

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 13-01-2023		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 17-Jul-2018 - 5-Sep-2021	
4. TITLE AND SUBTITLE Final Report: Developing a Science for Statistical ORAMs			5a. CONTRACT NUMBER W911NF-18-1-0312		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 611102		
6. AUTHORS			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Princeton University The Trustees of Princeton University 1 Nassau Hall Princeton, NJ 08544 -2001			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 70729-NC-YIP.12		
12. DISTRIBUTION AVAILABILITY STATEMENT Distribution Approved for public release; distribution is unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			Prateek Mittal
					19b. TELEPHONE NUMBER 609-258-0814

RPPR Final Report

as of 25-Jan-2023

Agency Code: 21XD

Proposal Number: 70729NCYIP
INVESTIGATOR(S):

Agreement Number: W911NF-18-1-0312

Name: Prateek Mittal
Email: pmittal@princeton.edu
Phone Number: 6092580814
Principal: Y

Organization: **Princeton University**

Address: The Trustees of Princeton University, Princeton, NJ 085442001

Country: USA

DUNS Number: 002484665

EIN: 210634501

Report Date: 05-Dec-2021

Date Received: 13-Jan-2023

Final Report for Period Beginning 17-Jul-2018 and Ending 05-Sep-2021

Title: Developing a Science for Statistical ORAMs

Begin Performance Period: 17-Jul-2018

End Performance Period: 05-Sep-2021

Report Term: 0-Other

Submitted By: Prateek Mittal

Email: pmittal@princeton.edu

Phone: (609) 258-0814

Distribution Statement:

STEM Degrees:

STEM Participants:

Major Goals: In this proposal, we aim to develop a scientific foundation for protecting the security/privacy of programs memory access patterns. We note that the lack of progress in low-bandwidth ORAMs, despite considerable efforts from the security community, is an undeniable indicator that we need a fundamentally new approach for the problem. Towards this end, we propose to trade-off reduction in bandwidth at the cost of statistical privacy guarantees (such as differential privacy) in oblivious RAMs. The intersection of differential privacy and oblivious RAMs opens up a new design space for our research community, and has the potential to reduce bandwidth overheads by orders of magnitude, including a constant overhead design point. In this context, we will investigate the following research challenges:

****Statistical Privacy for ORAMs.**** We will investigate metrics from the domain of statistical data privacy, that can rigorously quantify privacy leakage from memory access patterns. In particular, we will explore and formalize the notion of a differentially private ORAM that provides statistical privacy guarantees, and which to the extent of our knowledge, is the first of its kind. We will extend the differential privacy framework developed by Dwork et al. with its epsilon, delta differential privacy modification. In the current formulation of an ORAM (perfect security), the output is computationally indistinguishable for any two input sequences. In a differentially private ORAM, we will characterize the effect of a small change in the ORAM input to the change in the probability distribution at the output.

- ****Scientific Foundation for Security-Performance Trade-offs.**** We will explore a new scientific theory, along with the design of associated algorithms, that can achieve desired tradeoffs between ORAM bandwidth and ORAM security. Our theory will allow ORAM protocols to be tailored as per the needs and constraints of the user application, serving as an enabler for practical deployment. As a novel design point, our theory will investigate ORAM mechanisms with only a constant bandwidth overhead (providing orders of magnitude reduction compared to state-of-the-art).

- ****Theoretical and Systems Evaluation.**** We will evaluate the security and performance of our approach both theoretically and using a real-world prototype implementation. From a theoretical perspective, we will perform a rigorous security analysis of our proposed construction, and compute the fundamental limits of privacy-performance tradeoffs enabled by our overall approach. From a systems perspective, we will implement our approach in a real-world prototype and experimentally analyze its performance.

Accomplishments: **** Machine Learning with Differentially Private Labels: Mechanisms and Frameworks (PETS 2022)**.** Label differential privacy is a relaxation of differential privacy for machine learning scenarios where the

RPPR Final Report as of 25-Jan-2023

labels are the only sensitive information that need to be protected in the training data. For example, imagine a survey from a participant in a university class about their vaccination status. Some attributes of the students are publicly available but their vaccination status is sensitive information and must remain private. Now if we want to train a model that predicts whether a student has received vaccination using only their public information, we can use label-DP. Recent works on label-DP use different ways of adding noise to the labels in order to obtain label-DP models. In this work, we present novel techniques for training models with label-DP guarantees by leveraging unsupervised learning and semi-supervised learning, enabling us to inject less noise while obtaining the same privacy, therefore achieving a better utility-privacy trade-off. Our experiments show that these techniques can significantly outperform the prior work.

**** Mitigating Membership Inference Attacks by Self-Distillation Through a Novel Ensemble Architecture (USENIX Security 2022)**** Membership inference attacks are a key measure to evaluate privacy leakage in machine learning (ML) models. It is important to train ML models that have high membership privacy while largely preserving their utility. In this work, we propose a new framework to train privacy-preserving models that induce similar behavior on member and non-member inputs to mitigate membership inference attacks.

**** Athena: Probabilistic Verification of Machine Unlearning (PETS 2022)**** The right to be forgotten, also known as the right to erasure, is the right of individuals to have their data erased from an entity storing it. The status of this long held notion was legally solidified recently by the General Data Protection Regulation (GDPR) in the European Union. As a consequence, there is a need for mechanisms whereby users can verify if service providers comply with their deletion requests. In this work, we take the first step in proposing a formal framework, called Athena, to study the design of such verification mechanisms for data deletion requests – also known as machine unlearning – in the context of systems that provide machine learning as a service (MLaaS).

****DP-Cryptography: Marrying Differential Privacy and Cryptography in Emerging Applications (CACM 2021).**** Differential privacy (DP) has arisen as the state-of-the-art metric for quantifying individual privacy when sensitive data are analyzed, and it is starting to see practical deployment in organizations such as the US Census Bureau, Apple, Google, etc. There are two popular models for deploying differential privacy - standard differential privacy (SDP), where a trusted server aggregates all the data and runs the DP mechanisms, and local differential privacy (LDP), where each user perturbs their own data and perturbed data is analyzed. Due to security concerns arising from aggregating raw data at a single server, several real world deployments in industry have embraced the LDP model. However, systems based on the LDP model tend to have poor utility - "a gap" in the utility achieved as compared to systems based on the SDP model. In this work, we survey and synthesize emerging directions of research at the intersection of differential privacy and cryptography. First, we survey solutions that combine cryptographic primitives like secure computation and anonymous communication with differential privacy to give alternatives to the LDP model that avoid a trusted server as in SDP but close the gap in accuracy. These primitives introduce performance bottlenecks and necessitate efficient alternatives. Second, we synthesize work in an area we call "DP-Cryptography" - cryptographic primitives that are allowed to leak differentially private outputs. These primitives have orders of magnitude better performance than standard cryptographic primitives. DP-cryptographic primitives are perfectly suited for implementing alternatives to LDP, but are also applicable to scenarios where standard cryptographic primitives do not have practical implementations. Through this unique lens of research taxonomy, we survey ongoing research in these directions while also providing novel directions for future research.

****FALCON: Honest-Majority Maliciously Secure Framework for Private Deep Learning (PETS 2021)**** This paper aims to enable training and inference of neural networks in a manner that protects the privacy of sensitive data. We propose FALCON - an end-to-end 3-party protocol for fast and secure computation of deep learning algorithms on large networks. FALCON presents three main advantages. It is highly expressive. To the best of our knowledge, it is the first secure framework to support high capacity networks with over a hundred million parameters such as VGG16 as well as the first to support batch normalization, a critical component of deep learning that enables training of complex network architectures such as AlexNet. Next, FALCON guarantees security with abort against malicious adversaries, assuming an honest majority. It ensures that the protocol always completes with correct output for honest participants or aborts when it detects the presence of a malicious adversary. Lastly, FALCON presents new theoretical insights for protocol design that make it highly efficient and allow it to outperform existing secure deep learning solutions. Compared to prior art for private inference, we are about 8x faster than SecureNN (PETS '19) on average and comparable to ABY3 (CCS '18). We are about 16-200x more communication efficient than either of

RPPR Final Report as of 25-Jan-2023

these. For private training, we are about 6x faster than SecureNN, 4.4x faster than ABY3 and about 2-60x more communication efficient. This is the first paper to show via experiments in the WAN setting, that for multi-party machine learning computations over large networks and datasets, compute operations dominate the overall latency, as opposed to the communication.

Training Opportunities: The project has provided training opportunities for Sameer Wagh, who graduated from Princeton and joined UC Berkeley as a postdoctoral scholar, Daniel Cullina, who joined PSU as a tenure-track faculty, and Changchang Liu, a female PhD student who graduated from Princeton and joined IBM Research. Additional PhD students on the project include Xinyu Tang. The project has also provided training opportunities for undergraduate students, including under-represented minorities such as Kavya Chandran and Tinashe Handina.

Results Dissemination: Our work has been disseminated to venues such as ACM Conference on Computer and Communications Security (ACM CCS 2019), Communications of the ACM (CACM 2021), and the Privacy Enhancing Technologies Symposium (PETS 2021, 2022), and USENIX Security 2022.

Honors and Awards: Our work on differentially private cryptography was featured in Communications of the ACM, which is the flagship magazine of the ACM computing society. Our FALCON paper was the recipient of the Facebook Systems for Machine Learning Award, and the Qualcomm Innovation Award. **More broadly, PI Mittal was the recipient of the Casper Bowden Runnerup Award for Outstanding Research in Privacy Enhancing Technologies for three years in a row (2020, 2021, and 2022). His USENIX Security 2018 and USENIX Security 2021 papers on securing Internet domain validation impacted the real-world deployment of Let's Encrypt, and resulting in secure issuance of over 1 billion TLS certificates. **

Protocol Activity Status:

Technology Transfer: Our work on membership inference attacks (USENIX Security 2021) was integrated in Google's TensorFlow deployment. More broadly, the PI's work on security of domain validation (USENIX Security 2018, USENIX Security 2021) was deployed at Let's Encrypt, the world's largest CA and led to secure issuance of over 1 billion TLS certificates.

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Proceedings on Privacy Enhancing Technologies

Publication Identifier Type: DOI

Publication Identifier: 10.1515/popets-2018-0032

Volume: 2018

Issue: 4

First Page #: 64

Date Submitted: 10/3/19 12:00AM

Date Published: 10/1/18 8:00AM

Publication Location:

Article Title: Differentially Private Oblivious RAM

Authors: Sameer Wagh, Paul Cuff, Prateek Mittal

Keywords: Oblivious RAM, Differential Privacy

Abstract: In this work, we investigate if statistical privacy can enhance the performance of ORAM mechanisms while providing rigorous privacy guarantees. We propose a formal and rigorous framework for developing ORAM protocols with statistical security viz., a differentially-private ORAM (DP-ORAM). We present Root ORAM, a family of DP-ORAMs that provide a tunable, multi-dimensional trade-offs between the desired band-width overhead, local storage and system security. We theoretically analyze Root ORAM to quantify both its security and performance. We experimentally demonstrate the benefits of Root ORAM.

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

RPPR Final Report as of 25-Jan-2023

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Proceedings on Privacy Enhancing Technologies

Publication Identifier Type: DOI

Publication Identifier: 10.2478/popets-2019-0045

Volume: 2019

Issue: 3

First Page #: 233

Date Submitted: 10/3/19 12:00AM

Date Published: 7/1/19 4:00AM

Publication Location:

Article Title: Investigating Statistical Privacy Frameworks from the Perspective of Hypothesis Testing

Authors: Changchang Liu, Xi He, Thee Chanyaswad, Shiqiang Wang, Prateek Mittal

Keywords: differential privacy; hypothesis testing

Abstract: Over the last decade, differential privacy (DP) has emerged as the gold standard of a rigorous and prov- able privacy framework. However, there are very few practical guidelines on how to apply differential privacy in practice, and a key challenge is how to set an appropriate value for the privacy parameter? In this work, we employ a statistical tool called hypothesis testing for discovering useful and interpretable guidelines for the state-of-the-art privacy-preserving frameworks. We formalize and implement hypothesis testing in terms of an adversary's capability to infer mutually exclusive sensitive information about the input data (such as whether an individual has participated or not) from the output of the privacy-preserving mechanism. We quantify the success of the hypothesis testing using the precision- recall-relation, which provides an interpretable and natural guideline for practitioners and researchers on selecting epsilon.

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Proceedings on Privacy Enhancing Technologies

Publication Identifier Type: DOI

Publication Identifier: 10.2478/popets-2019-0003

Volume: 2019

Issue: 1

First Page #: 26

Date Submitted: 10/3/19 12:00AM

Date Published: 1/1/19 10:00AM

Publication Location:

Article Title: RON-Gauss: Enhancing Utility in Non-Interactive Private Data Release

Authors: Thee Chanyaswad, Changchang Liu, Prateek Mittal

Keywords: differential privacy, non-interactive private data release, random orthonormal projection

Abstract: A key challenge facing the design of differential privacy in the non-interactive setting is to maintain the utility of the released data. To overcome this challenge, we utilize the Diaconis-Freedman-Meckes (DFM) effect, which states that most projections of high-dimensional data are nearly Gaussian. Hence, we propose the RON-Gauss model that leverages the novel combination of dimensionality reduction via random orthonormal (RON) projection and the Gaussian generative model for synthesizing differentially-private data. We analyze how RON-Gauss benefits from the DFM effect, and present multiple algorithms for a range of machine learning applications, including both unsupervised and supervised learning. Furthermore, we rigorously prove that (a) our algorithms satisfy the strong ϵ -differential privacy guarantee, and (b) RON projection can lower the level of perturbation required for differential privacy.

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

RPPR Final Report as of 25-Jan-2023

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Proceedings on Privacy Enhancing Technologies

Publication Identifier Type: DOI

Publication Identifier: 10.2478/popets-2019-0025

Volume: 2019

Issue: 2

First Page #: 166

Date Submitted: 10/3/19 12:00AM

Date Published: 4/1/19 4:00AM

Publication Location:

Article Title: DPSelect: A Differential Privacy Based Guard Relay Selection Algorithm for Tor

Authors: Hans Hanley, Yixin Sun, Sameer Wagh, Prateek Mittal

Keywords: Differential Privacy, Max-Divergence, BGP Hijack Attacks

Abstract: In this work, we make the following contributions. First, we identify a novel means to quantify privacy leakage in guard selection algorithms using the metric of Max-Divergence. Max-Divergence ensures that probabilistic privacy loss is within strict bounds while also providing composability over time. Second, we utilize Max-Divergence and multiple notions of entropy to understand privacy loss in the worst-case for Counter-RAPTOR. Our worst-case analysis provides a fresh perspective to the field, as prior work such as Counter-RAPTOR only analyzed average case-privacy loss. Third, we propose modifications to Counter-RAPTOR that incorporate worst-case Max-Divergence in its design.

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Proceedings on Privacy Enhancing Technologies

Publication Identifier Type: DOI

Publication Identifier: 10.2478/popets-2019-0035

Volume: 2019

Issue: 3

First Page #: 26

Date Submitted: 10/3/19 12:00AM

Date Published: 7/1/19 4:00AM

Publication Location:

Article Title: SecureNN: 3-Party Secure Computation for Neural Network Training

Authors: Sameer Wagh, Divya Gupta, Nishanth Chandran

Keywords: Secure Multi-Party Computation, Privacy-preserving deep learning

Abstract: Neural Networks (NN) provide a powerful method for machine learning training and inference. To effectively train, it is desirable for multiple parties to combine their data – however, doing so conflicts with data privacy. In this work, we provide novel three-party secure computation protocols for various NN building blocks such as matrix multiplication, convolutions, Rectified Linear Units, Maxpool, normalization and so on. This enables us to construct three-party secure protocols for training and inference of several NN architectures such that no single party learns any information about the data. Experimentally, we implement our system over Amazon EC2 servers in different settings.

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

RPPR Final Report as of 25-Jan-2023

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Privacy Enhancing Technologies Symposium, 2021

Publication Identifier Type: DOI

Publication Identifier: 10.2478/popets-2021-0011

Volume: Issue: 1

First Page #:

Date Submitted: 1/13/23 12:00AM

Date Published:

Publication Location:

Article Title: FALCON: Honest-Majority Maliciously Secure Framework for Private Deep Learning

Authors: Sameer Wagh, Shruti Tople, Fabrice Benhamouda, Eyal Kushilevitz, Prateek Mittal, Tal Rabin

Keywords: privacy-preserving ML

Abstract: This paper aims to enable training and inference of neural networks in a manner that protects the privacy of sensitive data. We propose FA L C O N – an end-to-end 3-party protocol for fast and secure computation of deep learning algorithms on large networks. FA L C O N presents three main advantages. It is highly expressive. To the best of our knowledge, it is the first secure framework to support high capacity networks with over a hundred million parameters such as VGG16 as well as the first to support batch normalization, a critical component of deep learning that enables training of complex network architectures such as AlexNet. Next, FA L C O N guarantees security with abort against malicious adversaries, assuming an honest majority. It ensures that the protocol always completes with correct output for honest participants or aborts when it detects the presence of a malicious adversary.

Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors

Acknowledged Federal Support: Y

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: ACM Conference on Computer and Communications Security

Publication Identifier Type:

Publication Identifier:

Volume: Issue:

First Page #:

Date Submitted: 8/3/20 12:00AM

Date Published: 11/11/19 5:00AM

Publication Location:

Article Title: Privacy Risks of Securing Machine Learning Models against Adversarial Examples

Authors: Liwei Song, Reza Shokri, Prateek Mittal

Keywords: privacy, machine learning

Abstract: The arms race between attacks and defenses for machine learning models has come to a forefront in recent years, in both the security community and the privacy community. However, one big limitation of previous research is that the security domain and the privacy domain have typically been considered separately. It is thus unclear whether the defense methods in one domain will have any unexpected impact on the other domain. In this paper, we take a step towards resolving this limitation by combining the two domains. In particular, we measure the success of membership inference attacks against six state-of-the-art defense methods that mitigate the risk of adversarial examples (i.e., evasion attacks). Membership inference attacks determine whether or not an individual data record has been part of a model's training set. The accuracy of such attacks reflects the information leakage of training algorithms about individual members of the training set.

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

RPPR Final Report as of 25-Jan-2023

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Communications of the ACM

Publication Identifier Type: DOI

Publication Identifier: 10.1145/3418290

Volume: Issue:

First Page #:

Date Submitted: 1/13/23 12:00AM

Date Published:

Publication Location:

Article Title: DP-Cryptography: Marrying Differential Privacy and Cryptography in Emerging Applications

Authors: Sameer Wagh, Xi He, Ashwin Machanavajhala, Prateek Mittal

Keywords: statistical privacy, cryptography

Abstract: Differential privacy (DP) has arisen as the state-of-the-art metric for quantifying individual privacy when sensitive data are analyzed, and it is starting to see practical deployment in organizations such as the US Census Bureau, Apple, Google, etc. There are two popular models for deploying differential privacy - standard differential privacy (SDP), where a trusted server aggregates all the data and runs the DP mechanisms, and local differential privacy (LDP), where each user perturbs their own data and perturbed data is analyzed. Due to security concerns arising from aggregating raw data at a single server, several real world deployments in industry have embraced the LDP model. However, systems based on the LDP model tend to have poor utility - "a gap" in the utility achieved as compared to systems based on the SDP model. In this work, we survey and synthesize emerging directions of research at the intersection of differential privacy and cryptography.

Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: Y

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: USENIX Security Symposium

Publication Identifier Type:

Publication Identifier:

Volume: Issue:

First Page #:

Date Submitted: 1/13/23 12:00AM

Date Published: 8/1/22 4:00AM

Publication Location:

Article Title: Mitigating Membership Inference Attacks by Self-Distillation Through a Novel Ensemble Architecture

Authors: Xinyu Tang, Saeed Mahloujifar, Liwei Song, Virat Shejwalkar, Milad Nasr, Amir Houmansadr, Prateek M

Keywords: Membership Inference

Abstract: Membership inference attacks are a key measure to evaluate privacy leakage in machine learning (ML) models. It is important to train ML models that have high membership privacy while largely preserving their utility. In this work, we propose a new framework to train privacy-preserving models that induce similar behavior on member and non-member inputs to mitigate membership inference attacks. Our framework, called SELENA, has two major components. The first component and the core of our defense is a novel ensemble architecture for training. This architecture, which we call Split-AI, splits the training data into random subsets, and trains a model on each subset of the data. We use an adaptive inference strategy at test time: our ensemble architecture aggregates the outputs of only those models that did not contain the input sample in their training data.

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

RPPR Final Report as of 25-Jan-2023

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Privacy Enhancing Technologies Symposium
Publication Identifier Type: **Publication Identifier:**
Volume: **Issue:** **First Page #:**
Date Submitted: 1/13/23 12:00AM **Date Published:** 7/1/22 4:00AM
Publication Location:
Article Title: Machine Learning with Differentially Private Labels: Mechanisms and Frameworks
Authors: Xinyu Tang, Milad Nasr, Saeed Mahloujifar, Virat Shejwalkar, Liwei Song, Amir Houmansadr, Prateek M
Keywords: Label Differential Privacy
Abstract: Label differential privacy is a relaxation of differential privacy for machine learning scenarios where the labels are the only sensitive information that need to be protected in the training data. For example, imagine a survey from a participant in a university class about their vaccination status. Some attributes of the students are publicly available but their vaccination status is sensitive information and must remain private. Now if we want to train a model that predicts whether a student has received vaccination using only their public information, we can use label-DP. Recent works on label-DP use different ways of adding noise to the labels in order to obtain label-DP models. In this work, we present novel techniques for training models with label-DP guarantees by leveraging unsupervised learning and semi-supervised learning, enabling us to inject less noise while obtaining the same privacy, therefore achieving a better utility-privacy trade-off.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: Y

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Privacy Enhancing Technologies Symposium
Publication Identifier Type: DOI **Publication Identifier:** 10.56553/popets-2022-0072
Volume: **Issue:** **First Page #:**
Date Submitted: 1/13/23 12:00AM **Date Published:** 7/1/22 4:00AM
Publication Location:
Article Title: Athena: Probabilistic Verification of Machine Unlearning
Authors: David M. Sommer, Liwei Song, Sameer Wagh, Prateek Mittal
Keywords: Machine Unlearning
Abstract: The right to be forgotten, also known as the right to erasure, is the right of individuals to have their data erased from an entity storing it. The status of this long held notion was legally solidified recently by the General Data Protection Regulation (GDPR) in the European Union. As a consequence, there is a need for mechanisms whereby users can verify if service providers comply with their deletion requests. In this work, we take the first step in proposing a formal framework, called Athena, to study the design of such verification mechanisms for data deletion requests— also known as machine unlearning – in the context of systems that provide machine learning as a service (MLaaS).
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: Y

RPPR Final Report

as of 25-Jan-2023

Partners

,

I certify that the information in the report is complete and accurate:

Signature: Prateek Mittal

Signature Date: 1/13/23 1:22PM

Synthesizing the emerging directions of research at the intersection of differential privacy and cryptography.

BY SAMEER WAGH, XI HE, ASHWIN MACHANAVAJJHALA, AND PRATEEK MITTAL

DP-Cryptography: Marrying Differential Privacy and Cryptography in Emerging Applications

ON FEB 15, 2019, John Abowd, chief scientist at the U.S. Census Bureau, announced the results of a *reconstruction attack* that they proactively launched using data released under the 2010 Decennial Census.¹⁹ The decennial census released billions of statistics about individuals like “how many people of the age 10–20 live in New York City” or “how many people live in four-person households.” Using only the data publicly released in 2010, an internal team was able to correctly reconstruct records of address (by census block), age, gender, race, and ethnicity for 142 million

people (about 46% of the U.S. population), and correctly match these data to commercial datasets circa 2010 to associate personal-identifying information such as names for 52 million people (17% of the population).

This is not specific to the U.S. Census Bureau—such attacks can occur in any setting where statistical information in the form of deidentified data, statistics, or even machine learning models are released. That such attacks are possible was predicted over 15 years ago by a seminal paper by Irit Dinur and Kobbi Nissim¹²—releasing a sufficiently large number of aggregate statistics with sufficiently high accuracy provides sufficient information to reconstruct the underlying database with high accuracy. The practicality of such a large-scale reconstruction by the U.S. Census Bureau underscores the grand challenge that public organizations, industry, and scientific research faces: How can we safely disseminate results of data analysis on sensitive databases?

An emerging answer is *differential privacy*. An algorithm satisfies differential privacy (DP) if its output is insensitive to adding, removing or changing one record in its input database. DP is considered the “gold standard” for privacy for a number of reasons. It provides a persuasive mathematical proof of privacy to individuals with several rigorous interpretations.^{25,26} The DP guarantee is composable and repeating

» key insights

- **Local Differential Privacy is increasingly being embraced as the primary model of deployment of differential privacy, albeit at a heavy accuracy cost.**
- **Cryptographic primitives can help bridge the utility gap between systems deployed in the local differential privacy model and standard differential privacy model, but the increased utility may come at the cost of performance.**
- **DP-cryptographic primitives, which are relaxed notions of cryptographic primitives that leak differentially private outputs, permit implementations that are orders of magnitude faster than the regular primitives.**



invocations of differentially private algorithms lead to a graceful degradation of privacy. The U.S. Census Bureau was the first big organization to adopt DP in 2008 for a product called OnTheMap,²⁹ and subsequently there have been deployments by Google, Apple, Microsoft, Facebook, and Uber.^{2,11,17,18,36}

DP is typically implemented by collecting data from individuals in the clear at a trusted data collector, then applying one or more differentially private algorithms, and finally releasing the outputs. This approach, which we call *standard differential privacy (SDP)*, works in cases like the U.S. Census Bureau where there is a natural trusted data curator. However, when Google wanted to monitor and analyze the Chrome browser properties of its user base to detect security vulnerabilities, they chose a different model called *local differential privacy (LDP)*. In LDP, individuals perturb their records *before* sending them to the server, obviating the need for a trusted data curator. Since the server only sees perturbed records, there is no centralized database of sensitive information that is susceptible to an attack or subpoena requests from governments. The data that Google was collecting—browser fingerprints—uniquely identify individuals. By using LDP, Google was not liable to storing these highly identifying user properties. Due to these attractive security properties, a number of real-world applications of DP in the industry—Google’s RAPPOR,¹⁷ Apple Diagnostics² and Microsoft Telemetry¹¹—embrace the LDP model.

However, the improved security properties of LDP come at a cost in terms of utility. DP algorithms hide the presence or absence of an individual by adding noise. Under the SDP model, counts over the sensitive data, for example, “number of individuals who use the bing.com search engine,” can be released by adding a noise independent of the data size. In the LDP model, noise is added to *each individual record*. Thus, answering the same count query requires adding $O(\sqrt{N})$ error (Theorem 2.1 from Chen et al.¹⁰) for the same level of privacy, where N is the number of individuals participating in the statistic. In other words, under the LDP model, for a database of a billion people, one can only learn properties that



When used in practice, practical trust assumptions are made that enable the deployment of differential privacy-based systems.



are common to at least 30,000 people $O(\sqrt{N})$. In contrast, under SDP, one can learn properties shared by as few as a 100 people ($O(1)$ including constants¹⁵). Thus, the LDP model operates under more practical trust assumptions than SDP, but as a result incurs a significant loss in data utility. In this work, we review literature in this domain under two categories:

► **Cryptography for DP:** We review a growing line of research that aims to use cryptographic primitives to bridge the gap between SDP and LDP. In these solutions, the trusted data curator in SDP is replaced by cryptographic primitives that result in more practical trust assumptions than the SDP model, and better utility than under the LDP model. Cryptographic primitives such as anonymous communication and secure computation have shown significant promise in improving the utility DP implementations while continuing to operate under the practical trust assumptions that are accepted by the security community.

► **DP for cryptography:** Differential privacy is typically applied to settings that involve complex analytics over large datasets. Introducing cryptographic primitives results in concerns about the feasibility of practical implementations at that scale. This has given rise to a second line of work that employs differential privacy as a tool to speed up cryptographic primitives, thereby pushing the frontiers of their practical deployments. While the original cryptographic primitives are defined with respect to perfect privacy, under differential privacy, it is OK to learn distributional information about the underlying dataset. We explore in depth the following cryptographic primitives: secure computation and secure communication and show how in the context of differential privacy one can build “leaky” but efficient implementations of these primitives.

These lines of work both reflect exciting directions for the computer science community. We begin by giving a brief technical introduction to DP. We then discuss the “Cryptography for DP” and “DP for cryptography” paradigms. Finally, we provide concrete ideas for future work as well as open problems in the field through the lens of combining differential privacy and cryptography.

Differential Privacy

Differential privacy¹³ is a state-of-the-art privacy metric for answering queries from statistical databases while protecting individual privacy. Since its inception, there has been considerable research in both the theoretical foundations^{12,14} as well as some real-world DP deployments.^{2,17} The rigorous mathematical foundation and the useful properties of DP have led to an emerging consensus about its use among the security and privacy community.

DP definition. Informally, the privacy guarantees of differential privacy can be understood as follows: Given any two databases, otherwise identical except one of them contains random data in place of data corresponding to any *single* user, differential privacy requires that the response mechanism will behave approximately the same on the two databases. Formally,

Definition 1. Let M be a randomized mechanism that takes a database instance D and has a range O . We say M is (ϵ, δ) -differentially private, if for any neighboring databases (D_1, D_2) that differ in the data of a single user, and for any $S \subseteq O$, we have

$$\Pr[M(D_1) \in S] \leq e^\epsilon \Pr[M(D_2) \in S] + \delta$$

DP enjoys some important properties that make it a useful privacy metric. First, the privacy guarantees of DP have been thoroughly studied using various metrics from statistics and information theory such as hypothesis testing and Bayesian inference.^{25,26} Thus, the semantic meaning of its privacy guarantees is well understood. DP also has a number of composition properties which enable the analysis of privacy leakage for complex algorithms. In particular, sequential composition addresses the impossibility result by Dinur and Nissim¹² and quantifies the degradation of privacy as the number of sequential accesses to the data increases. The post-processing theorem (a special case of sequential composition) ensures the adversary cannot weaken the privacy guarantees of a mechanism by transforming the received response. The end-to-end privacy guarantee of an algorithm over the entire database can thus be established using the above composition theorems and more advanced theorems.¹⁵

Differentially private mechanisms.

Next, we review two classic differentially private mechanisms—the Laplace mechanism and the Randomized Response mechanism—with the following scenario: A data analyst would like to find out how many users use drugs illegally. Such a question would not elicit any truthful answers from users and hence we require a mechanism that guarantees (a) response privacy for the users and (b) good utility extraction for the data analyst.

Laplace mechanism: The Laplace mechanism¹³ considers a trusted data curator (SDP model) who owns a table of N truthful records of users, for example, each record indicates whether a user uses drugs illegally. If a data analyst would like to learn how many users use drugs illegally, the data curator (trusted) computes the true answer of this query and then perturbs it with a random (Laplace distributed) noise that is sufficient to provide privacy. The magnitude of this noise depends on the largest possible change on the query output—also known as the sensitivity of the query—if the data corresponding to a single user is changed.

Randomized response mechanism: Randomized response was first introduced by Warner in 1965 as a research technique for survey interviews. It enabled respondents to answer sensitive questions (about topics such as

sexuality, drug consumption) while maintaining the confidentiality of their responses. An analyst interested in learning aggregate information about sensitive user behavior would like to query this function on a database that is *distributed* across N clients with each client having its own private response $x_1, \dots, x_N$. Instead of releasing x_i directly, the clients release a perturbed version of their response y_i , thus maintaining response privacy. The analyst collects these perturbed responses and recovers meaningful statistics using reconstruction techniques.

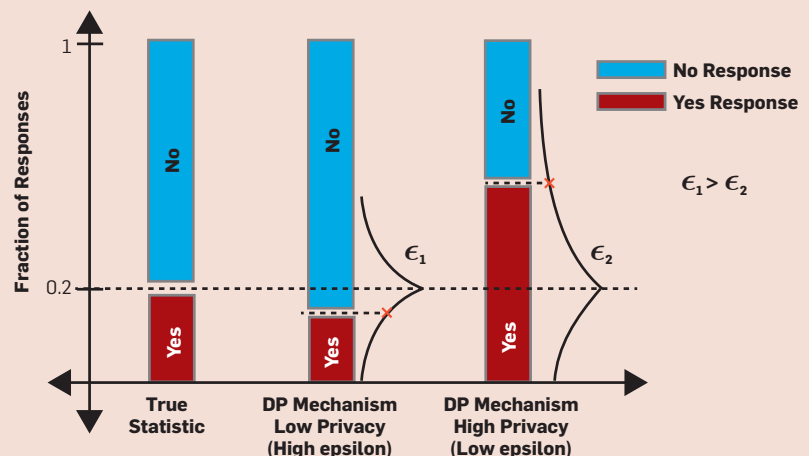
Both these approaches have gained popularity in many applications of differential privacy due to their simplicity as well as the rigorous privacy guarantee on user data. Figure 1 shows the behavior of DP mechanisms for two different privacy values in reference to the true statistic. A less private response results in a more accurate query result while a more private response results in a less accurate query result.

Cryptography for Differential Privacy

By itself, DP is a guarantee on a mechanism and hence is “independent” of the deployment scenario. However, when used in practice, practical trust assumptions are made that enable the deployment of differential privacy-based systems. Here,

Figure 1. Differentially private mechanisms randomize query response to achieve privacy.

If the true response to a query such as “What fraction of users use drugs illegally?” was 20%, then a high privacy response mechanism (low ϵ value) will add a lot of noise yielding low utility. On the contrary, if a low privacy response mechanism was used (high ϵ value), the response will be very close to 20% yielding high utility.



we consider two popular deployment scenarios for differential privacy—Standard Differential Privacy (SDP, graphically represented in Figure 2D) and Local Differential Privacy (LDP, graphically represented in Figure 2A). SDP relies on the need for a trusted data aggregator who follows the protocol. However, in practice, a trusted data aggregator may not always exist. LDP, on the other hand, does not require a trusted data aggregator.^a With the advent of privacy regulations, such as GDPR and FERPA, large organizations such as Google increasingly embrace the LDP model thereby avoiding the liability of storing such sensitive user data. This approach also insures data collectors from potential theft or subpoenas from the government. For these reasons, LDP is frequently a more attractive deployment scenario. However, the utility of the statistics released in LDP is poorer than that in SDP. Consequently, there is a gap in the trust assumptions and the utility achieved by mechanisms in SDP and LDP: high trust assumptions, high utility in SDP and lower trust assumptions, lower utility in LDP. We ask the following question:

a Differentially private federated learning is simply a special case of the LDP deployment scenario.

Can cryptographic primitives help in bridging the gap that exists between mechanisms in the SDP model and the LDP model?

An emerging direction of research has been to explore the use of cryptography to bridge the trust-accuracy gap and obtain the best of both worlds: high accuracy without assuming trusted data aggregator. We explore in depth two concrete examples of the role of cryptography in bridging this gap—anonymous communication, and secure computation and encryption.

Key challenges. There exists a big gap in the accuracy and trust achieved by known mechanisms in the SDP setting with a trusted data curator (Figure 2D) and LDP without such a trusted curator (Figure 2A). Achieving the utility as in the SDP setting while operating under practical trust assumptions such as those in LDP has proven to be a tough challenge. Cryptographic primitives show promise in solving this challenge.

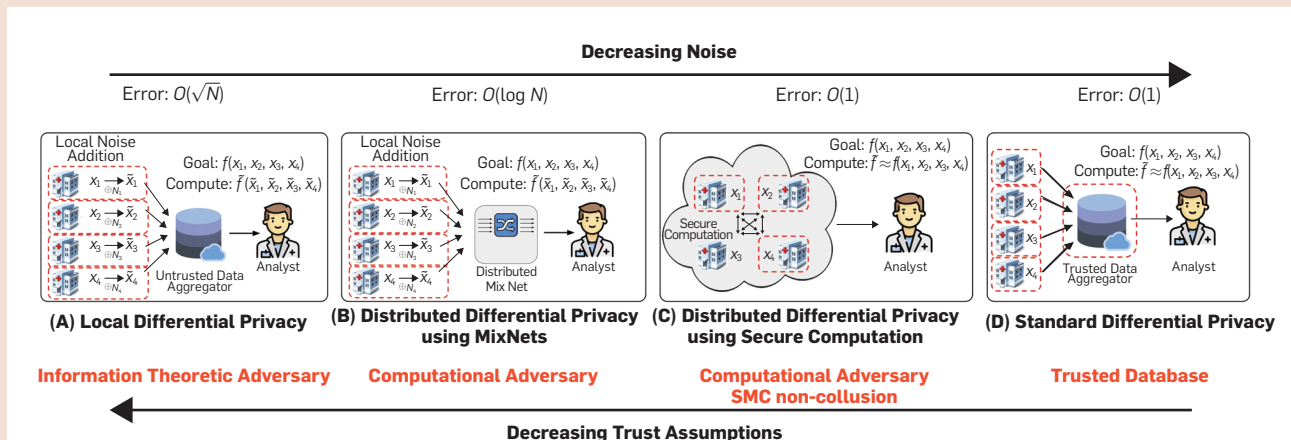
Improve accuracy via anonymous communication. In LDP, each data owner independently perturbs their own input (for example, using the randomized response mechanism) before the aggregation on an untrusted server. This results in a large noise in the final output, $O(\sqrt{N})$ for the case of statistical

counting queries.¹⁰ Applications such as Google's RAPPOR,¹⁷ Apple Diagnostics,² and Microsoft Telemetry,¹¹ which use this LDP deployment model operate under more practical trust assumptions yet suffer from poor accuracy/utility. Recent works^{8,10,16} show the use of an anonymous communication channel can help improve the accuracy of statistical counting query for LDP and thereby eliminate the need for a trusted data curator. We will use one of these systems called Prochlo^{8,16} to illustrate the key idea of how anonymous communication can help improve the accuracy of such applications.

Case Study: Prochlo. Anonymous communication channels, first proposed by Chaum in 1981,⁹ are systems that enable a user to remain unidentifiable from a set of other users (called the anonymity set). A larger anonymity set corresponds to a greater privacy guarantee. Examples of such systems include Mixnets, which use proxies to mix communications from various users. In order to circumvent the limitations of LDP, Google explored the use of an anonymous communication channel to improve the accuracy of queries under DP. The proposed technique is called Prochlo^{8,16} and it consists of three steps as shown in Figure 2B: Encode, Shuffle, and Analyze (ESA).

Figure 2. Various deployment scenarios of differential privacy and the underlying trust assumptions in each of them.

(D) Standard Differential Privacy (SDP) assumes a trusted database, and is thus able to achieve high accuracy, such as, $O(1)$ error. (A) Local Differential Privacy (LDP) on the other hand, does not rely on the use of a trusted database but achieves lower accuracy, that is, $O(\sqrt{N})$ error. The goal is to achieve utility of the SDP setting while operating under more practical assumptions such as the LDP setting (that is, no trusted database). (B) and (C) show how different cryptographic primitives can be used to improve the utility of DP deployments under such practical assumptions.



The first encoding step is similar to LDP where data owners randomize their input data independently. The second step uses an anonymous communication channel to collect encoded data into batches during a lengthy time interval and shuffles this data to remove the linkability between the output of the communication channel and the data owners. Last, the anonymous, shuffled data is analyzed by a data analyst.

The shuffling step is the crucial link in achieving anonymous communication by breaking linkability between the user and their data. This step strips user-specific metadata such as time stamps or source IP addresses, and batches a large number of reports before forwarding them to data analysts. Additional thresholding in this step will discard highly unique reports (for example, a long API bit-vector) to prevent attackers with sufficient background information from linking a report with its data owner. Hence, attacks based on traffic analysis and longitudinal analysis can be prevented, even if a user contributes to multiple reports. Prochlo implements this shuffling step using trusted hardware as proxy servers to avoid reliance on external anonymity channel. Furthermore, this shuffling step can amplify the privacy guarantee of LDP and hence improves the accuracy of the analysis, even when there is a single invocation from a user. We will next show the intuition for this use case.

Accuracy improvement. To illustrate how anonymous communication can help improve accuracy, let us look at a simple example of computing the sum of boolean values from N data owners, $f: \Sigma_{i=1}^N x_i$, where $x_i \in \{0, 1\}$. In LDP, each data owner reports a random bit with probability p or reports the true bit with probability $1 - p$ to achieve ϵ -LDP. When using additional anonymous communication channels, the data owners can enhance their privacy by hiding in a large set of N values, since the attackers (aggregator and analyst) see only the anonymized set of reports $\{\tilde{x}_1, \dots, \tilde{x}_N\}$. The improved privacy guarantee can be shown equivalent to a simulated algorithm that first samples a value s from a binomial distribution $B(N, p)$ to simulate the number of data owners who report a random bit, and then samples a subset



Cryptographic primitives provide strong privacy guarantees. However, deployment of certain cryptographic primitives in practical systems is limited due to the large overhead of these primitives.



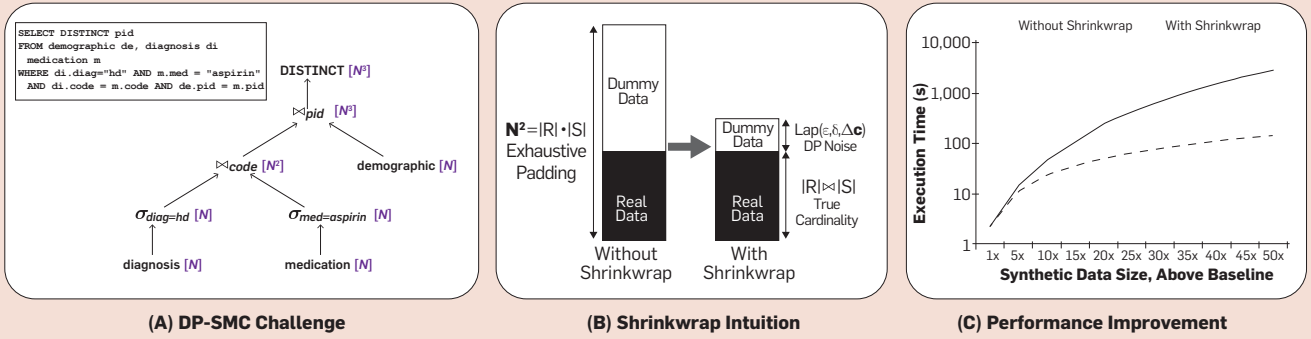
of responses for these s data owners from $\{\tilde{x}_1, \dots, \tilde{x}_N\}$. The randomness of these sampling processes can amplify the privacy parameter based on a well-studied sub-sampling argument.^{3,23} Therefore, given the value of the privacy parameter, the required noise parameter can be scaled down and hence the corresponding error can be reduced to $O(\sqrt{\log N})$. However, these bounds depend on the specific deployment scenarios. For instance, it is shown in Balle et al.⁴ that anonymous communication with a single message per data owner cannot yield expected error less than $O(N^{1/6})$. On the other hand, works such as Balle⁵ and Kasiviswanathan et al.²⁶ show that with a constant number of messages per data owner, it is possible to reduce the error for real-valued DP summation to $O(1)$. Note that these accuracy improvements assume that there is no collusion between the analyst and the anonymous communication, otherwise, the privacy guarantee will fall back to the same as LDP.

In reference to Figure 2, these works demonstrate the improvement in going from Figure 2A to Figure 2B showing a trade-off between accuracy and trust assumptions.

Improve trust via encryption and secure computation. SDP requires the use of a trusted data aggregator to achieve high accuracy. A number of works have explored the use of encryption and secure computation to eliminate the need for this trusted data aggregator.^{1,6,33} The key challenge here is to maintain the same level of accuracy as in SDP. We will use one of these proposed systems called DJoin to demonstrate the use of secure computation to enable high accuracy computation without the need for a trusted data aggregator. There is a complementary synergy between secure computation and DP and thus their combination achieves a strong privacy protection. For instance, secure computation ensures all parties learn only the output of the computation but nothing else while DP bounds the information leakage of individuals in the output of the computation, resulting in a system that is better than the use of secure computation or DP alone.

Case Study: DJoin. Consider a simple setting where two parties would like to compute the intersection size of their

Figure 3. (A) Exhaustive padding of intermediate results in an oblivious query evaluation; (B) Effect of Shrinkwrap on intermediate result sizes when joining tables R and S; (C) Aspirin count with synthetic data scaling. Executed using Circuit model. $\epsilon = 0.5$, $\epsilon = .00005$.



data while preserving DP for both datasets. If each party does not trust each other, how can we ensure a constant additive error as if they trust each other? It is well known that the lower bound for the error of this query is $\sqrt{N}$, where N is the data size of each party,³⁰ if we want to ensure the view of each party satisfies differential privacy. However, if we assume both parties are computationally bounded, a constant additive error can be achieved.

DJoin³³ offers a concrete protocol for achieving DP under this assumption. This protocol applies private set-intersection cardinality technique to privately compute the noisy intersection set of the two datasets. First, party A defines a polynomial over a finite field whose roots are the elements owned by A. Party A then sends the homomorphic encryptions of the coefficients to party B, along with its public key. Then the encrypted polynomial is evaluated at each of Party B's inputs, followed by a multiplication with a fresh random number. The number of zeros in the results is the true intersection size between A and B. To provide DP, party B adds a number of zeros (differentially private noise of $O(1)$ independent of data size) to the results and sends the randomly permuted results back to party A. Party A decrypts the results and counts the number of zeros. Party A also adds another copy of differentially private noise to the count and sends the result it back to party B. In other words, both parties add noise to their inputs to achieve privacy. However, the final protocol output has only an error of $O(1)$, which is the same as the SDP setting.

Trust improvement. Using secure computation and encryptions achieves

a constant additive error like SDP and prevents any party from seeing the other party's input in the clear. However, this requires an additional assumption of all parties being computationally bounded in the protocol. Hence, the type of DP guarantee achieved in DJoin is known as computational differential privacy.³² In addition, most of the existing protocols consider honest-but-curious adversaries who follow the protocol specification or consider malicious adversaries with an additional overhead to enforce honest behavior, that is, verify that the computation was performed correctly.

In reference to Figure 2, these works demonstrate the improvement in going from Figure 2D to Figure 2C eliminating the need for a trusted data aggregator.

Differential Privacy for Cryptography

As we discussed earlier, cryptographic primitives show promise in bridging the utility gap between SDP and LDP. However, the large overhead of implementing these conventional cryptographic primitives forms a bottleneck for the deployment of such systems. This motivates the need to enhance the performance of such cryptographic primitives. We ask the following question:

“Can we formulate leaky versions of cryptographic primitives for enhancing system performance while rigorously quantifying the privacy loss using DP?”

DP-cryptographic primitives^{7,37,38} are significant for two reasons. First, since the final privacy guarantees of such systems are differential privacy, it is natural to relax the building blocks

such as cryptographic primitives to provide differentially private guarantees. Secondly, the composability properties of DP allow for rigorous quantification of the privacy of the end-to-end system. We showcase benefits of “DP-cryptographic” systems through two detailed case studies on secure computation and secure communication.

Key challenges. Cryptographic primitives provide strong privacy guarantees. However, deployment of certain cryptographic primitives in practical systems is limited due to the large overhead of these primitives. Relaxing the privacy guarantees in a manner that is amenable to rigorous quantification is difficult and differential privacy can be well utilized to provide a solution to this problem to improve performance overhead.

Improve performance of cryptographic computation primitives. Cryptographic computation primitives such as Fully Homomorphic Encryption (FHE) and secure Multi-Party Computation (MPC) enable private computation over data. Over the past few years, there has been tremendous progress in making these primitives practical—a promising direction is MPC, which allows a group of data owners to jointly compute a function while keeping their inputs secret. Here, we show the performance improvement on MPC based private computation, in particular, *differentially private query processing*.

Case Study: Shrinkwrap. Shrinkwrap⁷ is a system that applies DP throughout an SQL query execution to improve performance. In secure computation, the computation overheads depend on the largest possible data

size so that no additional information is leaked. For example, two parties would like to securely compute the answer for the SQL query shown in Figure 3A. This query asks for the number of patients with heart disease who have taken a dosage of “aspirin.” Figure 3A expresses this query as a directed acyclic graph of database operators. For example, the first filter operator takes N records from the two parties and outputs an intermediate result that has patients with heart disease (hd). To hide the selectivity (fraction of records selected) of this operator, the baseline system must pad the intermediate result to its maximum possible size, which is the same as the input size. Exhaustive padding will also be applied to the intermediate output of the two joins and result in an intermediate result cardinality of N^3 and a high-performance overhead. However, if the selectivity of the filter is 10^{-3} , cryptographic padding adds a $1000\times$ overhead. Is there a way to pad fewer dummies to the intermediate result while ensuring a provable privacy guarantee?

Shrinkwrap helps reduce this overhead by padding each intermediate output of the query plan to a differentially private cardinality rather than to the worst case. As shown in Figure 3B, without Shrinkwrap, the output of a join operator with two inputs, each of size N is padded to a size of N^2 . With Shrinkwrap, the output is first padded to the worst size and the output is

sorted such that all the dummies are at the end of the storage. This entire process is executed obliviously. Then Shrinkwrap draws a non-negative integer value with a general Laplace mechanism⁷ and truncates the storage at the end. This approach reduces the input size of the subsequent operators and thereby their I/O cost. We can see from Figure 3C that Shrinkwrap provides a significant improvement in performance over the baseline without DP padding for increasing database sizes.

The relaxed privacy in the secure computation of Shrinkwrap can be quantified rigorously⁷ using computational differential privacy. Assuming all parties are computationally bounded and work in the semi-honest setting, it can be shown that data owners have a computational differentially private view over the input of other data owners; when noisy answers are returned to the data analyst, the data analyst has a computational differentially private view over the input data of all the data owners.

Improve performance of cryptographic communication primitives. Anonymous communication systems aim to protect user identity from the communication recipient and third parties. Despite considerable research efforts in this domain, practical anonymous communication over current Internet architecture is proving to be a challenge. Even if the message contents are encrypted, the packet metadata is

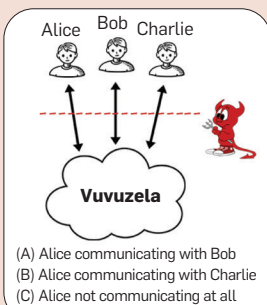
difficult to hide. On one end, systems such as Dissent³⁹ offer strong privacy guarantees yet can scale only to a limited number of participants. On the other end, practical deployed systems such as Tor are vulnerable to traffic analysis and other attacks, limiting their use due to the non-rigorous nature of their privacy guarantees. We will show a case study that uses DP to reduce the communication cost while offering rigorous privacy guarantee. We denote this primitive differentially private anonymous communication.

Case Study: Vuvuzela. Vuvuzela³⁷ is an anonymous communication system that uses DP to enable a highly scalable system with relaxed yet rigorously quantified privacy guarantees. Vuvuzela provides indistinguishable traffic patterns to clients who are actively communicating with other clients, and clients who are not communicating with anyone. In reference to Figure 4, an adversary is unable to distinguish the following three scenarios: Alice not communicating; Alice communicating with Bob; and, Alice communicating with Charlie. In each of the scenarios, a Vuvuzela client’s network traffic appears indistinguishable from the other scenarios.

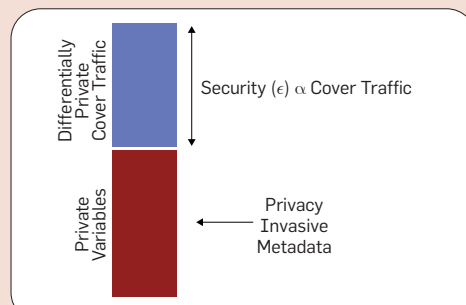
Vuvuzela employs a number of servers $S_1, \dots, S_n$ where at least one of the servers is assumed to be honest. Clients send (and receive) messages to (and from) the first server, which in turn is connected to the second server and so

Figure 4. Vuvuzela is a secure messaging system.

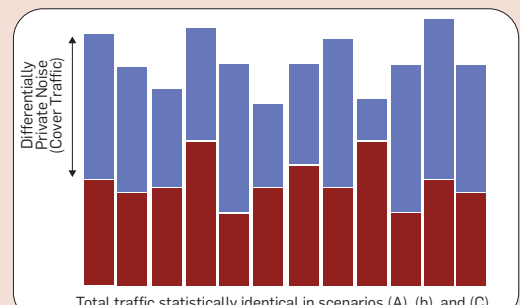
An adversary who can observe and tamper with all network traffic cannot distinguish whether Alice is messaging Bob, Charlie, or is simply not communicating. Vuvuzela uses differential privacy to add noise and mask the privacy invasive metadata, thereby provably hiding information about user communication patterns. Vuvuzela achieves a throughput of 68,000 messages per second for a million users scaling linearly with number of users.



(A) Schematic for Vuvuzela



(B) Performance improvement of Vuvuzela



(C) Security proportional to cover traffic

on. The client creates a layered encryption of its message m , that is, $\text{Enc}_{s_1}(\dots \text{Enc}_{s_n}(m))$, where $\text{Enc}_s(\cdot)$ is the encryption under the key of server S . The clients leave messages at virtual locations in a large space of final destinations (called dead drops), where the other legitimate client can receive it. To hide if a client is communicating or not, a client not in an active conversation makes fake requests to appear indistinguishable from a client in an active conversation. If two clients are in active conversation, they exchange messages via the same random dead drop.

Vuvuzela's threat model assumes at least one server is honest and the adversary is a powerful network level adversary (observing all network traffic) potentially corrupting all other servers.^b The only computation hidden from the adversary is the local computation performed by the honest server which unlinks users' identifiers from the dead drops and adds cover (dummy) traffic. As a consequence, the adversary can only observe the number of single or double exchange requests at the dead drop locations. Each Vuvuzela server adds cover traffic using a Laplace distribution to randomize the number of single dead drops and the number of double dead drops, which is observable by the adversary. Such random cover traffic addition along with the assumption of at least one honest server provides DP guarantees for the observed variables. In other words, Vuvuzela adds noise (cover network traffic) to the two observables (by the adversary) viz. the number of dead drops with one exchange request, and the number of dead drops with two exchange requests, thereby providing communication privacy to clients. This privacy relaxation enables Vuvuzela to scale to a large number of users—it can achieve a throughput of 68,000 messages per second for a million users. Systems such as Stadium,³⁵ and Karaoke²⁷ further improve upon Vuvuzela and scale to even larger sets of users.

Limitations of differentially private cryptography. We caution readers against careless combinations of differential privacy and cryptographic primitives. First, the limitations of both DP as well as cryptographic primitives apply to DP

cryptographic primitives. For instance, an open question is deciding an appropriate level for the privacy budget. Most applications that utilize DP to improve the performance of cryptographic systems involve a trade-off between the level of privacy achieved and the performance of the systems. More generally, differentially private cryptographic systems open up new trade-offs in a privacy-performance-utility space. For instance, in the case of Shrinkwrap, weaker privacy guarantee directly leads to lower performance overhead (privacy performance trade-off while keeping the accuracy level of the query answer constant). On the other hand, systems such as RAPPOR allow for approximate computation of statistics and primarily provide a privacy-utility trade-off. Second, designers need to carefully consider the suitability of these hybrid techniques in their applications as these combinations involve more complex trust assumptions and hence a more complicated security analysis. We remind the reader that while proposing newer DP systems for cryptography, it is imperative to understand the meaning of the privacy guarantees for the application in context. In other words, differential privacy for cryptography may not be the right thing to do in all cases; however, it is well motivated when the goal is to build a differentially private system. Finally, composition results, which bound the privacy loss for a sequence of operations need to be independently studied.

Discussion and Open Questions

Here, we provide directions for future work highlighting important and emerging open questions in the field. We discuss open challenges in deploying differential privacy in the real world—realistic datasets, alternative models and trust assumptions, and other DP-cryptographic primitives. Finally, we caution readers against callous combinations of differential privacy and cryptography.

Differential privacy frameworks—SDP, LDP, and beyond. Over the past decade, there has been significant progress in enabling applications in the standard differential privacy model. For instance, there have been research efforts in attuning DP to handle realistic challenges such as multi-dimensional and complex data—involving graphs,

time series, correlated data.^{24,28} Similarly, there has been work in designing a tailored DP mechanism that is optimized for particular application setting to achieve good accuracy.^{22,31} Prior work has explored combinations of sequential and parallel composition, dimensionality reduction, and sensitivity bound approximations to achieve good accuracy in the SDP model. However, much work needs to be done in adapting state-of-the-art techniques in SDP to more complex deployment scenarios such as LDP. For instance, an open question is the following:

“Is there an algorithm that can efficiently search the space of DP algorithms in the LDP setting for the one that answers the input query with the best accuracy?”

Research advances have demonstrated such mechanisms for the SDP model,^{22,31} however, the discovery of such mechanisms in the LDP setting remains an open question. On a similar note, it is unclear how nuanced variants of DP that have been proposed to handle these more complex databases^{24,28} in the SDP setting translate into LDP or more complex deployment settings.

Differential privacy in practice—Trust assumptions vs accuracy gap.

We have seen how deployments of DP that differ in the trust assumptions provide approximately the same privacy guarantee, but with varying levels of accuracy. In particular, we looked at two popular deployment scenarios: SDP and LDP. There exist other trust assumptions that we have not covered in this article in detail. For instance, Google's recently proposed Prochlo system⁸ uses trusted hardware assumptions to optimize utility of data analytics. On a similar note, Groce et. al.²¹ consider yet another model—where the users participating are malicious. This is the first work to explore a malicious adversarial model in the context of DP and the development of better accuracy mechanisms for such a model is an open research question. More concretely, we can ask:

“What other models of deployment of differential privacy exist and how

^b Even Tor, a practical anonymous communication system, does not protect against such network level adversaries.³⁴

do we design mechanisms for them? Can other technologies such as MPC, FHE, trusted hardware opens up new opportunities in mechanism design?”

An interesting theoretical question is to characterize the separation between different trust models in terms of the best accuracy achievable by a DP algorithm under that model. For instance, McGregor et. al.³⁰ provide separation theorems, that is, gaps in achievable accuracy between (information-theoretic) differential privacy and computational differential privacy for two-party protocols. We ask:

In the Mixnets model (Figure 2B), what is the lower bound on the error for aggregate queries over relational transformations (like joins and group-by) over the data records? An example of such an aggregate is the degree distribution of a graph that reports the number of nodes with a certain degree.

Relaxing cryptographic security via DP: The emerging paradigm of leaky yet differentially private cryptography leads to a number of open questions for the research community. So far, the research community has explored the intersection of differential privacy and cryptographic primitives in limited contexts such as ORAM, MPC, and anonymous communication. However, there exists a broader opportunity to explore the trade-offs of DP cryptographic primitives in contexts such as program obfuscation, zero-knowledge proofs, encrypted databases, and even traffic/protocol morphing. Here, we can ask:

“What other cryptographic primitives can benefit in performance from a privacy relaxation quantified rigorously using differential privacy? How can we design such relaxed primitives?”

In the context of differentially private data analysis, there is a trade-off between privacy and utility. In the context of differentially private cryptographic primitives and resulting applications, there is a broader trade-off space between privacy, utility, and performance. Another open question is:

“What lower bounds exist for overhead of cryptographic

primitives when the privacy guarantees are relaxed using DP?”

Another challenge is how to design optimized protocols that achieve desired trade-offs in the new design space of differentially private cryptography. The trade-off space between privacy, utility, and performance is non-trivial, especially for complex systems. An interesting research question is:

“How to correctly model the trade-off space of real systems so that system designers can decide whether it is worth sacrificing some privacy or utility for a better performance?”

References

- Agarwal, A., Herlihy, M., Kamara, S., and Moataz, T. Encrypted Databases for Differential Privacy. *IACR Cryptology ePrint Archive*, 2018.
- Apple is using Differential Privacy to help discover the usage patterns of a large number of users without compromising individual privacy. <https://apple.co/3otHYKw>
- Balle, B., Barthe, G., and Gaboardi, M. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in Neural Information Processing Systems 31*. S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.), Curran Associates, Inc., 2018, 6277–6287.
- Balle, B., Bell, J., Gascon, A., and Nissim, K. The privacy blanket of the shuffle model. In *Proceedings of the Annual Intern. Cryptology Conf.* Springer, 2019, 638–667.
- Balle, B., Bell, J., Gascon, A., and Nissim, K. Private summation in the multi-message shuffle model. *arXiv preprint arXiv:2002.00817*, 2020.
- Bater, J., Elliott, G., Eggen, C., Goel, S., Kho, A., and Rogers, J. SMCQL: Secure querying for federated databases. In *Proceedings of the VLDB Endowment 10*, 6 (2017), 673–684.
- Bater, J., He, X., Ehrich, W., Machanavajhala, A., and Rogers, J. Shrinkwrap: Efficient SQL query processing in differentially private data federations. In *Proceedings of the VLDB Endowment*, 2018.
- Bittau, A. et al. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 2017 Symp. on Operating Systems Principles*.
- Chaum, D.L. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM* 24, 2 (Feb. 1981), 84–90.
- Cheu, A., Smith, D., Ullman, J., Zeber, D., and Zhilyaev, M. Distributed differential privacy via shuffling. *Theory and Practice of Differential Privacy*, 2018.
- Ding, B., Kulkarni, J., and Yekhanin, S. Collecting telemetry data privately. In *Proceedings of the 2017 Annual Conf. on Neural Information Processing Systems*.
- Dinur, I. and Nissim, K. Revealing information while preserving privacy. In *Proceedings of the ACM SIGMOD-SIGACT-SIGART Symp. on Principles of Database Systems*. ACM, 2003.
- Dwork, C. Differential privacy. *Automata, Languages and Programming*. Springer, 2006.
- Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Theory of Cryptography Conf.* Springer, 2006, 265–284.
- Dwork, C. A. Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 2014.
- Erlingsson, U., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., and Thakurta, A. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Annual ACM-SIAM Symp. Discrete Algorithms*, 2019, 2468–2479.
- Erlingsson, U., Pihur, V., and Korolova, A. RAPPOR: Randomized aggregatable privacy preserving ordinal response. In *Proceedings of the ACM Conf. on Computer and Communications Security*, 2014.
- Facebook Privacy-Protected URLs light Table Release. <https://bit.ly/3kKSqXY>.
- Garfinkel, S., Abowd, J., and Martindale, C. Understanding database reconstruction attacks on public data. *Commun. ACM* 62 (2019), 46–53.
- Ghazi, B., Manurangsi, P., Pagh, R., and Velingker, A. Private aggregation from fewer anonymous messages. In *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2020, 798–827.
- Groce, A., Rindal, P., and Rosulek, M. Cheaper private set intersection via differentially private leakage. *Privacy Enhancing Technologies Symposium*, 2019.
- Johnson, N., Near, J., and Song, D. Towards practical differential privacy for SQL queries. In *Proceedings of the VLDB Endowment*, 2018.
- Kasiviswanathan, S., Lee, H., Nissim, K., Raskhodnikova, S., and Smith, A. What can we learn privately? *SIAM J. Comput.* 40, 3 (June 2011), 793–826.
- Kasiviswanathan, S., Nissim, K., Raskhodnikova, S., and Smith, A. Analyzing graphs with node differential privacy. *Theory of Cryptography*. Springer, 2013, 457–476.
- Kasiviswanathan, S. and Smith, A. On the ‘semantics’ of differential privacy: A Bayesian Formulation. *J. Privacy and Confidentiality*, 2014.
- Kifer, D. and Machanavajhala, A. Pufferfish: A framework for mathematical privacy definitions. In *Proceedings of the ACM Trans. Database Systems* 39, 1 (2014).
- Lazar, D., Gilad, Y., and Zeldovich, N. Karaoke: Distributed private messaging immune to passive traffic analysis. In *USENIX Symp. on Operating Systems Design and Implementation*, 2018.
- Liu, C., Chakraborty, S., and Mittal, P. Dependence makes you vulnerable: Differential privacy under dependent tuples. In *Proceedings of the 2016 Symp. on Network and Distributed System Security*.
- Machanavajhala, A., Kifer, D., Abowd, J., Gehrke, J., and Vilhuber, L. Privacy: Theory meets practice on the map. In *Proceedings of the 2008 IEEE Intern. Conf. on Data Engineering*.
- McGregor, A., Mironov, I., Pitassi, T., Reingold, O., Talwar, K., and Vadhan, S. The limits of two-party differential privacy. In *Proceedings of the 2010 Symp. on Foundations of Computer Science*. IEEE.
- McKenna, R., Miklau, G., Hay, M., and Machanavajhala, A. Optimizing error of high-dimensional statistical queries under differential privacy. In *Proceedings of the VLDB Endowment 11*, 10 (2018), 1206–1219.
- Mironov, I., Pandey, O., Reingold, O., and Vadhan, S. Computational differential privacy. In *Proceedings of the 29th Annual Intern. Cryptology Conf. Advances in Cryptology*. Springer-Verlag, Berlin, Heidelberg, 2009, 216–242.
- Narayan, A. and Haeberlen, A. DJoin: Differentially private join queries over distributed databases. In *Proceedings of the 2012 USENIX Symp. on Operating Systems Design and Implementation*.
- Sun, Y., Edmundson, A., Vanbever, L., Li, O., Rexford, J., Chiang, M., and Mittal, P. Raptor: Routing attacks on privacy in Tor. In *Proceedings of the 2015 USENIX Security Symp.*
- Tyagi, N., Gilad, Y., Leung, D., Zaharia, M., and Zeldovich, N. Stadium: A distributed metadata-private messaging system. In *Proceedings of the 2017 Symp. on Operating Systems Principles*.
- Uber Releases Open Source Project for Differential Privacy; <https://bit.ly/2RV16hX>.
- van den Hooff, J., Lazar, D., Zaharia, M., and Zeldovich, N. Vuvuzela: Scalable private messaging resistant to traffic analysis. In *Proceedings of the 2015 Symp. on Operating Systems Principles*.
- Wagh, S., Cuff, P. and Mittal, P. Differentially private oblivious ram. In *Proceedings on Privacy Enhancing Technologies 4* (2018), 64–84.
- Wolinsky, D., Corrigan-Gibbs, H., Ford, B., and Johnson, A. Dissent in numbers: Making strong anonymity scale. In *Proceedings of the 2012 USENIX Symp. on Operating Systems Design and Implementation*.

Sameer Wagh (snwagh@gmail.com) is a post-doc researcher at the University of California, Berkeley, CA, USA.

Xi He is an assistant professor in the Cheriton School of Computer Science at the University of Waterloo, Ontario, Canada.

Ashwin Machanavajhala is an associate professor and director of Graduate Studies in the Department of Computer Science at Duke University, Durham, NC, USA.

Prateek Mittal is an associate professor in the Department of Electrical Engineering at Princeton University, Princeton, NJ, USA.