

Partner Institution(s):
- Universidad Javeriana Cali, Colombia
- Universidad del Valle, Colombia

Duration: August 2022 - July 2026

Description: The role of social networks is relatively paradoxical. On the one hand, the world is more interconnected, we have better access to information and to different opinions. On the other hand, they can shape the opinions of users on an unprecedented scale, giving rise to wide polarization and leading to fractures within society. In this project, we aim to design models to analyze this phenomenon using formal methods and adapting economic and statistical models that had already tackled similar problems. The main objective is to explain the phenomenon of belief formation under the effect of cognitive biases, and show that social networks can be designed differently, adjusting algorithms for biases that will impact the flow of influence and thus mitigate the phenomenon of polarization.

9.2 International research visitors

9.2.1 Visits of international scientists

Filippo Galli

Status PhD student

Institution of origin: Scuola Normale Superiore, Pisa

Country: Italy

Dates: 1 January 2022 - 28 February 2022

Context of the visit: Collaboration with Sayan Biswas, Catuscia Palamidessi anad Gangsoo Zeong on privacy protection in machine learning

Mobility program/type of mobility: Internship

Mario Alvim

Status Assistant professor

Institution of origin: Universidade Federal de Minas Gerais (UFMG), Belo Horizonte

Country: Brazil

Dates: 12 February 2022 - 15 March 2022

Context of the visit: Collaboration with Catuscia Palamidessi, Ramon Gonze and Mireya Jurado on a Bayesian interpretation of the shuffle model.

Mobility program/type of mobility: Research visit

Ramon Gonze

Status Master student

Institution of origin: Universidade Federal de Minas Gerais (UFMG), Belo Horizonte

Country: Brazil

Dates: 12 February 2022 - 26 March 2022

Context of the visit: Collaboration with Catuscia Palamidessi, Mario Alvim, and Mireya Jurado on a Bayesian interpretation of the shuffle model.

Mobility program/type of mobility: Research visit

Mireya Jurado

Status PhD student

Institution of origin: Florida International University

Country: USA

Dates: 12 February 2022 - 26 March 2022

Context of the visit: Collaboration with Catuscia Palamidessi, Mario Alvim and Ramon Gonze on a Bayesian interpretation of the shuffle model.

Mobility program/type of mobility: Research visit

Daniele Gorla

Status Associate professor

Institution of origin: Università di Roma "La Sapienza"

Country: Italy

Dates: 17 August 2022 - 16 September 2022

Context of the visit: Collaboration with Catuscia Palamidessi and Federica Granese on the back-box estimation of the degree of privacy in differential privacy mechanisms.

Mobility program/type of mobility: Research visit

Gerardo Sarria

Status Associate professor

Institution of origin: Pontificia Universidad Javeriana Cali.

Country: Colombia

Dates: 30 Sept 2022 - 15 October 2022

Context of the visit: Collaboration with Frank Valencia funded by the ECOS Nord project FACTS two work on a Foundational Approach to Cognition.

Mobility program/type of mobility: Research visit

9.2.2 Visits to international teams

Federica Granese

Visited institution: École de technologie supérieure (ÉTS) in Montreal

Country: Canada

Dates: October 2022 - February 2023

Context of the visit: Collaboration with José Dolz on Error detection in image segmentation tasks

Mobility program/type of mobility: Internship

Héber Arcolezi

Visited institution: The University of British Columbia (UBC), Vancouver

Country: Canada

Dates: October 2022 - December 2022

Context of the visit: Collaboration with Profs. Mathias Lécuyer and Sébastien Gambs

Mobility program/type of mobility: Research visit

9.3 European initiatives

9.3.1 Horizon Europe

ELSA

Participants: Catuscia Palamidessi, Gangsoo Zeong.

Web Page: Link

Title: European Lighthouse on Secure and Safe AI

Duration: From September 1, 2022 to August 31, 2025

Partners:

- Institut National de Recherche in Informatique et Automatique (INRIA), France
- Pal Robotics SL, Spain
- Yooz SAS, France
- Helsingin Yliopisto, Finland
- Pluribus One SRL, Italy
- Kungliga Tekniska Hoegskolan (KTH), Sweden
- European Molecular Biology Laboratory (EMBL), Germany
- The University of Birgmigham (UoB), United Kingdom
- Università degli Studi di Cagliari (UNICA), Italy
- Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland
- Valeo Comfort and Driving Assistance, France
- NVIDIA Switzerland AG, Switzerland
- The Alan Turing Institute, United Kingdom
- Fondazione Istituto Italiano di Tecnologia (IIT), Italy
- Eidgenoessiche Technische Hochschule Zürich (ETH Zürich), Switzerland
- Lancaster University, United Kingdom
- Politecnico di Torino (POLITO), Italy
- Università degli Studi di Milano (UMIL), Italy
- CISPA - Helmholtz-Zentrum fur Informationssicherheit GGMBH, Germany
- Leonardo - Società per Azioni, Italy
- The University of Oxford (UOXF), United Kingdom
- Università degli Studi di Genova (UNIGE), Italy
- Max-Planck-Gesellschaft zur Forderung der Wissenschaften EV (MPG), Germany
- Centre de Visio per Computador (CVC), Spain
- Università degli Studi di Modena e Reggio Emilia (UNIMORE), Italy
- Consorzio Interuniversitario Nazionale per l'Informatica (CINI), Italy

Inria contact: Catuscia Palamidessi

Coordinator: Mario Fritz, CISPA

Description: In order to reinforce European leadership in safe and secure AI technology, we are proposing a virtual center of excellence on safe and secure AI that will address major challenges hampering the deployment of AI technology. These grand challenges are fundamental in nature. Addressing them in a sustainable manner requires a lighthouse rooted in scientific excellence and rigorous methods. We will develop a strategic research agenda which is supported by research programmes that focus on “technical robustness and safety”, “privacy preserving techniques and infrastructures” and “human agency and oversight”. Furthermore, we focus our efforts to detect, prevent and mitigate threats and enable recovery from harm by 3 grand challenges: “Robustness guarantees and certification”, “Private and robust collaborative learning at scale” and “Human-in-the-loop decision making: Integrated governance to ensure meaningful oversight” that cut across 6 use cases: health, autonomous driving, robotics, cybersecurity, multi-media, and document intelligence. Throughout our project, we seek to integrate robust technical approaches with legal and ethical principles supported by meaningful and effective governance architectures to nurture and sustain the development and deployment of AI technology that serves and promotes foundational European values. Our initiative builds on and expands the internationally recognized, highly successful and fully operational network of excellence ELLIS (European Laboratory for Learning and Intelligent Systems). We build ELSA on its 3 pillars: research programmes, a set of research units, and a PhD/postdoc programme, thereby connecting a network of over 100 organizations and more than 337 ELLIS fellows and scholars (113 ERC grants) committed to shared standards of excellence. We will not only establish a virtual center of excellence, but all our activities will be also inclusive and open to input, interactions and collaboration of AI researchers and industrial partners in order to drive the entire field forward.

9.3.2 H2020 projects

HYPATIA

Participants: Catuscia Palamidessi, Sami Zhioua, Héber Arcolezi, Hamid Jalalzai, Majid Zolfaghari, Sayan Biswas, Ruta Binkyte-Sadauskiene, Ganesh Del Grosso, Natasha Fernandes, Federica Granese, Karima Makhlouf, Carlos Pinzon Henao.

Web Page: Link

Title: Privacy and Utility Allied

Duration: From October 1, 2019 to September 30, 2024

Partners: Institut National de Recherche in Informatique et Automatique (INRIA), France

Inria contact: Catuscia Palamidessi

Coordinator: Catuscia Palamidessi

Description: With the ever-increasing use of internet-connected devices, such as computers, smart grids, 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. Undeniably, the big-data technology provides enormous benefits to industry, individuals and society, ranging from improving business strategies and boosting quality of service to enhancing scientific progress. On the other hand, however, the collection and manipulation of personal data raises alarming privacy issues. Both the experts and 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. 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: (1) can be applied and controlled directly by the user, thus avoiding the need of a trusted party, (2) are robust with respect to combination of information from different sources, and (3) provide an optimal trade-off between privacy and utility. We intend to pursue these goals by developing a new framework for privacy based on the addition of controlled noise to individual data, and associated methods to recover the useful statistical information, and to protect the quality of service.

9.3.3 Other european programs/initiatives

CRYPTECS

Participants: Catuscia Palamidessi, Kostas Chatzikokolakis, Andreas Athanasiou.

Web Page: Link

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 (France) and Sven Trieflinger (Germany)

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.

9.4 National initiatives

iPOP

Participants: Catuscia Palamidessi, Sami Zhioua, Héber Arcolezi, Sayan Biswas, Ruta Binkyte-Sadauskiene, Karima Makhlouf.

Web Page: Link

Title: Interdisciplinary Project on Privacy

Program: PEPR Cybersecurity

Duration: 1 October 2022 - 30 September 2028

Coordinators: Antoine Boutet (Insa-Lyon) - Vincent Roca (Inria)

Partners:

- Inria
- CNRS
- CNIL
- INSA-Centre Val de Loire (CVL)
- INSA-Lyon
- Université Grenoble Alpes
- Université de Lille
- Université Rennes 1
- Université de Versailles Saint-Quentin-en-Yvelines

Inria COMETE contact: Catuscia Palamidessi

Description: Digital technologies provide services that can greatly increase quality of life (e.g. connected e-health devices, location based services or personal assistants). However, these services can also raise major privacy risks, as they involve personal data, or even sensitive data. Indeed, this notion of personal data is the cornerstone of French and European regulations, since processing such data triggers a series of obligations that the data controller must abide by. This raises many multidisciplinary issues, as the challenges are not only technological, but also societal, judiciary, economic, political and ethical. The objectives of this project are thus to study the threats on privacy that have been introduced by these new services, and to conceive theoretical and technical privacy-preserving solutions that are compatible with French and European regulations, that preserve the quality of experience of the users. These solutions will be deployed and assessed, both on the technological and legal sides, and on their societal acceptability. In order to achieve these objectives, we adopt an interdisciplinary approach, bringing together many diverse fields: computer science, technology, engineering, social sciences, economy and law.

FedMalin

Participants:	Catuscia Palamidessi, Sami Zhioua, Héber Arcolezi, Sayan Biswas, Ruta Binkyte-Sadauskiene, Karima Makhlouf.

Web Page: Link

Title: Federated MAchine Learning over the INternet

Program: Inria Challenge

Duration: 1 October 2022 - 30 September 2026

Coordinators: Aurélien Bellet and Giovanni Neglia

Partners:

- ARGO (Inria Paris)
- COATI (Inria Sophia)
- COMETE (Inria Saclay)
- EPIONE (Inria Sophia)
- MAGNET (Inria Lille)
- MARACAS (Inria Lyon)
- NEO (Inria Sophia)

- SPIRALS (Inria Lille)
- TRIBE (Inria Saclay)
- WIDE (Inria Rennes)

Inria COMETE contact: Catuscia Palamidessi

Description: In many use-cases of Machine Learning (ML), data is naturally decentralized: medical data is collected and stored by different hospitals, crowdsensed data is generated by personal devices, etc. Federated Learning (FL) has recently emerged as a novel paradigm where a set of entities with local datasets collaboratively train ML models while keeping their data decentralized. FedMalin aims to push FL research and concrete use-cases through a multidisciplinary consortium involving expertise in ML, distributed systems, privacy and security, networks, and medicine. We propose to address a number of challenges that arise when FL is deployed over the Internet, including privacy and fairness, energy consumption, personalization, and location/time dependencies. FedMalin will also contribute to the development of open-source tools for FL experimentation and real-world deployments, and use them for concrete applications in medicine and crowdsensing.

9.5 Regional initiatives

LOST2DNN

Participants: Catuscia Palamidessi, Ganesh Del Grosso, Federica Granese, Pablo Piantanida.

Web Page: Link

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.

10 Dissemination

Participants: Catuscia Palamidessi, Frank Valencia, Sami Zhioua, Héber Arcolezi, Ruta Binkyte-Sadauskiene, Karima Makhlouf, Carlos Pinzon Henao.

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

The team Comete has organized a workshop on ethical AI. Campus de l'Ecole Polytechnique, Palaiseau, 30 September 2022

10.1.2 Scientific events: selection

Catuscia Palamidessi has been member of the selection committee of the following conferences and workshops:

- FOSSACS 2024. The 27th International Conference on Foundations of Software Science and Computation Structures. (Part of ETAPS 2024.) Luxembourg, 2024.
- CSF 2023. The 36th IEEE Computer Security Foundations Symposium. Dubrovnik, Croatia, July 10 - 14, 2023.
- LICS 2023. The Thirty-Eighth Annual ACM/IEEE Symposium on Logic in Computer Science. Boston, USA, 26–29 June 2023.
- FACS 2022. The 18th International Conference on Formal Aspects of Component Software. Oslo, Norway, 10-11 November 2022.
- 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.
- SDS 2023. The 10th IEEE Swiss Conference on Data Science. Zurich, Switzelrland, June 22 – 23, 2023.
- WIL 2023. The 7th Women in Logic Workshop. Rome, Italy, June 1st , 2023.
- FTSCS 2022. The 8th International Workshop on Formal Techniques for Safety-Critical Systems. Auckland, New Zealand, December 7, 2022.
- 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.

Frank Valencia has been member of the selection committee of the following conferences and workshops:

- FOSSACS 2023. The 26th International Conference on Foundations of Software Science and Computation Structures. (Part of ETAPS 2023.) Paris, 22-27 April 2023.
- ICLP 2023. 39th International Conference on Logic Programming. London, July 9 - 15, 2023
- CLEI 2022. The 48th International Latin American Conference on Informatics. Armenia, Colombia, 17-21 October 2022.
- ICLP-DC 2022 The 38th International Conference on Logic Programming (ICLP) - DC. Haifa, Israel, July 31-August 8, 2022.

10.1.3 Journals

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

- (2022-) Member of the Editorial Board of the ACM Transactions on Privacy and Security (TOPS), ACM.
- (2021-22) 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.
- (2006-) Member of the Editorial Board of Mathematical Structures in Computer Science, Cambridge University Press.

10.1.4 Invited talks

Catuscia Palamidessi has given the following invited talks:

- FLOC 2022. Keynote speaker at the Federated Logic Conference. Haifa, Israel, July-August 2022.
- AFCP 2022. The NeurIPS workshop on Algorithmic Fairness through the Lens of Causality and Privacy. New Orleans, USA, December 3, 2022.
- Talk on Differential Privacy at the Collège de France. 24 March 2022.

Frank Valencia has given the following invited talk:

- IEEE ComSoc SIoA SIG Invited Speaker at the IEEE Social Networks Technical Committee SIG Seminar on Modelling Bias and Polarization in Social Networks. 28 October 20222.

10.1.5 Leadership within the scientific community

Since July 2002 Catuscia palamidessi is Chair of SIGLOG, the ACM Special Interest Group on Logic and Computation.

10.1.6 Scientific expertise

Catuscia Palamidessi is/has been:

- (2021-) Member of the Board of Trustees of the IMDEA Software Institute.
- (2019-22) 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.
- (2023) Member of the committee for a faculty position at the ETH Zurich, Switzerland.
- (2023) Member of the committee for a professor position at the INSA Lyon, France.
- (2022-23) Member of the committee for the UK-US Prize Challenge on Advancing Privacy-Preserving Federated Learning. Sponsored by NIST and NSF.
- (2022) Member of the review panel for the Italian Ministry of Universities and Research research grants.
- (2022) Member of the review panel for the Swiss National Science Foundation advanced grants.
- (2022) Member of the review panel for the Estonian Research Council grants.
- (2022) Member of the review panel for the Swedish Research Council consolidator grants.
- (2022) Member of the committee for the CCS Test-of-Time Award.
- (2022) Member of the committee for a professor position at the Université Sorbonne Paris Nord, France.

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

Frank Valencia has been teaching a Discrete Math course and Computability course at la Pontificia Universidad Javeriana. Each course consists 42 hours of lectures.

Catuscia Palamidessi has given the following courses:

- Course on Privacy at the ELSA Summer school on Trustworthy AI. Saarbrücken, Germany. September 2022.
- Course on Privacy and Fairness at the IDESSAI, the second Inria-DFKI European Summer School on AI. Saarbrücken, Germany. August-September 2022.

10.2.2 Supervision

In COMETE we are supervising the following PhD students:

- (2023-) Ramon Gonze. IPP and Federal University of Minas Gerais. Co-supervised by Catuscia Palamidessi and by Mario Alvim. Thesis subject: Bayesian guarrantees of differential privacy.
- (2022-) Andreas Athanasiou. IPP and University of Athens. Co-supervised by Catuscia Palamidessi and by Kostantinos Chatzikokolakis. Thesis subject: The shuffle model for d-privacy.
- (2021-) Karima Makhlouf. IPP. Thesis subject: Relation between privacy and fairness in machine learning. Co-supervised by Catuscia Palamidessi and by Héber Arcolezi.
- (2020-) Ruta Binkite-Saudaskiene. IPP. Co-supervised by Catuscia Palamidessi and by Sami Zhioua. 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.

- (2020-) Sayan Biswas. 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 by Pablo Piantanida. Thesis subject: Privacy in Machine Learning.

In 2022, in COMETE we have supervised the following master students and interns:

- Simon Sebastian. Bacelor student, IPP, France. Co-supervised by Carlos Pinzon and Catuscia Palamidessi. From May 2022 until August 2022.
- Cesara Petrui. Bacelor student, IPP, France. Co-supervised by Carlos Pinzon and Catuscia Palamidessi. From May 2022 until August 2022.
- Filippo Galli. Visiting PhD student, ENS Pisa, Italy. Supervised by Catuscia Palamidessi. From Sept 2021 until March 2022.

10.2.3 Other advisory activity

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

- (2022-24) Member of the advising committee for Martina Cinquini, PhD student supervised by Salvatore Ruggieri, University of Pisa, Italy.
- (2012-) External member of the scientific council for the PhD in Computer Science at the University of Pisa, Italy.
- (2020-22) Member of the advising committee of Abhishek Sharma, PhD student supervised by Maks Ovsjanikov, IPP, France.

10.2.4 Juries

Catuscia Palamidessi has been member of the jury in the following PhD defenses:

- Guilherme Alves (LORIA, France). Member of the committee board at the PhD defense. Title of the thesis: *Hybrid Approaches for Algorithmic Fairness.* Advised by Miguel Couceiro and Amedeo Napoli. Defended in December 2022.
- Elli Anastasiadi (Reykjavik University, Island). Member of the committee board at the PhD defense. Title of the thesis: *Runtime and Equational Verification of Concurent Systems.* Advised by Luca Aceto and Anna Ingólfsdóttir. Defended in October 2022.
- Arthur Américo (Queen Mary University of London, UK). PhD thesis reviewer and member of the committee at the PhD defense. Title of the thesis: *From Quantitative Information Flow to Information Theory and Quantum Leakage.* Advised by Pasquale Malacaria. Defended in September 2022.
- Vincent Grari (University of Sorbonne, France). Member of the committee at the PhD defense. Title of the thesis: *Adversarial mitigation to reduce unwanted biases in machine learning.* Advised by Marcin Detyniecki and Sylvain Lamprier. Defended in June 2022.

10.3 Popularization

10.3.1 Articles and contents

Frank Valencia has interviewed for the following media:

- Chut! Interview *Les réseaux sociaux fournissent le milieu idéal pour la prolifération des biais cognitifs.* Jan 12, 2023.
- Inria Website Interview on Social networks: Can mathematical modeling help reduce polarization of opinions? May 11, 2022.
- ToT Curses Article on *Un modéle mathématique pour réduire la polarisation des opinions.* May 18, 2022.
- Cali24horas Radio Interview on Polarization on Social Media. March 17, 2022.

10.3.2 Interventions

Catuscia Palamidessi is invited speaker at EMW 2023, the ETAPS Mentoring Workshop. This workshop aims at encouraging graduate students and senior undergraduate students to pursue careers in programming language research, and at educating them on the research career. 23 April 2023

10.4 Administrative responsibilities

Since 2021, Catuscia Palamidessi is president of the Commission Scientifique of INRIA Saclay.

Since 2020, Frank D. Valencia is a member of the *Conseil de laboratoire* of LIX, Ecole Polytechnique.

11 Scientific production

11.1 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.

11.2 Publications of the year

International journals

[11] M. S. Alvim, B. Amorim, S. Knight, S. Quintero and F. Valencia. 'A Formal Model for Polarization under Confirmation Bias in Social Networks'. In: *Logical Methods in Computer Science* (1st Dec. 2022). URL: https://hal.archives-ouvertes.fr/hal-03872692.

[12] M. S. Alvim, K. Chatzikokolakis, Y. Kawamoto and C. Palamidessi. 'Information Leakage Games: Exploring Information as a Utility Function'. In: *ACM Transactions on Privacy and Security* (2022). DOI: 10.1145/3517330. URL: https://hal.archives-ouvertes.fr/hal-03091413.

[13] H. H. Arcolezi, J.-F. Couchot, B. Al Bouna and X. Xiao. 'Improving the utility of locally differentially private protocols for longitudinal and multidimensional frequency estimates'. In: *Digital Communications and Networks* (July 2022). DOI: 10.1016/j.dcan.2022.07.003. URL: https://hal.inria.fr/hal-03727621.

[14] H. H. Arcolezi, J.-F. Couchot, D. Renaud, B. Al Bouna and X. Xiao. 'Differentially private multivariate time series forecasting of aggregated human mobility with deep learning: Input or gradient perturbation?' In: *Neural Computing and Applications* 34 (3rd June 2022), pp. 13355–13369. DOI: 10.1007/s00521-022-07393-0. URL: https://hal.inria.fr/hal-03689723.

International peer-reviewed conferences

[15] H. H. Arcolezi, J.-F. Couchot, S. Gambs, C. Palamidessi and M. Zolfaghari. 'Multi-Freq-LDPy: Multiple Frequency Estimation Under Local Differential Privacy in Python'. In: ESORICS 2022 - European Symposium on Research in Computer Security. Vol. 13556. Lecture Notes in Computer Science. Copenhague, Denmark: Springer Nature Switzerland, 24th Sept. 2022, pp. 770–775. DOI: 10.1007/978-3-031-17143-7_40. URL: https://hal.inria.fr/hal-03816212.

[16] R. Binkytė-Sadauskienė, K. Makhlouf, C. Pinzón, S. Zhioua and C. Palamidessi. 'Causal Discovery for Fairness'. In: *Proceedings of the 3rd Workshop on Algorithmic Fairness through the Lens of Causality and Privacy (AFCP 2022)*. Algorithmic Fairness through the Lens of Causality and Privacy (AFCP) - Workshop affiliated with Neural Information Processing systems (NeurIPS). New Orleans, United States, 3rd Dec. 2022. URL: https://hal.inria.fr/hal-03911551.

[17] S. Biswas, G. Cormode and C. Maple. 'Impact of sampling on locally differentially private data collection'. In: *Proceedings of the Eight Conference on Competitive Advantage in the Digital Economy (CADE)*. CADE 2022 - Competitive Advantage in the Digital Economy. Venice, Italy: Institution of Engineering and Technology; IEEE, 13th June 2022, pp. 64–70. DOI: 10.1049/icp.2022.2042. URL: https://hal.science/hal-03846611.

[18] S. Biswas, K. Jung and C. Palamidessi. 'Tight Differential Privacy Blanket for the Shuffle Model'. In: *Proceedings of the Eight International Conference on Competitive Advantage in the Digital Economy (CADE 2022)*. CADE 2022 - Competitive Advantage in the Digital Economy. Venice, Italy: IET Digital Library; IEEE Xplore, 13th June 2022, pp. 61–63. DOI: 10.1049/icp.2022.2041. URL: https://hal.science/hal-03846624.

[19] G. Del Grosso, H. Jalalzai, G. Pichler, C. Palamidessi and P. Piantanida. 'Leveraging Adversarial Examples to Quantify Membership Information Leakage'. In: *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, United States: IEEE, 18th June 2022, pp. 10389–10399. DOI: 10.1109/CVPR52688.2022.01015. URL: https://hal.science/hal-03919891.

[20] N. Fernandes, A. Mciver, C. Palamidessi and M. Ding. 'Universal Optimality and Robust Utility Bounds for Metric Differential Privacy'. In: 2022 IEEE 35th Computer Security Foundations Symposium (CSF). Haifa, Israel: IEEE, 7th Aug. 2022, pp. 348–363. DOI: 10.1109/CSF54842.2022.9919647. URL: https://hal.inria.fr/hal-03909798.

[21] F. Granese, M. Picot, M. Romanelli, F. Messina and P. Piantanida. 'MEAD: A Multi-Armed Approach for Evaluation of Adversarial Examples Detectors'. In: *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2022)*. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Grenoble, France, 30th June 2022. URL: https://hal.inria.fr/hal-03909893.

[22] C. Pinzón, C. Palamidessi, P. Piantanida and F. Valencia. 'On the Impossibility of non-Trivial Accuracy in Presence of Fairness Constraints'. In: Proceedings of the AAAI 36th Conference on Artificial Intelligence. Vol. 36. Proceedings 7. Vancouver / Virtual, Canada, 30th June 2022, pp. 7993–8000. DOI: 10.1609/aaai.v36i7.20770. URL: https://hal.science/hal-03452324.

Conferences without proceedings

[23] F. Galli, S. Biswas, K. Jung, T. Cucinotta and C. Palamidessi. 'Group privacy for personalized federated learning'. In: International Workshop on Federated Learning: Recent Advances and New Challenges in Conjunction with NeurIPS 2022 (FL-NeurIPS'22). New Orleans, United States, 2nd Dec. 2022. URL: https://hal.inria.fr/hal-03907130.

Reports & preprints

[24] G. Alves, F. Bernier, M. Couceiro, K. Makhlouf, C. Palamidessi and S. Zhioua. *Survey on Fairness Notions and Related Tensions*. 10th June 2022. URL: https://hal.archives-ouvertes.fr/hal-03484009.

[25] H. H. Arcolezi, C. Pinzón, C. Palamidessi and S. Gambs. *Frequency Estimation of Evolving Data Under Local Differential Privacy*. 23rd Dec. 2022. URL: https://hal.inria.fr/hal-03911550.

[26] S. Biswas and C. Palamidessi. *PRIVIC: A privacy-preserving method for incremental collection of location data*. 2022. URL: https://hal.inria.fr/hal-03968692.

[27] H. Jalalzai, E. Kadoche, R. Leluc and V. Plassier. *Membership Inference Attacks via Adversarial Examples*. 22nd Dec. 2022. URL: https://hal.inria.fr/hal-03910286.

[28] K. Makhlouf, S. Zhioua and C. Palamidessi. *Identifiability of Causal-based Fairness Notions: A State of the Art*. 3rd Jan. 2023. URL: https://hal.archives-ouvertes.fr/hal-03920431.

[29] C. Pinzón, S. Quintero, S. Ramírez, C. Rueda and F. Valencia. *Counting and Computing Join-Endomorphisms in Lattices (Revisited)*. 25th July 2022. URL: https://hal.inria.fr/hal-03864755.

[30] S. Quintero, C. Pinzón, S. Ramírez and F. Valencia. *On the Computation of Distributed Knowledge as the Greatest Lower Bound of Knowledge*. 20th Aug. 2022. URL: https://hal.inria.fr/hal-03864537.

[31] S. Simon, C. Petrui, C. Pinzón and C. Palamidessi. *Minimizing Information Leakage under Padding Constraints*. 23rd Dec. 2022. URL: https://hal.inria.fr/hal-03911552.

11.3 Cited publications

[32] 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].

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] F. Granese, D. Gorla and C. Palamidessi. 'Enhanced Models for Privacy and Utility in Continuous-Time Diffusion Networks'. In: *International Journal of Information Security* 20.5 (2021), pp. 673–782. DOI: 10.1007/s10207-020-00530-7. URL: https://hal.inria.fr/hal-03094843.

[39] M. Hardt, E. Price and N. Srebro. 'Equality of Opportunity in Supervised Learning'. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*. NIPS'16. Barcelona, Spain: Curran Associates Inc., 2016, pp. 3323–3331.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.