ISSN 2186-7437

NII Shonan Meeting Report

No. 164

Differential Privacy and its Applications

Marco Gaboardi
Jun Sakuma
Thomas Steinke

October 28–November 01, 2024

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Differential Privacy and its Applications

Organizers:
Marco Gaboardi (Boston University)
Jun Sakuma (Institute of Science Tokyo)
Thomas Steinke (Google DeepMind)

October 28–November 01, 2024

NII publishes online an "NII Shonan Meeting Report" (ISSN 2186-7437) for each NII Shonan Meeting. The organizers of a meeting are requested to write a report immediately after their meeting. The items that the organizers need to include in the report include, but not limited to,

- A cover in the standard format;
- The title of the meeting, the names of the organizers, and the dates of the meeting;
- Background and introduction descriptions of the research field(s): What the workshop was about, why the workshop was held, and what they would try to solve;
- An overview of the meeting;
- A collection of the abstracts of the participants' talks;
- A list of participants.
- A brief schedule of the meeting;
- A summary of discussions;
- A summary of new findings; and
- A summary of identified issues and future directions.

The paper size needs to be A4. The organizers may use a different format, but this LaTeX style is still needed to generate a standard cover. The rest of this guide introduces how to use the standard LaTeX style.

Background and introduction

Differential privacy is a mathematically-formalized standard for the privacy-preserving analysis and use of data: it permits a broad array of data analysis methods while simultaneously offering a strong guaranteed bound on the increase in harm that any user suffers as a result of participating in a differentially private data analysis. Over the past decade, a rich literature has developed and differential privacy has slowly been accepted as a "gold standard" for data privacy. This has culminated in its adoption by government agencies like the US Census Bureau and companies like Apple, Google, and Microsoft.

The development of differential privacy has been given impetus by the increasing scale of private data being collected and used. Traditional de-identification techniques have suffered high-profile failures as they struggle to keep pace with the abundance of information available to potential attackers. The distinguishing feature of differential privacy is that it makes no assumption about the background information known to a potential attacker.

One of the reasons for the success of differential privacy is that it can be integrated into algorithmic and machine learning techniques to provide a variety of data analysis tools that are at the same time private and accurate. Furthermore, these can be composed in a modular way to develop sophisticated systems for using private data. In addition, differential privacy has also found application in areas of data analysis where privacy is not necessarily a concern, such as in ensuring fairness of algorithms and in guaranteeing statistical validity in adaptive data analysis.

Differential privacy is studied by researchers in different research areas of computer science, such as algorithms, cryptography, machine learning, programming languages, security, databases, as well as from several areas of statistics, data science, information theory, law and policy making, and social science. This broad community often work in the different areas in isolation or with limited communication between areas. We believe that for differential privacy to be successful, a conjunct effort of researchers in all these areas is needed.

The overall goal of the meeting was to foster the discussion between researchers in academia and industry that are working in different areas of differential privacy and its applications. The common ground between the different participants was the interest in the different aspects of data privacy. On the applications side, we had discussions around the different tools and assumptions that are needed to put differential privacy to work in practice. On the theoretical side instead, we had discussions around foundational issues about algorithms and models for privacy. We also explored the applicability of the most recent techniques developed in the setting of differential privacy to problems in different research areas, such as guaranteeing statistical validity in adaptive data analysis.

The meeting was an occasion to acquire, as a community of researchers interested in data privacy, a common understanding of the technical challenges and the needs for making differential privacy practical. Moreover, the meeting offered to early-career researchers an opportunity to present and discuss their work with established researchers in the field.

Overview of the meeting

The meeting was structured around invited long talks aimed at stimulate the discussion around specific topics. Besides the long talks, the program included short talks by participants, an open problems session and breakout sessions. Besides, the meeting offered several breaks and unstructured time for participants to interact and engage in impromptu discussions.

The invited long talks were the following:

- Rachel Cummings (Columbia University, USA) - Centering Policy and Practice: Research Gaps around Usable Differential Privacy
- Reza Shokri (National University of Singapore, Singapore) - Inference Attacks in Machine Learning
- Natasha Fernandes (Macquarie University, Australia) - Quantitative information flow and Differential Privacy
- Mark Bun (Boston University, USA) - Private PAC Learning: Why are you like this?
- Pasin Manurangsi (Google, Thailand) - Aggregation in Shuffle Differential Privacy: Near-Optimal Accuracy with Vanishing Communication Overhead

The breakout sessions were on the following topics:

- The Successor of Differential Privacy (how to think more in general about relaxation and other measure of privacy)
- Differential privacy for AI
- Mechanization of Differential Privacy (implementations and verification)

Overview of Talks

Private PAC Learning – Why are you like this?

Mark Bun, Boston University

This talk is an homage to differentially private classification in the PAC model. I'll survey some of the understanding we've built of the feasibility, sample complexity, and computational complexity of private PAC learning over the last ∼5 years. As the technical part of the talk, I'll attempt to give an accessible proof of the lower bound for privately learning thresholds due to Alon, Livni, Malliaris, and Moran (STOC '19), making use of the benefit of hindsight.

Centering Policy and Practice: Research Gaps around Usable Differential Privacy

Rachel Cummings, Columbia University

As a mathematically rigorous framework that has amassed a rich theoretical literature, differential privacy is considered by many experts to be the gold standard for privacy-preserving data analysis. Others argue that while differential privacy is a clean formulation in theory, it poses significant challenges in practice. Both perspectives are, in our view, valid and important. To bridge the gaps between differential privacy's promises and its real-world usability, researchers and practitioners must work together to advance policy and practice of this technology. In this paper, we outline pressing open questions towards building usable differential privacy and offer recommendations for the field, such as developing risk frameworks to align with user needs, tailoring communications for different stakeholders, modeling the impact of privacy-loss parameters, investing in effective user interfaces, and facilitating algorithmic and procedural audits of differential privacy systems.
Joint work with Jayshree Sarathy
`https://arxiv.org/abs/2406.12103`

Quantitative Information Flow and Differential Privacy

Natasha Fernandes, Macquarie University

Quantitative information flow (QIF) is an information-theoretic framework for analysing information leaks from secure systems. In its current form, QIF derives from foundational work by Geoffrey Smith [FOSSACS'09] and the g-leakage framework proposed by Alvim et al. [CSF'21] and has developed into a powerful mathematical toolkit for the analysis of probabilistic systems. QIF allows the analysis of attacks against secure systems using a Bayesian, information-theoretic adversarial model, and includes some useful tools for robust reasoning. More recently QIF has been applied to the study of differential privacy and metric differential privacy [Chatzikokolakis et al., CSF'19].

In this talk I will introduce the QIF framework and present some recent results on how local differential privacy can be interpreted with respect to Bayesian attackers. In particular, we find an interpretation for the epsilon of LDP as a capacity bound for the class of max-case and average-case attackers [Fernandes et al., CSF'24]. We also explain the non-monotonicity of utility and epsilon in terms of refinement, with an experimental example [Alvim et al., CCS'23].

Beyond Adaptivity

Katrina Ligett, The Hebrew University of Jerusalem

Recent work on the problem of overfitting due to adaptivity in data analysis has made progress on (differential privacy-adjacent) algorithmic tools that can mitigate the risks of adaptivity, formal lower bounds that illustrate bounds on what is achievable in the worst case, and models formalizing adaptivity. However (loosely), the algorithms we have are not practical, the lower bounds we have are not realistic, and the models we have model only a small part of the scientific process. This talk explores directions for future work in all three of these dimensions.

Private Everlasting Prediction

Kobbi Nissim, Georgetown University

Learning is one of the most important tasks applied to data. When applied to sensitive personal data, preserving a strong notion of privacy such as differential privacy is desirable. A rich line of study has revealed significant theoretical and practical gaps between private learning (i.e., with differential privacy) and non-private learning.

We will explore private prediction as an alternative to private learning. Instead of outputting a hypothesis, a predictor answers a stream of classification queries. Dwork and Feldman demonstrated that answering a single classification query requires a significantly lower sample complexity than private learning of the same concept class. We define private everlasting prediction (PEP), a model where an unlimited number of classification queries may be answered privately. We also present a generic PEP construction in the PAC model. The sample complexity of the initial training sample in our construction is quadratic (up to polylog factors) in the VC dimension of the concept class.

Joint work with Moni Naor (Weizmann Institute), Uri Stemmer (Tel-Aviv University), and Chao Yan (Georgetown University)

The Blahut-Arimoto mechanisms for Metric Privacy

Catuscia Palamidessi,

Metric privacy [PETS 2013] is a generalization of both central and local differential privacy and can be applied to any domain provided with a natural notion of metric. In practice, it has been mostly used in the local model, and especially in the case of location privacy, where it is known under the name of geo-indistinguishability [CCS 2013]. The standard mechanisms for metric privacy are the (multi-dimensional) Laplace and its discrete version. These mechanisms, however, have a drawback, namely, they do not protect the outliers, i.e., the data points that are isolated w.r.t. the metric.

In this talk, I will propose a solution to the problem of the outliers: the Blahut-Arimoto (BA) mechanism. This is based on an algorithm developed in rate-distortion theory, which is a particular branch of information theory. The algorithm, which is of iterative nature, approximates in the limit a channel that is Pareto optimal w.r.t. the minimization of mutual information and the distortion (the expected distance between the input and the output). A parameter β decides the balance between the two. It turns out that the channel is a metric privacy mechanism, with $\epsilon = 2\beta$. Furthermore, I will present a de-noising mechanism, the iterative Bayesian update (IBU) that experimentally is shown to work well in association with BA: assuming that the data in input to the BA are sampled i.i.d from some distribution π, IBU allows us to estimate the original distribution π from the frequencies of the noisy data in output, which is the basic notion of statistical utility. We also show a surprising result, namely that IBU and BA are dual of each other, which probably explains why IBU and BA work well together.

Based on a joint work with Sayan Biswas [PETS 2024].

Elephants Do Not Forget: Differential Privacy with State Continuity for Privacy Budget

Benjamin Rubinstein, The University of Melbourne

Current implementations of differentially-private (DP) systems either lack support to track the global privacy budget consumed on a dataset, or fail to faithfully maintain the state continuity of this budget. We show that failure to maintain a privacy budget enables an adversary to mount replay, rollback and fork attacks - obtaining answers to many more queries than what a secure system would allow. As a result the attacker can reconstruct secret data that DP aims to protect - even if DP code runs in a Trusted Execution Environment (TEE). We propose ElephantDP, a system that aims to provide the same guarantees as a trusted curator in the global DP model would, albeit set in an untrusted environment. Our system relies on a state continuity module to provide protection for the privacy budget and a TEE to faithfully execute DP code and update the budget. To provide security, our protocol makes several design choices including the content of the persistent state and the order between budget updates and query answers. We prove that ElephantDP provides liveness (i.e., the protocol can restart from a correct state and respond to queries as long as the budget is not exceeded) and DP confidentiality (i.e., an attacker learns about a dataset as much as it would from interacting with a trusted curator). Our implementation and evaluation of the protocol use Intel SGX as a TEE to run the DP code and a network of TEEs to maintain state continuity. Compared to an insecure baseline, we observe 1.1-3.2× overheads and lower relative overheads for complex DP queries.
Joint work with Jiankai Jin, Chitchanok Chuengsatiansup, Toby Murray, Yuval Yarom, Olga Ohrimenko

Analytically estimating the accuracy of differentially private ratios

Alejandro Russo, Chalmers University of Technology / DPella AB

Error bounding average calculations pose significant challenges as the noise distribution arises from the ratio between two random variables, for example, one for the noisy sum and one for the noisy count. Analytically computing this ratio's distribution is non-trivial. According to some existing DP-solutions, average calculation accuracy may be addressed by "eliminating" one random variable (e.g., see Open DP resize transformation). Since the record count is treated as constant and the noise distribution of a random variable divided by a constant is known, error estimation for the average calculation becomes straightforward. While a resourceful approach, this method relies on the human analyst's guess, which may be inaccurate due to the analyst's disconnection from the dataset and does not account for errors in guessing. We present a novel method to determines the accuracy of differentially private ratios, such as, averages or sum ratios, in the form of probabilistic error bounds by using a noisy count of the size of the dataset and a set of pre-condition checks to determine, with high probability, whether the conditions for applying error (uncertanty) propagation equations (often used in physic) are met. This introduce a new way of determining the accuracy of differentially private ratios.

3 Information Theory Views of Differential Privacy

Anand Sarwate, Rutgers, The State University of New Jersey

Differential privacy uses a particular way of quantifying the difference between two distributions that has connections to several key topics studied in information theory. In this talk I survey some of these connections and the questions they raise the breakout discussions that will follow the talk. Based on the hypothesis testing interpretation of differential privacy, several natural questions arise around the right model for privacy, the right noise distributions to consider, and the right frameworks for multiterminal systems. The (ϵ, δ) tradeoff curve for a mechanism can be written in terms of f-divergences which can be used to design noise distributions and other approaches to composition. Strong data processing inequalities and contraction coefficients for f-divergences can also give a different way of thinking about iterative methods. These three views suggest that there may be other views as well.

Differential Privacy in Isabelle/HOL

Tetsuya Sato, Institute of Science Tokyo

Formulations of differential privacy are easy to understand, but differential privacy of actual mechanisms is complicated to determine. One of the reasons for this is that small changes in database programs can break their differential privacy. Therefore, formal verification of differential privacy has been studied for over a decade.

In this talk, we introduce our very ongoing work of formalization of differential privacy (in the continuous setting) in a proof assistant, Isabelle/HOL.

Our current progress is as follows: we formalize the definition of DP and its basic composition theorems; we formalize the (ideal) Laplace mechanism; we give a formal proof of the differential privacy of the report noisy max mechanism given in the seminal textbook "Algorithmic Foundations of Differential Privacy" of Dwork and Roth.

Joint work with Yasuhiko Minamide and is based on joint works with Shin-ya Katsumata.

Inference Attacks in Machine Learning

Reza Shokri, National University Singapore

We design and develop algorithms and tools to audit privacy risks in machine learning models. We define a model's information leakage about its training data as the dependence of its prediction distribution on whether a particular data point was included in the training set. We design reconstruction attacks in a leave-one-out setting to show that models memorize specific details of some training records in such a way that it could be extracted by an adversary. [Ye, Borovykh, Hayou, Shokri 2024]

Analyzing privacy risks in leave-one-out settings is computationally expensive. So, we design membership inference attacks to test whether an adversary can detect the presence of a specific data point in a model's training set. The effectiveness of such attacks provides a lower bound on a model's vulnerability to memorization. We introduce a robust membership inference attack that outperforms prior attacks in all benchmarks, and offers a more accurate measure of a model's information leakage. [Zarifzadeh, Liu, Shokri 2024]

We also introduced the ML Privacy Meter tool (privacy-meter.com) that is an open-source privacy auditing tool based on the state-of-the-art membership inference attacks.

We discussed that membership inference attacks are powerful tests to trace model predictions to training data. However, their results need to be interpreted based on the exact indistinguishability games in which they are defined. We show that exact membership inference attacks may fail to capture privacy risks when training data or model parameters are slightly modified (e.g., using simple data transformations). Under these conditions, MIAs might fail to detect the presence of nearby data points in the training set that share sensitive information. To provide a more accurate assessment of privacy risks, we propose range membership inference attacks, which evaluate if any data point within the training set contains sensitive information about a given target. [Tao, Shokri 2024]

In the end, we discussed some challenges of analyzing and protecting privacy in large language models. For training a privacy-preserving model, according to differential privacy, it is important to have well-defined boundaries between units of data. This seems to be a challenge in text. We discuss a number of such issues in [Brown, Lee, Mireshghallah, Shokri, Tramer 2022].

We also covered the privacy expectations from language models as AI agents, where the users might blindly trust them to correctly interpret the input prompts and produce information which respects their privacy as contextual integrity. [Mireshghallah, Kim, Zhou, Tsvetkov, Sap, Shokri, Choi, 2024].

Privately Evaluating Untrusted Black-Box Functions

Adam Smith, Boston University

We provide tools for sharing sensitive data in situations when the data curator does not know in advance what questions an (untrusted) analyst might want to ask about the data. The analyst can specify a program that they want the curator to run on the dataset. We model the program as a black-box function f. We study differentially private algorithms, called privacy wrappers, that, given black-box access to a real-valued function f and a sensitive dataset x, output an accurate approximation to f(x). The dataset x is modeled as a finite subset of a possibly infinite set U, in which each entry x represents data of one individual. A privacy wrapper calls f on the dataset x and on some subsets of x and returns either an approximation to f(x) or a nonresponse symbol. The wrapper may also use additional information (that is, parameters) provided by the analyst, but differential privacy is required for all values of these parameters. Correct setting of these parameters will ensure better accuracy of the privacy wrapper. The bottleneck in the running time of our privacy wrappers is the number of calls to f, which we refer to as queries. Our goal is to design privacy wrappers with high accuracy and small query complexity.

We introduce a novel setting, called the automated sensitivity detection setting, where the analyst supplies only the black-box function f and the intended (finite) range of f. In contrast, in the previously considered setting, which we refer to as the claimed sensitivity bound setting, the analyst also supplies additional parameters that describe the sensitivity of f. We design privacy wrappers for both settings and show that our wrappers are nearly optimal in terms of accuracy and locality (i.e., the depth of the local neighborhood of the dataset x they explore, which is a proxy for query complexity). In the claimed sensitivity bound setting, we provide the first accuracy guarantees that have no dependence on the size of the universe U. We also re-interpret and analyze previous constructions in our framework, and use them as comparison points. In addition to addressing the black-box privacy problem, our private mechanisms provide feasibility results for differentially private release of general classes of functions.

Joint work with Ephraim Linder, Sofya Raskhodnikova, and Thomas Steinke.

Auditing Privacy Mechanisms via Label Inference Attacks

Marika Swanberg, Boston University

We propose reconstruction advantage measures to audit label privatization mechanisms. A reconstruction advantage measure quantifies the increase in an attacker's ability to infer the true label of an unlabeled example when provided with a private version of the labels in a dataset (e.g., aggregate of labels from different users or noisy labels output by randomized response), compared to an attacker that only observes the feature vectors, but may have prior knowledge of the correlation between features and labels. We consider two such auditing measures: one additive, and one multiplicative. These incorporate previous approaches taken in the literature on empirical auditing and differential privacy. The measures allow us to place a variety of proposed privatization schemes—some differentially private, some not—on the same footing. We analyze these measures theoretically under a distributional model which encapsulates reasonable adversarial settings. We also quantify their behavior empirically on real and simulated prediction tasks. Across a range of experimental settings, we find that differentially private schemes dominate or match the privacy-utility tradeoff of more heuristic approaches.

Instance-Optimal Private Density Estimation in the Wasserstein Distance

Kunal Talwar, Apple

Estimating the density of a distribution from samples is a fundamental problem in statistics. In many practical settings, the Wasserstein distance is an appropriate error metric for density estimation. For example, when estimating population densities in a geographic region, a small Wasserstein distance means that the estimate is able to capture roughly where the population mass is. In this work we study differentially private density estimation in the Wasserstein distance. We design and analyze instance-optimal algorithms for this problem that can adapt to easy instances.

For distributions P over $\mathbb{R}$, we consider a strong notion of instance-optimality: an algorithm that uniformly achieves the instance-optimal estimation rate is competitive with an algorithm that is told that the distribution is either P or Q_P for some distribution Q_P whose probability density function (pdf) is within a factor of 2 of the pdf of P. For distributions over $\mathbb{R}^2$, we use a different notion of instance optimality. We say that an algorithm is instance-optimal if it is competitive with an algorithm that is given a constant-factor multiplicative approximation of the density of the distribution. We characterize the instance-optimal estimation rates in both these settings and show that they are uniformly achievable (up to polylogarithmic factors). Our approach for $\mathbb{R}^2$ extends to arbitrary metric spaces as it goes via hierarchically separated trees. As a special case our results lead to instance-optimal private learning in TV distance for discrete distributions.

Joint work with Vitaly Feldman, Audra McMillan and Satchit Sivakumar.

A Framework for Differential Privacy Against Timing Attacks

Salil Vadhan, Harvard University / OpenDP

The standard definition of differential privacy (DP) ensures that a mechanism's output distribution on adjacent datasets is indistinguishable. However, real-world implementations of DP can, and often do, reveal information through their runtime distributions, making them susceptible to timing attacks. In this work, we establish a general framework for ensuring differential privacy in the presence of timing side channels. We define a new notion of timing privacy, which captures programs that remain differentially private to an adversary that observes the program's runtime in addition to the output. Our framework enables chaining together component programs that are timing-stable followed by a random delay to obtain DP programs that achieve timing privacy. Importantly, our definitions allow for measuring timing privacy and output privacy using different privacy measures. We illustrate how to instantiate our framework by giving programs for standard DP computations in the RAM and Word RAM models of computation. Furthermore, we show how our framework can be realized in code through a natural extension of the OpenDP Programming Framework. Lastly, we mention some preliminary positive results on the possibility of achieving pure differential privacy in the unbounded setting.

Joint work with Zachary Ratliff [paper].

List of Participants

- Hiromi Arai, RIKEN
- Borja de Balle Pigem, Google DeepMind
- Mark Bun, Boston University
- Yang Cao, Institute of Science Tokyo
- Rachel Cummings, Columbia University
- Vitaly Feldman, Apple
- Natasha Fernandes, Macquarie University
- Kazuto Fukuchi, University of Tsukuba
- Marco Gaboardi, Boston University
- Shin-ya Katsumata, Kyoto Sangyo University
- Katrina Ligett, Hebrew University
- Pasin Manurangsi, Google Research
- Annabelle McIver, Macquarie University
- Carroll Morgan, University of New South Wales
- Takao Murakami, The Institute of Statistical Mathematics (ISM)
- Kobbi Nissim, Georgetown University
- Catuscia Palamidessi, Inria
- Guy Rothblum, Apple
- Benjamin Rubinstein, University of Melbourne
- Alejandro Russo, Chalmers University of Technology / DPella AB
- Jun Sakuma, Institute of Science Tokyo
- Anand Sarwate, Rutgers University
- Tetsuya Sato, Institute of Science Tokyo
- Reza Shokri, National University of Singapore
- Adam Smith, Boston University
- Jessica Sorrell, Johns Hopkins University
- Marika Swanberg, Boston University
- Kunal Talwar, Apple
- Hiroshi Unno, Tohoku University
- Salil Vadhan, Harvard University / OpenDP
- Xiaokui Xiao, National University of Singapore

Meeting Schedule

Check-in Day: October 27 (Sun)

19:00-21:00 Welcome Banquet

Day1: October 28 (Mon)

07:30 – 09:00 Breakfast
09:00 – 09:30 Welcome and Introductions
09:30 – 10:30 Rachel Cummings - Centering Policy and Practice: Research Gaps around Usable Differential Privacy
10:30 – 11:00 Break
11:00 – 12:00 2 short talks
Kunal Talwar - Instance-optimal statistical estimation
Katrina Ligett - Thinking about adaptivity
12:00 – 14:00 Lunch
14:00 – 15:30 3 short talks
Kobbi Nissim - Private Everlasting Prediction
Adam Smith - Privately Evaluating Untrusted Black-Box Functions
Guy Rothblum - Verifiable Data Science
15:30 – 16:00 Break
16:00 – 17:00 Open Problem Session
18:00 – 19:30 Dinner followed by Social Activities

Day2: October 29 (Tue)

07:30 – 09:00 Breakfast
09:00 – 10:00 Reza Shokri - Inference Attacks in Machine Learning
10:00 – 10:30 Break
10:30 – 12:00 2 short talks
Salil Vadhan - A Framework for Differential Privacy Against Timing Attacks
Catuscia Palamidessi - The Blahut-Arimoto mechanism for metric privacy
12:00 – 14:00 Lunch
14:00 – 14:30 1 short talk
Ben Rubinstein: Elephants Do Not Forget: Differential Privacy with State Continuity for Privacy Budget
14:30 – 15:30 Break-out group discussions
The Successor of Differential Privacy (how to think more in general about relaxation and other measure of privacy)
Differential privacy for AI
15:30 – 16:00 Break
16:00 – 16:30 Group Discussions Summary Meeting (group presentations)
16:30 – 17:00 Marika Swanberg: Auditing Privacy Mechanisms via Label Inference Attacks
18:00 – 19:30 Dinner followed by Social Activities

Day3: October 30 (Wed)

07:30 – 09:00 Breakfast
09:00 – 10:00 Natasha Fernandes - Quantitative information flow and Differential Privacy

10:00 – 10:30 Break
10:30 – 12:00 3 short talks
Shin-ya Katsumata - Divergences on Computational Effects
Tetsuya Sato - Differential Privacy in Isabelle/HOL
Yang Cao - Towards Explaining, Choosing, and Negotiating Parameters of Differential Privacy
12:00 – 13:30 Lunch
13:30 – 20:45 Excursion to Kamakura and Dinner

Day4: October 31 (Thu)

07:30 – 09:00 Breakfast
09:00 – 10:00 Mark Bun - Private PAC Learning: Why are you like this?
10:00 – 10:30 Break
10:30 – 12:00 2 short talks
Jessica Sorrell - Separating DP and Replicability
Alejandro Russo: Analytically estimating the accuracy of differentially private ratios
12:00 Photo session
12:10 – 14:30 Lunch
14:30 – 15:00 1 short talk
Anand Sarwate - 3 Information Theory Views of Differential Privacy
15:00 – 16:00 Break-out group discussions
Mechanization of Differential Privacy (implementations and verification)
16:00 – 16:30 Break
16:30 – 17:00 Continuing group discussions and Summary Meeting (group presentations)
18:00 – 19:30 Dinner followed by Social Activities

Day5: November 1 (Fri)

07:30 – 09:00 Breakfast
09:00 – 10:00 Pasin Manurangsi - Aggregation in Shuffle Differential Privacy: Near-Optimal Accuracy with Vanishing Communication Overhead
10:00 – 10:30 Break
10:30 – 11:30 3 short talks
Vitaly Feldman - Trade-offs in Data Memorization: Learn more, remember less
Takao Murakami - Augmented Shuffle Protocols for Accurate and Robust Frequency Estimation under Differential Privacy
Xiaokui Xiao - Synthesizing Relational Data with Differential Privacy
12:00 – 12:05 Closing Discussions
12:05 – 13:30 Lunch

Summary of breakout discussions

Differential Privacy and AI

Privacy Attacks on AI We discussed mainly the following question "Can LLMs help with record linkage/finding auxiliary information for privacy attacks?". We agreed that AI is going too fast for privacy experts to keep up—maybe attacks can demonstrate/motivate the need for better privacy considerations in this area.

Privacy of language models We discussed about fine tuning and about the fact that language models are stitching together information about individuals in a way that is problematic. DP can be a good notion for fine tuning in some situations (e.g. when one can identify the privacy unit) but not in others (e.g. where it is difficult to identify the privacy unit).

What is the next privacy definition?

What we would like to preserve: compositionality, and protection for inference about the individual, rather than the population.

What we would like to change: we would like to relax the worst case guarantee (problem: relaxations may break compositionality).

What we would like to add: 1) privacy risks as part of the communication of the strength of an algorithm, 2) whether the definition aligns with legislations, if meaningful, 3) a more general framework able to describe privacy risks.

Mechanization of Differential Privacy (implementations and verification)

The current landscape for verifying Differential Privacy (DP) algorithms is fragmented, with no single reference tool or system readily adoptable by the community. One way to enhance the assurance of DP implementations with minimal friction is to develop a comprehensive test suite that identifies common vulnerabilities, such as floating-point issues, overflow, and underflow in discrete mechanisms. Such a test suite would focus solely on the input and output behavior of DP implementations, making it widely applicable across different languages and systems. In contrast, formal verification tends to be language-dependent. Therefore, testing serves as a natural first step before advancing into verification. Verification should be viewed as providing different levels of assurance—from mechanically verifying DP algorithms to fully verifying the implementation code itself.

Figure 1: Participants during the excursion

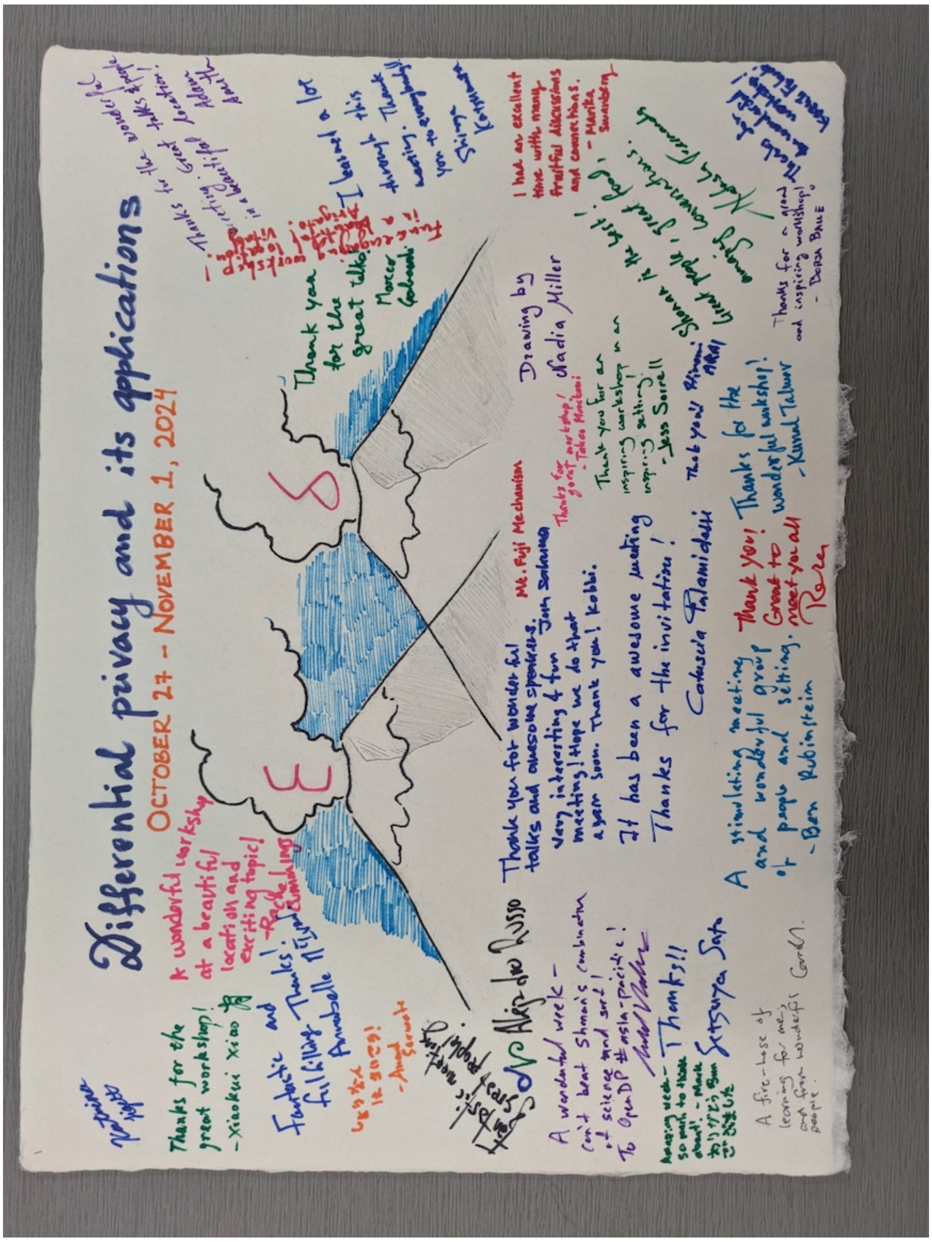

Figure 2: Poster of Seminar no. 164 “Differential Privacy and Applications”.