Age-Dependent Differential Privacy

Meng Zhang
ZJU-UIUC Institute, Zhejiang University
mengzhang@intl.zju.edu.cn

Ermin Wei
Northwestern University
ermin.wei@northwestern.edu

Randall Berry
Northwestern University
rberry@northwestern.edu

Jianwei Huang
The Chinese University of Hong Kong, Shenzhen
jianweihuang@cuhk.edu.cn

ABSTRACT

The proliferation of real-time applications has motivated extensive research on analyzing and optimizing data freshness in the context of *age of information*. However, classical frameworks of privacy (e.g., differential privacy (DP)) have overlooked the impact of data freshness on privacy guarantees, and hence may lead to unnecessary accuracy loss when trying to achieve meaningful privacy guarantees in time-varying databases. In this work, we introduce *age-dependent DP*, taking into account the underlying stochastic nature of a time-varying database. In this new framework, we establish a connection between classical DP and age-dependent DP, based on which we characterize the impact of data staleness and temporal correlation on privacy guarantees. Our characterization demonstrates that *aging*, i.e., using stale data inputs and/or postponing the release of outputs, can be a new strategy to protect data privacy in addition to noise injection in the traditional DP framework. Furthermore, to generalize our results to a multi-query scenario, we present a sequential composition result for age-dependent DP. We then characterize and achieve the optimal tradeoffs between privacy risk and utility. Finally, case studies show that, when achieving a target of an arbitrarily small privacy risk in a single-query case, the approach of combining aging and noise injection can achieve a bounded accuracy loss, whereas using noise injection only (as in the DP benchmark) will lead to an unbounded accuracy loss.

CCS CONCEPTS

• **Security and privacy** → **Privacy-preserving protocols**; • **Networks** → *Network performance analysis*; *Network privacy and anonymity*; • **Theory of computation** → Timed and hybrid models;

KEYWORDS

Age of Information, Differential Privacy

ACM Reference Format:
Meng Zhang, Ermin Wei, Randall Berry, and Jianwei Huang. 2022. Age-Dependent Differential Privacy. In *Abstract Proceedings of the 2022 ACM SIGMETRICS/IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS/PERFORMANCE '22 Abstracts), June 6–10, 2022, Mumbai, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3489048.3526953

1 INTRODUCTION

Fresh data has become indispensable for ubiquitous real-time applications, ranging from Internet-of-things (IoT) systems (e.g., healthcare wearables), cyber-physical systems (e.g., autonomous automobile systems), to financial services. For instance, real-time location and velocity information of motor vehicles is the key to realize reliable and safe autonomous driving. The increasing importance of fresh data has been driving research on a new metric, *Age of information (AoI)*, to measure the timeliness of the information that a receiver has about the status of a remote source [2].

An unpredecented amount of personal data has been generated in such real-time applications, which may severely compromise user privacy, as an adversary may infer information about a user based on such frequently generated data. The leakage of such information may lead to threatening consequences. For instance, stalkers may access real-time GPS location data from location-based service providers to track mobile users. Therefore, without a strong privacy-preserving scheme, an adversary may hack the real-time personal data to obtain illegal benefits.

To combat such privacy leakage attacks, a widely used analytical framework is *differential privacy (DP)* [3], which quantifies the level of individual privacy leakage due to releasing aggregate information from a database. The key idea of DP is to provide strong privacy guarantees by injecting tunable levels of noise into the aggregate information before its release, with the goal of maintaining a proper tradeoff between privacy and statistical utility of databases.

Existing techniques have largely overlooked the impact of data freshness for time-varying databases. Intuitively, as some data has diminishing values over time, releasing outdated data may lead to less privacy leakage if a user only focuses on protecting its real-time status. As an example, for a mobile user trying to protect its real-time location, the accuracy of an adversary's inference (hence the user's privacy leakage) will significantly reduce as the location data becomes outdated. This observation motivates us to answer the following question: *How should one quantify the impact of data timeliness on data privacy protection?*

J. Huang is also with the Shenzhen Institute of Artificial Intelligence and Robotics for Society.

This work is supported by the Zhejiang University/University of Illinois at Urbana-Champaign Institute Starting Fund, the Shenzhen Science and Technology Program (Project JCYJ20210324120011032), Guangdong Basic and Applied Basic Research Foundation (Project 2021B1515120008), the Shenzhen Institute of Artificial Intelligence and Robotics for Society, and NSF grant ECCS-2030251.

Corresponding authors are Randall Berry and Jianwei Huang.

SIGMETRICS/PERFORMANCE '22 Abstracts, June 6–10, 2022, Mumbai, India.

ACM ISBN 978-1-4503-9141-2/22/06.
https://doi.org/10.1145/3489048.3526953

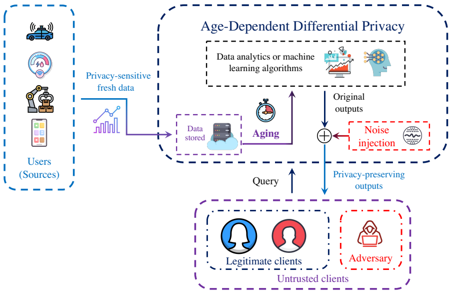

Figure 1: An age-dependent differential privacy mechanism.

2 MODEL

We consider a set $\mathcal{I} = \{1 \leq i \leq I\}$ of users and an infinite-horizon discrete-time model with time $t \in \mathbb{N}$. The system starts to operate at time $t = 0$, and $\mathbb{N}$ denotes the space of all non-negative integers. Each user i's privacy-sensitive data is captured by samples of a sequence of random variables, captured by a discrete-time stationary process $\{x_{i,t}\}_{t\in\mathbb{N}}$. We use $\{X_t\}_{t\in\mathbb{N}}$ to denote the aggregate process across all users, where $X_t \triangleq \{x_{i,t}\}_{i\in\mathcal{I}}$ is the database of all users at time t, belonging to the state space $\mathcal{X}$.

Motivated by the framework of age of information and DP, this work proposes an age-related generalization of DP that provides more meaningful guarantees for time-varying datasets. We name it *age-dependent differential privacy.*

Definition 1 (Age-Dependent Differential Privacy). *A single-query mechanism $M : \mathcal{X} \rightarrow \mathcal{Y}$ is (ϵ, t)-age-dependent DP if for all pairs $X, X' \in \mathcal{X}$ which differ in only one user's data:*

$$\Pr[M(X_0) \in \mathcal{W} | X_t = X] \leq \exp(\epsilon)\Pr[M(X_0) \in \mathcal{W} | X_t = X'], \quad (1)$$

for all $\mathcal{W} \subset \mathcal{Y}$, where the probability is over the randomness of both the output of mechanism M and the stochastic process of $\{X_t\}_{t\in\mathbb{N}}$.

Furthermore, whereas existing studies largely rely on injecting noise to achieve DP [3], our proposed framework provides a new design space for protecting privacy, namely *aging*. In particular, when t becomes large, X_0 is outdated and the conditional probability becomes indistinguishable, i.e., it achieves privacy protection perfectly. This also raises another challenge: *How should one characterize age-dependent privacy guarantees by leveraging aging along with the classical noise injection methods?*

3 RESULTS

We provide theoretic guarantees achieved by classical DP mechanisms when adopted in the new age-dependent DP framework. This establishes a connection between the classical DP notion and our proposed age-dependent variant. Therefore, it enables us to derive the achievable age-dependent privacy guarantees by exploiting both classical methods (e.g., noise injection) and aging (e.g., timing inputs and outputs).

Theorem 1. *If a mechanism M is ϵ_C-DP, then M is also $(\epsilon(t), t)$-age-dependent DP, where $\epsilon(t)$ satisfies*

$$\epsilon(t) = \ln\left(1 + \Delta(t) \cdot (\exp(\epsilon_C) - 1)\right), \ \forall t \in \mathbb{N}, \quad (2)$$

where $\Delta(t)$ is the maximal total variation distance, defined as

$$\Delta(t) \triangleq \max_{i\in\mathcal{I}} \max_{x_{i,0}, x'_{i,0} \in \mathcal{X}_i} \delta\left(\hat{P}_{i,t}(x_{i,0}, \cdot), \hat{P}_{i,t}(x'_{i,0}, \cdot)\right), \forall t \in \mathbb{N}, \quad (3)$$

and $\hat{P}_{i,t}(\cdot, \cdot)$ represents the t-step transition probability matrix of user i's reversed process.

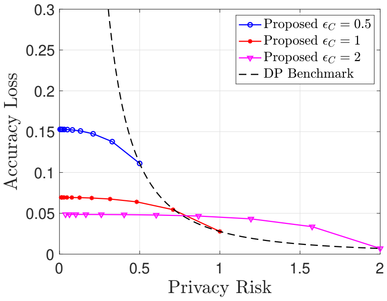

Figure 2: Comparison of proposed mechanisms and the classical DP benchmark.

The significance of Theorem 1 is two-fold. First, it establishes the connection between the guarantees achieved by the classical DP and the age-dependent DP. This showcases a methodology to attain privacy guarantees in (2), by combining noise injection and aging. We present an example of age-dependent DP mechanism in Figure 1. Second, it also indicates that $\Delta(t)$ is the only key feature needed for characterizing such a bound. For a specific process that satisfies certain mixing or geometric ergodicity (aperiodic and recurrent Markov chains on finite state spaces) properties, $\Delta(t)$ converges to zero at a certain rate (e.g., a geometric rate). In this case, since $\lim_{x\to 0} \ln(1 + x)/x = 1$, (2) further implies that the age-dependent privacy risk $\epsilon(t)$ converges to zero at the same rate as $\Delta(t)$. Finally, as shown in Fig. 2, proposed strategies that combine aging along with noise injection may lead to unbounded performance gains compared to the DP benchmark (which only uses noise injection).

Furthermore, the operation of real-time systems involving frequent data updates necessitates the understanding of the performances of privacy-preserving mechanisms with sequential queries. In this case, age-dependent DP mechanisms bring two new challenges, compared to the classical ones. First, our characterization further depends on the *timing* of both inputs (how stale the input databases for all queries) and outputs (when the outputs of all queries are releases). In contrast, the classical composition results only depend on the number of queries. The second challenge is to make the optimal tradeoff between privacy and utility, as it involves solving an optimization problem with an objective that cannot be directly expressed in a closed form. To this end, we construct multi-query mechanisms by *composing* single-query mechanisms, and derive the corresponding privacy guarantees over time. By exploiting a special structure of the optimal solutions, we are then able to make the optimal tradeoffs between privacy and utility.

REFERENCES

[1] M. Zhang, E. Wei, R. Berry, and J. Huang. 2022. Age-dependent differential privacy. https://www.dropbox.com/s/ud5195uetdvas13/AgeDependent.pdf?dl=0

[2] R. D. Yates, Y. Sun, D. R.Brown, S. K. Kaul, E. Modiano, and S. Ulukus. 2021. Age of information: An introduction and survey. *IEEE Journal on Selected Areas in Communications*, 39, 5, 1183-1210.

[3] C. Dwork. 2006. Differential privacy. *Proc. 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP)*, 4052, 1–12.