# Age-Dependent Differential Privacy

Meng Zhang , *Member, IEEE*, Ermin Wei , *Member, IEEE*, Randall Berry , *Fellow, IEEE*, and Jianwei Huang , *Fellow, IEEE*

*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, which may provide a new tool for 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 assume knowledge of the data process's statistical information and establish a connection between classical DP and age-dependent DP. We use this connection to characterize the impact of data staleness and temporal correlation on privacy guarantees. Our characterization reveals that the total variation distance is the sole essential statistical information. Moreover, we demonstrate that *aging*, which involves utilizing stale data inputs and/or delaying the release of outputs, can serve as a novel strategy for safeguarding data privacy, in addition to the traditional approach of injecting noise in the DP framework. Furthermore, to generalize our results to a multi-query scenario, we present a sequential composition result for age-dependent DP under any publishing and aging policies. We then characterize the optimal tradeoffs between privacy risk and utility and show how this can be achieved. Finally, case studies show that to achieve an arbitrarily small privacy risk in a single-query case, combing aging and noise injection only leads to a bounded accuracy loss, whereas using noise injection only (as in the benchmark case of DP) will lead to an unbounded accuracy loss.

*Index Terms*— Differential privacy (DP), age of information (AoI).

## I. INTRODUCTION

### A. Background

FRESH data has become indispensable for ubiquitous real-time applications, including Internet-of-things (IoT) systems (e.g., healthcare wearables), cyber-physical systems (e.g., autonomous transportation systems), and financial services. For instance, real-time location and velocity information of motor vehicles is the key to realize reliable and safe autonomous driving, and real-time stock quotes are essential for developing effective high-frequency trading strategies in financial markets. The increasing importance of fresh data has been driving research on a new metric, *Age of information (AoI)*, to measure the timeliness of information [2].

An unprecedented amount of personal data is generated in such real-time applications. This may severely compromise user privacy. Privacy sensitive information may include user identities and user properties, whose leakage may lead to undesirable consequences. For instance, burglars can use real-time electricity usage readings to infer whether houses are occupied; stalkers may access real-time GPS location data to track mobile users.

To combat privacy leakage, researchers have proposed numerous solutions, varying in terms of the level of data protection and implementation complexity (see a survey in [12]). 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. Apple [4], Google [5], Microsoft [7], and the U.S. Census Bureau [6] have adopted DP frameworks. 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.

Despite its privacy guarantees against arbitrary adversaries, existing DP techniques (and other related classes of privacy metrics such as $k$-Anonymity [8] and $t$-Closeness [9]) have largely overlooked the impact of data freshness for time-varying databases. Intuitively, as some data has diminishing value 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 will significantly drop as the location data becomes outdated. In a more concrete example in Section III, we show that the accuracy of an adversary's estimate diminishes quickly to a level of no privacy concerns due to data aging, whereas the classical DP framework can only provide a very loose upper
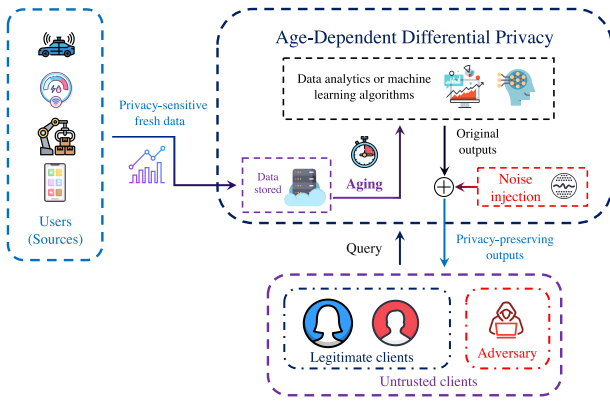
Fig. 1.   An age-dependent differential privacy mechanism.

bound (as high as $88\%$) for the probability an adversary's estimate is correct.

This observation motivates the following key question:

*Question 1:* How should one quantify the impact of data timeliness on data privacy protection?

Motivated by Question 1 and the notions of AoI 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*.

Our new framework is applicable to a wide range of privacy-sensitive applications with time-varying datasets, such as (i) location-based services [10], (ii) smart meter readings [11], (iii) medical information, and (iv) stock holdings in financial markets. Other types of privacy-sensitive information that are time-invariant (e.g., genome information) can be well understood under the classical DP notion, hence are not the main focus of this paper.

Whereas existing studies largely rely on injecting noise to achieve DP [3], our proposed framework provides a new direction to protect privacy, namely *aging*, which postpones the release of outputs. Figure 1 presents an example of an age-dependent differentially private mechanism that combines both the new method (aging) and the existing one (noise injection). This also raises another challenge in designing privacy-perservering mechanisms:

*Question 2:* How should one characterize age-dependent privacy guarantees by leveraging aging along with classical methods (e.g., noise injection)?

To answer this question, we provide theoretic guarantees achieved by classical DP mechanisms when adopted in our new age-dependent DP framework. This establishes a connection between the classical DP notion and our proposed age-dependent generalization. Additionally, 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).

The operation of real-time systems involving frequent data updates raises the necessity of understanding the performance of privacy-preserving mechanisms with sequential queries, a topic that has been studied for conventional DP mechanisms through work on sequential composition [63]. Age-dependent DP mechanisms bring two new challenges. First, our charac-

terization further depends on the *timing* of both inputs (how stale the input database is for all queries) and outputs (when to release the outputs of all queries). In contrast, the classical composition results only depend on the number of queries. The second challenge is that optimally trading off privacy and utility now depends on optimizing over such timing as well as any noise that is added, both of which may impact utility and privacy in different ways.[1] In light of the above challenges, our final key question is

*Question 3:* How should one characterize the optimal trade-offs between privacy and utility for multi-query mechanisms?

To this end, we construct multi-query mechanisms by a combination of single-query mechanisms, and derive the corresponding privacy guarantees over time considering the independent guarantees achieved by these single-query mechanisms and the corresponding timings of inputs and outputs. By deriving and exploiting a special structure of the optimal solutions, we are then able to formulate a tractable optimization problem that makes tradeoffs between privacy and utility.

### B. Contributions

We summarize the key contributions of this work in the following:

- *Age-dependent DP.* To the best of our knowledge, age-dependent DP is the first performance metric of privacy protection that accounts for the potentially diminishing privacy leakage of delayed data.
- *Privacy guarantee analysis.* By establishing the connection between the classical DP notion and our proposed age-dependent DP, we derive the achievable privacy guarantees when one exploits aging along with classical DP mechanisms. We also characterize the key factor that determines the decaying rate of age-dependent privacy risks over time.
- *Composition.* We further address how age-dependent DP guarantees compose over multiple queries, depending on the timing of both inputs and outputs. Even for infinitely many queries, we show that the peak privacy risk can be upper bounded under a characterizable condition.
- *Tradeoffs.* We present two case studies to understand tradeoffs between privacy and utility. Our numerical studies reveal that, when approaching an arbitrary small privacy risk in a single-query case, our proposed scheme (of combing aging and noise injection) only incurs a bounded accuracy loss, whereas the loss incurred by a benchmark (using noise injection only) grows unbounded.

We organize the rest of this paper as follows. In Section II, we review related studies. Section III describes the model. In Section IV, we introduce the definition of age-dependent DP for a single-query mechanism and characterize its properties. In Section V, we generalize our definition and results to

---

[1]As an example, while real-time status information, such as the locations of nearby vehicles, is crucial in autonomous driving, it is worth noting that certain algorithms, like path planning, are less reliant on up-to-the-minute map data. Similarly, machine learning algorithms used in autonomous driving may also exhibit less sensitivity to the timeliness of training datasets. Therefore, by allowing for slightly delayed updates, the system can strike a better balance between privacy preservation and maintaining an acceptable level of utility.

multi-query mechanisms and characterize how age-dependent DP guarantees compose. We further present two heuristic examples in Section VI and finally conclude in Section VII.

## II. RELATED WORK

### A. Age of Information

Many works in recent years considered AoI for various queueing systems (e.g., [34], [35], [36], [37], [38], [61], [62]) and wireless networks (see references [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [51], [52], [53], [54], [57], [60] and a survey in [2]). Many also considered the optimization of scheduling algorithms for wireless networks [43], [44], [52], [53], [54], [56], [57], [60]. For example, Talak et al. [43] studied the optimization of information freshness in wireless networks. Kadota et al. [44] studied the optimization of information freshness in wireless networks with throughput constraints. A few studies also considered the modeling of aged information and proposed variants of AoI (e.g., [45], [58], [59], [61]). For instance, Maatouk et al. in [58] proposed the notion of age of incorrect information, which captures both the values of freshness and correctness. Sun et al. [45] considered a family of general data freshness metrics represented by monotonic functions in AoI. Wang in [59] considered age of critical information, which reflects the information freshness quantified by changes in its critical levels. Yates [61] considered a network of observer nodes, in which timeliness at each node is measured by update versions. Several studies examined the economic value of information freshness from an economic perspective, investigating the interactions between sources and destinations through the integration of pricing and mechanism design [46], [47], [48]. These studies have primarily focused on the operational cost incurred by the frequent generation of status updates, while neglecting the consideration of privacy costs.

Most studies in this literature have not considered privacy protection. One exception is [50], in which Jin et al. studied the an age-minimal mobile crowd sensing system for location-based services while achieving DP. However, [50] did not consider an age-based variant of DP, nor the impact of data freshness or temporal correlation on privacy protection, which we do here.

### B. Differential Privacy

There has been a lot of work on DP (see [3]) and we will only survey some related studies here.

*1) Variants of Differential Privacy:* While the initial work on DP [3] makes no distributional assumption on the data, as shown in the no-free-lunch theorem for privacy [13], it is generally impossible to achieve a better privacy-utility tradeoff (than DP) without making assumptions on the knowledge of data distributions. Therefore, a number of existing variants of DP assumed specific knowledge of data generating distributions, which can provide more meaningful privacy protection under considered scenarios (e.g., [14], [15], [16], [17]). For instance, Rinaldo et al. [14] studied how data distributions affect the characterizations of privacy guarantees. Kifer et al. [17] proposed a Bayesian framework for defining privacy, the

Pufferfish. Yang et al. [15] further studied the influence of data correlations on privacy in the Pufferfish framework. *Our approach assumes the knowledge of an underlying probabilistic model for the time-varying data, whereas these variants mainly did not account for the impact of data timeliness.*

*2) Privacy Analysis of Dynamic Databases:* Related studies of DP for dynamic databases have focused on a limited class of problems (e.g., [19], [23], [24], [25], [30]). Both [19] and [24] adapted the notion of DP to streaming environments, where each entry in the database is a single bit, and bits arrive one per unit time. In [23], Smith et al. extended this technique to maintain private sums of real vectors arriving online in a stream. Mechanisms proposed in [19], [23], and [24] are non-adaptive, in the sense that they answer a single query repeatedly on a dynamic database. Cummings et al. extended in [25] the results to an adaptive analysis. *Nevertheless, this line of work did not characterize the impact of data timeliness or temporal correlation on data privacy protection. Reference [30] is the only exception, in which Cao et al. studied the privacy leakage accounting for continuous data release with time correlation (but not data timeliness). The composition result in [30] did not account for different data publishing and aging policies, which we consider in this work.*

### C. Time-Based Privacy Metrics

Another line of related work considers time-based metrics for privacy, which only includes a few studies [31], [32], [33]. In [32], Wright et al. proposed a time-based metric that measures the time until the adversary's success, which in fact assumed that the adversary will eventually succeed. On the other hand, Hoh et al. [31] considered *the mean time to confusion*, which measures the time during which the adversary's uncertainty stays below a confusion threshold. For an adversary aiming to track a target's location over time, Sampigethaya et al. [33] considered the maximum tracking time to measure the adversary's tracking ability. *These studies did not account for the impacts of the temporal correlation of data or data timeliness on data privacy.*

## III. MODEL AND PRELIMINARY

In this section, we introduce the model of the users, the adversary, and the stochastic processes of the users' privacy-sensitive and time-varying databases. After reviewing the classical notion of DP, we use an illustrative example to show that DP may not provide an informative description of privacy protection for time-varying databases. We summarize all key parameters in Table I.

### A. Model

*1) System Overview:* We consider a set $\mathcal{I} = \{1, \ldots, I\}$ 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 set of all non-negative integers.

TABLE I
NOTATION SUMMARY

| Notation | Definition |
|---|---|
| $P_i(\cdot,\cdot)$ | User $i$'s transition matrix |
| $P_{i,t}(\cdot,\cdot)$ | User $i$'s $t$-step transition matrix |
| $\hat{P}_{i,t}(\cdot,\cdot)$ | User $i$'s $t$-step transition matrix of the reversed process |
| $\pi_i(\cdot)$ | User $i$'s stationary distribution |
| $M(\cdot)$ | Single-query mechanism |
| $\mathcal{M}_t(\cdot,\cdot)$ | Multi-query mechanism |
| $\mathcal{S}$ | Publishing policy |
| $\mathcal{A}$ | Aging policy |
| $\Delta(t)$ | Maximal total variation distance |
| $\epsilon_C$ | Classical DP risk |
| $\epsilon(t)$ | Age-dependent privacy risk |

*2) User:* For each user $i$, its privacy-sensitive data is captured by a discrete-time stationary process $\{x_{i,t}\}_{t\in\mathbb{N}}$.[2] We assume this process take values in a finite state space, $\mathcal{X}_i$. Due to the stationarity, there exists a stationary probability distribution $\pi_i : \mathcal{X}_i \to [0,1]$, such that $\pi_i(x) = \Pr(x_{i,t} = x)$ for all $t \in \mathbb{N}$. We denote the $t$-step transition probability by $P_{i,t}(x,y) \triangleq \Pr[x_{i,t} = y | x_{i,0} = x]$ for all $t \in \mathbb{N}$ and all $x, y \in \mathcal{X}_i$. We further use $\hat{P}_{i,t}(x,y)$ to denote the $t$-step transition probability of the reversed process, which satisfies that

$$\hat{P}_{i,t}(x,y) = \frac{\pi_i(y)P_{i,t}(y,x)}{\pi_i(x)}, \forall x,y \in \mathcal{X}_i, t \in \mathbb{N}, i \in \mathcal{I}. \quad (1)$$

If $\hat{P}_{i,t}(x,y) = P_{i,t}(x,y)$ for all $t \in \mathbb{N}$ and $x, y \in \mathcal{X}_i$, then $\{x_{i,t}\}_{t\in\mathbb{N}}$ is *reversible* [64].

A class of random processes of particular interest is time-invariant Markov chains, which are characterized via a time-invariant transition probability matrix $P_i(x,y)$ for each user $i$ and states $x$ and $y$.

*3) Aggregate Database:* We use $\{X_t\}_{t\in\mathbb{N}}$ to denote the aggregate data process across all users, where $X_t \triangleq \{x_{i,t}\}_{i\in\mathcal{I}}$ is the random variable of the (aggregate) database of all users at time $t$, belonging to the (aggregate) state space $\mathcal{X} = \prod_{i\in\mathcal{I}} \mathcal{X}_i$. Throughout this paper, a *database* is referred to as the aggregate data of all users sampled at a time, $X_t$. Assuming that the processes are independent across users, then the corresponding $t$-step transition probability is $P_t(X,Y) = \prod_{i\in\mathcal{I}} P_{i,t}(x_i,y_i)$, the stationary probability is $\pi(X) \triangleq \prod_{i\in\mathcal{I}} \pi_i(x_i)$, and the $t$-step transition probability of the reversed process is $\hat{P}_t(X,Y)$, for all $X = \{x_i\}_{i\in\mathcal{I}}$, $Y = \{y_i\}_{i\in\mathcal{I}}$, and $t \in \mathbb{N}$.

*4) Event-Level Privacy Protection Goal:* Users seek to protect their data privacy against a type of adversaries that aim to infer about the current status $X_t$ at each time slot $t \in \mathbb{N}$. That is, the key focus is to protect each user's single data point

at time $t \in \mathbb{N}$, which is referred to as the *event-level* privacy protection [19], [30]. Answering a query using database $X_{t'}$ for some $t' < t$ thus leads to less privacy leakage than using database $X_t$, even if $X_{t'}$ and $X_t$ may be correlated.

The consideration of such adversaries is practical in a wide range of application scenarios. Examples include real-time electricity consumption data, real-time GPS data, and financial data. This can be distinguished from *user-level* privacy protection, where an adversary is interested in obtaining the entire history (or its subset) of a user's location database $\{X_t\}_{t\in\mathcal{T}}$ for some $\mathcal{T}$, e.g., to determine if a user ever visited a given location with a given time window. We will leave the modeling of this more general class of adversaries and the corresponding analysis for future work.

In our analysis, we consider the worst-case scenario where the adversary has knowledge of the underlying probability models $\{P_{i,t}(x,y)\}_{i\in\mathcal{I}}$, which provides a strong guarantee. We assume that the platform also has knowledge of $\{P_{i,t}(x,y)\}_{i\in\mathcal{I}}$ and can use it to determine how data can be released. The platform may gain such knowledge, for example, from analyzing historical data from the same set of users.

### B. Differential Privacy

Here we review the classical notion of DP [3], which can be viewed as if applying to databases that consist of a single element from a joint-state space $\mathcal{X}$:

*Definition 1 (Differential Privacy (DP)):* Given any $\epsilon_C > 0$, a mechanism (algorithm) $M : \mathcal{X} \to \mathcal{Y}$ is $\epsilon_C$-DP if for all pairs $X, X' \in \mathcal{X}$ which differ only in one user's data, the following inequality holds

$$\Pr[M(X) \in \mathcal{W}] \le \exp(\epsilon_C)\Pr[M(X') \in \mathcal{W}], \ \forall \mathcal{W} \subset \mathcal{Y}, \quad (2)$$

where the probability is taken over the randomness of the output of mechanism $M$.

We use subscript $C$ to indicate that $\epsilon_C$ stands for the classical privacy risk, so as to distinguish it from the privacy risk achievable in our framework.

DP requires that changing each user's data will have little impact on the output when $\epsilon$ is small. It provides a strong and mathematically rigorous robustness against an arbitrary adversary, but may come at the expense of the statistical utility (e.g., the accuracy) of the output.

To achieve DP, a well-known approach is to inject Laplace or Gaussian noise to the database output [3]. Formally, we review the following definitions:

*Definition 2 (Sensitivity):* For any function $f : \mathcal{X} \to \mathbb{R}^d$, the $\ell_1$-sensitivity of $f$ is defined as

$$S^{(f)} = \max_{X,X'} \|f(X) - f(X')\|_1, \quad (3)$$

where $X$ and $X'$ are neighbouring databases, i.e., $X$ and $X'$ differ only in one user's data.

*Definition 3 (Laplace Mechanism):* For any function $f : \mathcal{X} \to \mathbb{R}^D$, the Laplace mechanism (associated with $f$) is defined as

$$\mathcal{M}_L(X) = f(X) + \boldsymbol{n}, \quad (4)$$

---

[2] We note that data processes exhibit stationary in many applications such as location-based services (e.g., [27], [28], [29], [30]), physical activity measurements and electricity (e.g., [18]). As a preliminary step in understanding age-dependent privacy and developing the corresponding mechanisms, we will confine our analysis to the stationary setting and will leave non-stationary data processes for future work.

where $\boldsymbol{n} = \{n_k\}_{k=1}^{D}$ are independent $\mathrm{Laplace}(S^{(f)}/\epsilon_C)$ random variables. A random variable has a $\mathrm{Laplace}(b)$ distribution if its probability density function is

$$p_{\mathcal{L}}(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right). \tag{5}$$

*Proposition 1 (Classical Privacy Guarantee [3]):* For any function $f : \mathcal{X} \to \mathbb{R}^D$ with the sensitivity given in (3), the corresponding Laplace mechanism is $\epsilon_C$-DP.

*Proof:* Given any two adjacent inputs datasets $X, X' \in \mathcal{X}$ which differ only in one user's data, let $y_1 = f(X)$ and $y_2 = f(X')$. By defining $\mathcal{X}' \triangleq \{y - y_1 : y \in \mathcal{W}\}$, we have:

$$
\begin{aligned}
\frac{\Pr[\mathcal{M}_L(X) \in \mathcal{W}]}{\Pr[\mathcal{M}_L(X') \in \mathcal{W}]} &= \frac{p_{\mathcal{L}}(x|b)}{p_{\mathcal{L}}(x - (y_1 - y_2)|b)} \\
&= \frac{\frac{1}{2b}\exp(-|x|/b)}{\frac{1}{2b}\exp(-|x - (y_1 - y_2)|/b)} \\
&\leq \exp(|y_1 - y_2|/b) \\
&\leq \exp(S^{(f)}/b) = \exp(\epsilon_C), \tag{6}
\end{aligned}
$$

which completes the proof of Proposition 1. ∎

As an example, consider function $f(X) = \sum_{i \in \mathcal{I}} X_i/I$, where $X_i \in \{0, 1\}$ for every $i \in \mathcal{I}$. From Definition 2, we have $S^{(f)} = 1/I$. We use the variance of its outcome to characterize the $L_2$ accuracy loss (mean square error) of the Laplace mechanism in (4), given by $\mathrm{Var}(\mathcal{M}_L(X)) = 2/(\epsilon_C^2 I^2)$. This implies that a higher degree of privacy protection (i.e., a small $\epsilon_C$) requires a larger statistical utility loss (i.e., a larger variance $\mathrm{Var}(\mathcal{M}_L(X))$).

Next we present an example to illustrate that the classical DP privacy guarantee may not be practical for time-varying databases, as it does not capture the impact of data freshness on privacy leakage.

### C. A Motivating Example

In this subsection, we consider an illustrative example to motivate our alternative privacy notion.

Consider the following example.[3] There are two towns: $A$ and $B$. The databases $X, X'$ consist of two profiles of resident income values of town $A$. There is a billionaire $x'$ whose income is orders of magnitude higher than other individuals in either town. The billionaire occasionally travels between two towns, and the probability that the billionaire travels to the other town after each day is $10\%$. An adversary wishes to track down the current location of the billionaire. The adversary observes the mean income $w$ of town $A$ sanitised by an $\epsilon_C$-DP mechanism. The result $w$ is only published at time $t = 0$. Using a Bayesian estimator, the probability of the adversary being correct is

$$P(A) = \frac{1}{1 + \exp(-\epsilon_C)}. \tag{7}$$

For an $\epsilon_C$-DP mechanism with $\epsilon_C = 2$, the attack accuracy is as high as $P(A) = 88\%$.

However, if the adversary obtains this message after $t$ days (still using the same $\epsilon_C$-DP mechanism), we can express the

---

[3]Our example is the same as the motivational example in [16] except that we consider time-varying databases.

eventual accuracy $P_t(A)$ in the following:

$$
\begin{bmatrix} P_t(A) \\ 1 - P_t(A) \end{bmatrix} = \begin{bmatrix} P(A) \\ 1 - P(A) \end{bmatrix} \cdot \begin{bmatrix} 90\% & 10\% \\ 10\% & 90\% \end{bmatrix}^t. \tag{8}
$$

That is, the probability of a successful attack becomes $P_t(A) \approx 65\%$ when $t = 3$ days, $P_t(A) \approx 58\%$ when $t = 6$ days, and $P_t(A) \approx 53\%$ when $t = 10$ days. Therefore, the probability of a successful attack decreases over time and eventually converges to $50\%$ (which corresponds to a pure random guess).

Moreover, the probability of a successful attack when it receives $w$ immediately with $\epsilon_C = 0.3$ is approximately the same as when it observes $w$ after $t = 6$ days with $\epsilon_C = 2$. This shows that aging provides an alternative to increasing the noise variance for increasing privacy protection. This motivates us to design an age-dependent generalization that takes the data timeliness into account in the next sections.

## IV. AGE-DEPENDENT DIFFERENTIAL PRIVACY: THE SINGLE-QUERY CASE

In this section, we start with introducing the definition of *age-dependent differential privacy* for a *single-query mechanism*. In the age-dependent DP framework, we characterize the privacy guarantees achieved by any DP mechanism considering data freshness and the temporal correlation.

### A. Definitions

Throughout Section IV, we focus on *single-query (time-invariant) mechanisms* $M : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} = \prod_{i \in \mathcal{I}} \mathcal{X}_i$ denotes the state space of an aggregate database (at a specific time instance) and $\mathcal{Y}$ is the set of all possible outcomes. In Section V, we will provide an extension to multi-query mechanisms.

We first formally introduce age-dependent DP under the single-query case:

*Definition 4 (Age-Dependent Differential Privacy):* A single-query mechanism $M$ is $(\epsilon, t)$-age-dependent DP for a given random process $\{X_t\}$, if, for any pair $X, X' \in \mathcal{X}$ which differ only in one user's data, the following is true:

$$
\begin{aligned}
\Pr[M(X_0) \in \mathcal{W}|X_t = X] \\
\leq \exp(\epsilon)\Pr[M(X_0) \in \mathcal{W}|X_t = X'], \forall \mathcal{W} \subset \mathcal{Y}, \tag{9}
\end{aligned}
$$

where the probability takes into account the randomness of both the output of mechanism $M$ and the stochastic process of $\{X_t\}_{t \in \mathbb{N}}$.

Different from classical DP (Definition 1), age-dependent DP (Definition 4) also accounts for the evolution of the stochastic process $\{X_t\}_{t \in \mathbb{N}}$. This also implies that the classical $\epsilon_C$-DP is a special case of the age-dependent DP, i.e., $(\epsilon_C, t = 0)$-age-dependent DP. Due to the time homogeneity of mechanism $M$ and the process $\{X_t\}_{t \in \mathbb{N}}$, inequality (9) still holds when we replace $X_0$ and $X_t$ in (9) by $X_k$ and $X_{k+t}$, respectively, for any $k \in \mathbb{N}$.

In (9), we interpret $t$ as the *age* of the output $M(X_0)$. In particular, for a process $\{X_t\}_{t \in \mathbb{N}}$ satisfying some *ergodicity*

or *mixing* properties (see Appendix A), we have $\Pr[X_0|X_t] \approx \pi(X_0)$ as $t$ increases (i.e., $X_0$ becomes more outdated). In other words, as the output data $M(X_0)$ gets outdated, the probabilities conditional on $X_t = X$ and $X_t = X'$ become indistinguishable and hence mechanism $M$ asymptotically becomes perfectly privacy-preserving.

To better understand the connection between DP and age-dependent DP, we introduce a heuristic way to guarantee age-dependent DP by artificially aging. The main idea is to artificially age the input by $t$ units of time and sequentially apply an $\epsilon$-DP mechanism. That is, when provided with an input dataset $X$, we generate a new dataset $X_{-t}$ by employing the reversed process $\hat{P}_t$, i.e., $X_{-t} \sim \hat{P}_t(X, \cdot)$.

*Definition 5 (Artificially Aging Mechanism):* The artificially aging mechanism $M_A^t$ associated to any arbitrary mechanism $M$ and a given random process $\{X_t\}$, is given by

$$M_A^t(X) \triangleq M(X_{-t}), \tag{10}$$

where $X_{-t}$ is a random database generated by employing the reversed process $\hat{P}_t$, i.e., $X_{-t} \sim \hat{P}_t(X, \cdot)$.

The following result formally establishes the connection.

*Corollary 1:* An aging mechanism $M_A^t$ is $\epsilon_C$-DP if and only if $M$ is $(\epsilon_C, t)$-age-dependent DP.

### B. Properties

Age-dependent DP satisfies some basic properties of the classical DP, including post-processing resilience (see Appendix B). In this subsection, we will focus on presenting analytical results to understand how privacy guarantees evolve over time.

*1) Total Variation Distance:* To understand the temporal correlation of databases, we first present an important quantity to measure the statistical distance between two probability distributions.

*Definition 6 (Total Variation Distance):* The *total variation distance* between probability distributions $\mu$ and $\pi$ on a finite set $\mathcal{F}$ is defined as

$$\delta(\mu, \pi) = \max_{\mathcal{A} \subset \mathcal{F}} |\mu(\mathcal{A}) - \pi(\mathcal{A})| = \frac{1}{2} \sum_{A \in \mathcal{F}} |\mu(A) - \pi(A)|. \tag{11}$$

To derive age-dependent privacy guarantees for all mechanisms that satisfy the classical DP, we also need the following definition:

*Definition 7 (Maximal Total Variation Distance):* The maximal total variation distance $\Delta(t)$ (between any user's two $t$-step transition probability distributions of the reversed process) is defined as

$$\Delta(t) \triangleq \max_{i \in \mathcal{I}} \max_{x_{i,0}, x'_{i,0} \in \mathcal{X}_i} \delta\left(\hat{P}_{i,t}(x_{i,0}, \cdot), \hat{P}_{i,t}(x'_{i,0}, \cdot)\right), \forall t \in \mathbb{N}, \tag{12}$$

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

We will use $\Delta(t)$ to characterize the privacy risk, i.e., $\epsilon$ in Definition 4. A wide range of stochastic processes have a diminishing value of $\Delta(t)$ over time at some specific rate.

As an example, a wide range of Markov chains satisfy the following related and extensively studied property [65], [66]:

*Definition 8 (Geometric Ergodicity [65]):* A Markov chain $\{X_t\}_{t \in \mathbb{N}}$ with stationary distribution $\pi(\cdot)$ has the geometric ergodicity property if, for all $t \in \mathbb{N}$ and $Y \in \mathcal{X}$,

$$|\Pr(X_t = Y|X_0 = X) - \pi(Y)| \le a(X) \cdot \rho^{-t}, \tag{13}$$

for some function $a(X)$ and some decay coefficient $\rho < 1$.

In addition to ergodicity, another related notation with a potential property leading to a diminishing value of $\Delta(t)$ is *mixing*. We present details in Appendix A.

*2) Mechanism-Dependent Guarantee:* We start with the following theorem to characterize the formal privacy protection guarantee:

*Theorem 1 (Mechanism-Dependent Guarantee):* If a mechanism $M$ is $\epsilon_C$-DP, then it is also $(\epsilon(t), t)$-age-dependent DP, where $\epsilon(t)$ satisfies

$$\epsilon(t) = \ln\left(1 + \Delta(t) \cdot (\exp(\epsilon_C) - 1)\right), \ \forall t \in \mathbb{N}. \tag{14}$$

Theorem 1 characterizes an age-dependent privacy guarantee based on the maximal total variation distance $\Delta(t)$ of the underlying process and the classical DP guarantee $\epsilon_C$. The proof of Theorem 1 mainly involves showing that two transition probability matrices that are close (i.e., $\Delta(t)$ is small) lead to a bounded difference in the likelihoods of the observed value. We defer the complete proof to Section IV-C.

The significance of Theorem 1 is two-fold. First, it establishes the connection between the privacy guarantees achieved by classical DP and age-dependent DP. This provides a methodology to attain privacy guarantees in (14), by combining noise injection and aging. Second, it indicates that we only need the maximal total variation distance $\Delta(t)$ of the process $\{X_t\}_{t \in \mathbb{N}}$ to characterize such a bound. Estimating $\Delta(t)$ is not challenging in practice, and it can be done, for example, by referring to [72]. Furthermore, based on the Pinsker's inequality, we can bound the total variation distance between any two probability distributions $\mu$ and $\pi$ the Kullback–Leibler (KL) divergence [73]:

$$\delta(\mu, \pi) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(\mu \parallel \pi)}. \tag{15}$$

It is worth noting that the KL divergence is a well-studied metric, and its estimation has been extensively explored in the literature (e.g., [74]). Note that Theorem 1 holds for any arbitrary stationary processes $\{X_t\}_{t \in \mathbb{N}}$. When a process has some additional properties, such as satisfying 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 $\lim_{x \to 0} \ln(1+x)/x = 1$, (14) further implies that the age-dependent privacy risk $\epsilon(t)$ converges to zero at the same rate as $(\exp(\epsilon_C) - 1)\Delta(t)$. Finally, as we will demonstrate in Section VI, aging along with noise injection may not necessarily reduce the data utility comparing with noise injection alone.

It is also possible to achieve age-dependent DP without requiring mechanism $M$ to be $\epsilon_C$-DP, as we show next.

*Theorem 2 (Mechanism-Independent Guarantee):* If each agent $i$'s time-varying database is an irreducible, aperiodic, and reversible Markov chain $\{x_{i,t}\}_{t \in \mathbb{N}}$, any mechanism $M$ is $(\epsilon(t), t)$-age-dependent DP satisfying $\epsilon(t) = \mathcal{O}(\lambda_*^t)$. Here, $\lambda_* = \max\{\lambda_1, |\lambda_{m-1}|\}$, where $\lambda_1$ and $\lambda_{m-1}$ are specific eigenvalues of $P$ (see Lemma 3 in Appendix A).

We present the proof of Theorem 2 in Appendix D. In Theorem 2, we use a different proof technique from the one we use in Theorem 1, which cancels out the impact of the degree of DP achieved by $M$.

Theorem 2 implies that, even without noise injection, using aging alone can ensure an age-dependent privacy guarantee with a similar convergence rate.

### C. Proof of Theorem 1

In this subsection, we present the complete proof of Theorem 1. We first introduce lemmas to characterize the change in expected value when one introduces a small change in probability distribution (based on the maximal total variation distance $\Delta(t)$). We then bound the privacy risk $\epsilon(t)$ based on the privacy risk of DP, $\epsilon_C$, and $\Delta(t)$.

We start with an expected value version of a DP-like probability distribution bound:

*Lemma 1:* If $M$ is $\epsilon_C$-DP, then for any $z_i, z_i' \in \mathcal{X}_i$, we have

$$\mathbb{E}_{\mathbf{z}_{-i} \sim \hat{P}_{-i,t}(\mathbf{x}_{-i}, \cdot)}[\Pr(\mathcal{M}(z_i, \mathbf{z}_{-i}) \in \mathcal{W})]$$
$$\leq \exp(\epsilon_C)\mathbb{E}_{\mathbf{z}_{-i} \sim \hat{P}_{-i,t}(\mathbf{x}_{-i}, \cdot)}[\Pr(\mathcal{M}(z_i', \mathbf{z}_{-i}) \in \mathcal{W})], \quad (16)$$

where $\hat{P}_{-i,t}(\mathbf{x}_{-i}, \mathbf{z}_{-i}) = \prod_{j \neq i} \hat{P}_t(x_j, z_j)$ for all $t \in \mathbb{N}$, $\mathbf{x}_{-i}$, $\mathbf{z}_{-i}$, and $i \in \mathcal{I}$.

We present the proof in Appendix C.

For presentation simplicity, we define $f(z_i) = \mathbb{E}_{\mathbf{z}_{-i} \sim \hat{P}_{-i,t}(\mathbf{x}_{-i}, \cdot)}[\Pr(\mathcal{M}(z_i, \mathbf{z}_{-i}) \in \mathcal{W})]$ and $p(z_i) = \hat{P}_{i,t}(x_i', z_i)$ for all $i \in \mathcal{I}$. We set $\underline{z}_i \triangleq \arg\min_{z_i \in \mathcal{X}_i} f(z_i)$ and $\bar{z}_i \triangleq \arg\max_{z_i \in \mathcal{X}_i} f(z_i)$. Therefore, Lemma 1 implies that

*Corollary 2:* If $M$ is $\epsilon_C$-DP, then every $z_i \in \mathcal{X}_i$ satisfies $1 \leq f(z_i)/f(\underline{z}_i) \leq \exp(\epsilon_C)$.

The following Lemma 2 characterizes an upper bound for the change in the expected value of $x_i \in \mathcal{X}_i$ when the probability distribution has a small change (characterized by the maximal total variation distance $\Delta(t)$):

*Lemma 2:* Assume that every $x_i \in \mathcal{X}_i$ satisfies $\underline{x}_i \leq x_i \leq \bar{x}_i$. For a probability distribution $p(x_i)$ satisfying $\sum_{x_i \in \mathcal{X}_i} p(x_i) = 1$, consider the following optimization problem:

$$\max_{\boldsymbol{\delta} = \{\delta(x_i) : x_i \in \mathcal{X}_i\}} \sum_{x_i \in \mathcal{X}_i} (p(x_i) + \delta(x_i))x_i, \quad (17a)$$

$$\text{s.t.} \quad \frac{1}{2} \sum_{x_i \in \mathcal{X}_i} |\delta(x_i)| = \Delta(t), \sum_{x_i \in \mathcal{X}_i} \delta(x_i) = 0. \quad (17b)$$

The optimal solution is $\delta^*(\bar{x}_i) = \Delta(t)$, $\delta^*(\underline{x}_i) = -\Delta(t)$, and $\delta^*(x_i') = 0$ for all other $x_i' \in \mathcal{X}_i$. The maximal objective value of (17) is

$$\sum_{x_i \in \mathcal{X}_i} p(x_i)x_i + \Delta(t)(\bar{x}_i - \underline{x}_i). \quad (18)$$

*Proof:* We prove Lemma 2 by contradiction. Suppose that there exists an optimal solution $\boldsymbol{\delta}^*$ such that $\delta^*(x_i) \neq 0$ for a $x_i \notin \{\bar{x}_i, \underline{x}_i\}$. We consider the following two cases:

- Suppose that there exists an $x_i' \notin \{\bar{x}_i, \underline{x}_i\}$ such that $\delta^*(x_i') > 0$. We can always construct a new solution $\bar{\boldsymbol{\delta}} = \{\bar{\delta}(x_i) : x_i \in \mathcal{X}_i\}$ such that $\bar{\delta}(x_i') = 0$, $\bar{\delta}(\bar{x}_i) = \delta^*(x_i') + \delta^*(\bar{x}_i)$, and $\bar{\delta}(x_i) = \delta^*(x_i)$ for all other $x_i \notin \{\bar{x}_i, x_i'\}$. It is easy to check that $\bar{\boldsymbol{\delta}}$ is feasible. In addition, since $x_i' < \bar{x}_i$, we see $\sum_{x_i \in \mathcal{X}_i} x_i \delta^*(x_i) < \sum_{x_i \in \mathcal{X}_i} x_i \bar{\delta}(x_i)$. Therefore, it contradicts with existence of $x_i' \notin \{\bar{x}_i, \underline{x}_i\}$ such that $\delta^*(x_i') > 0$ at the optimal solution.

- Suppose that there exists an $x_i' \notin \{\bar{x}_i, \underline{x}_i\}$ such that $\delta^*(x_i') < 0$. We can always construct a new solution $\bar{\boldsymbol{\delta}} = \{\bar{\delta}(x_i) : x_i \in \mathcal{X}_i\}$ such that $\bar{\delta}(x_i') = 0$, $\bar{\delta}(\underline{x}_i) = \delta^*(x_i') + \delta^*(\underline{x}_i)$, and $\bar{\delta}(x_i) = \delta^*(x_i)$ for all other $x_i \notin \{\underline{x}_i, x_i'\}$. It is easy to check that $\bar{\boldsymbol{\delta}}$ is feasible. In addition, since $\underline{x}_i < x_i'$, we see $\sum_{i \in \mathcal{I}} x_i \delta^*(x_i) < \sum_{i \in \mathcal{I}} x_i \bar{\delta}(x_i)$. Therefore, it contradicts with the existence of $x_i' \notin \{\bar{x}_i, \underline{x}_i\}$ such that $\delta^*(x_i') < 0$ at the optimal solution.

Combining the above two cases, we show that all optimal solutions $\boldsymbol{\delta}^*$ satisfy $\delta^*(x_i') = 0$ for all $x_i' \notin \{\bar{x}_i, \underline{x}_i\}$. It is readily verified that such an optimal solution should be $\delta^*(\bar{x}_i) = \Delta(t)$, $\delta^*(\underline{x}_i) = -\Delta(t)$, which proves Lemma 2. ∎

That is, for any probability distribution $p'$ on $\mathcal{X}_i$, such that that total variation distance satisfies $\delta(p, p') \leq \Delta(t)$, Lemma 2 provides an upper bound of $\sum_{x_i \in \mathcal{X}} p'(x_i)x_i$ in (18).[4]

For all $z_i \in \mathcal{X}_i$, all pairs of $x_i'$ and $x_i \in \mathcal{X}$, and all $\hat{P}_{i,t}$ with the maximal total variation distance given in (12), it follows that, for all users $i$,

$$\ln\left(\frac{\Pr[M(X_0) \in \mathcal{W}|X_t = X]}{\Pr[M(X_0) \in \mathcal{W}|X_t = X']}\right)$$
$$= \ln\left(\frac{\mathbb{E}_{z_i \sim \hat{P}_{i,t}(x_i, \cdot)}[f(z_i)]}{\mathbb{E}_{z_i \sim \hat{P}_{i,t}(x_i', \cdot)}[f(z_i)]}\right)$$
$$\overset{(a)}{\leq} \ln\left(\frac{\sum_{z_i \in \mathcal{X}_i} p(z_i)f(z_i) + \Delta(t)f(\bar{z}_i) - \Delta(t)f(\underline{z}_i)}{\sum_{z_i} p(z_i)f(z_i)}\right)$$
$$\overset{(b)}{\leq} \ln\left(\frac{\sum_{z_i \in \mathcal{X}_i} p(z_i)f(\underline{z}_i) + \Delta(t)f(\bar{z}_i) - \Delta(t)f(\underline{z}_i)}{\sum_{z_i \in \mathcal{X}_i} p(z_i)f(z_i) - \sum_{z_i \in \mathcal{X}_i} p(z_i)(f(z_i) - f(\underline{z}_i))}\right)$$
$$= \ln\left(\frac{\Delta(t)f(\bar{z}_i) + (1 - \Delta(t))f(\underline{z}_i)}{f(\underline{z}_i)}\right)$$
$$= \ln\left(1 + \frac{\Delta(t)f(\bar{z}_i) - \Delta(t)f(\underline{z}_i)}{f(\underline{z}_i)}\right)$$
$$\overset{(c)}{\leq} \ln\left(1 + \Delta(t)(\exp(\epsilon_C) - 1)\right), \quad (19)$$

where $(a)$ is from Lemma 2, $(b)$ is due to the fact that $a/b < (a-c)/(b-c)$ when $a > b > c > 0$, and $(c)$ is from Corollary 2. This completes the proof of Theorem 1.

### D. Summary

This section introduced the notion of age-dependent DP. We characterized how the age-dependent privacy risk evolves

---

[4]In this case, (18) is only an upper bound, but not necessarily the same maximal value, as we still need to consider the non-negative constraint that $1 \geq p'(x_i) \geq 0$ for all $x_i \in \mathcal{X}_i$, which is not captured by Lemma 2.
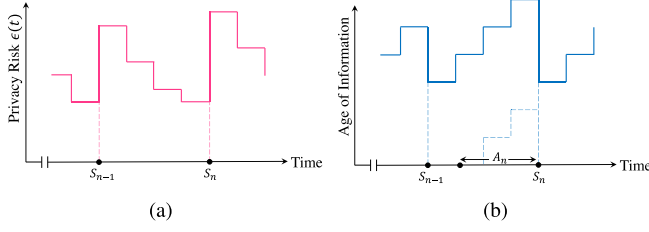
Fig. 2. Illustration of (a) privacy risks of age-dependent DP and (b) age of information for a multi-query mechanism.

over time and how to achieve age-dependent DP by exploiting the classical DP and aging. Our analysis reveals that the key factor of $\{X_t\}_{t \in \mathbb{N}}$ that determines the decaying rate is $\Delta(t)$. The analysis in this section is only applicable to a single-query mechanism $M$, whereas the deployment of real-time applications in practice relies on publishing data updates frequently. This motivates us to study the more general case of multiple queries next.

## V. AGE-DEPENDENT DIFFERENTIAL PRIVACY: SEQUENTIAL COMPOSITION

In this section, we introduce a more general notion of age-dependent DP for multi-query mechanisms (e.g., for real-time applications that publish updates frequently) and present a sequential composition theorem. We will further discuss how to best trade off privacy and utility (characterized by age of information).

### A. The General Definition and Composition

In a multi-query scenario, the history of all published outputs are visible to the adversary, which brings a new challenge in characterizing the privacy risk over time. A fundamental problem in the DP literature is how the overall privacy level degrades when combining multiple queries, where each query meets a certain DP guarantee. Such a combination of multiple queries is known as *composition*. The characterization of classical sequential composition results typically relies on the total number of queries (e.g., [63]). As an example, Dwork in [3] showed that:

*Proposition 2 (Basic Composition [3]):* Let $M_n$ be an $\epsilon_{C,n}$-DP mechanism for all $n \in \{1, \ldots, N\}$. The composition $M'(X)$ defined as $M'(X) = \{M_n(X)\}_{n=1}^{N}$ is $\epsilon_C$-DP for $\epsilon_C = \sum_{n=1}^{N} \epsilon_{C,n}$.

Proposition 2 can only provide a loose privacy guarantee in our case since it does not capture the impacts of data freshness and the temporal correlation. Another challenge of analyzing time-varying databases comes from the need to consider timings of both inputs and outputs: *how stale is each input database used for each query* and *when each output is published*.

We construct a multi-query mechanism based on incorporating an *aging policy* and a *publishing policy* into potentially infinitely many single-query mechanisms. The aging policy determines the age of each input dataset, while the publishing policy dictates when the outcome of each individual query mechanism should be made public. Specifically, letting $n$

denote the query index, we use $\mathcal{A} = \{A_n\}_{n \in \mathbb{N}}$ to denote the aging policy, where $A_n$ indicates the age of the $n$-th input database used for the $n$-th query. We further use $\mathcal{S} = \{S_n\}_{n \in \mathbb{N}}$ to denote the publishing policy, i.e., $S_n$ represents the time to publish the $n$-th outcome. That is, $S_n - A_n$ representing the time stamp of the input dataset used for the $n$-th query as shown in Fig. 2, and the outcome of the $n$-the query is denoted by $M_n(X_{S_n - A_n})$. Given a publishing policy, we call the time interval $[S_n, S_{n+1})$ as the $n$-th *epoch*.

Further, given the aging policy and the publishing policy, let $\mathcal{H}_t$ be the history of all used input datasets up to time $t$, given by

$$\mathcal{H}_t(\mathcal{S}, \mathcal{A}) \triangleq \{X_{S_n - A_n} : n \in \mathbb{N}, S_n \leq t\}, \quad (20)$$

and let

$$N_t \triangleq \sum_{n \geq 0} \mathbb{1}(S_n \leq t) \quad (21)$$

be the number of queries $N_t$ up to time $t$. We now define the general class of (potential multi-query) mechanisms as follows:

*Definition 9 (Multi-query Mechanism):* A multi-query mechanism $\mathcal{M}_t$ is given by

$$\mathcal{M}_t(\mathcal{H}_t(\mathcal{S}, \mathcal{A})) \triangleq \{M_n(X_{S_n - A_n}) : n \in \mathbb{N}, S_n \leq t\}, \quad (22)$$

where $M_n : \mathcal{X} \to \mathcal{Y}$ is a single-query mechanism used for the $n$-th query.

Note that we label the multi-query mechanism with a time stamp $t$ so that $\mathcal{M}_t(\mathcal{H}_t(\mathcal{S}, \mathcal{A}))$ stands for the history of all outputs published no later than time $t$. We next present the general definition of age-dependent DP associated to $\mathcal{M}_t$, characterized by not only $t$ and $\epsilon$, but also the publishing and the aging policies:

*Definition 10 (Age-Dependent Differential Privacy):* A multi-query mechanism $\mathcal{M}_t$ is $(\epsilon(t), t, \mathcal{S}, \mathcal{A})$-age-dependent DP for a given random process $\{X_t\}$ if for all $t$, the following inequality holds:

$$\Pr[\mathcal{M}_t(\mathcal{H}_t(\mathcal{S}, \mathcal{A})) \in \mathcal{W}_t | X_t = X]$$
$$\leq \exp(\epsilon) \Pr[\mathcal{M}_t(\mathcal{H}_t(\mathcal{S}, \mathcal{A})) \in \mathcal{W}_t | X_t = X'], \quad (23)$$

for each pair $X, X' \in \mathcal{X}$ which differ only in one user's data, and for all output histories $\mathcal{W}_t \subset \mathcal{Y}^{N_t}$. We let $S_0 = 0$ and $\epsilon(t) = 0$ for all $t \in [0, S_1)$.

To characterize how the information freshness evolves over time given $\mathcal{S}$ and $\mathcal{A}$, we consider the following definition of age of information:

*Definition 11 (Age of Information [2]):* Given the publishing policy $\mathcal{S}$ and the aging policy $\mathcal{A}$, we define age of information AoI(t) as

$$\text{AoI}(t+1) \triangleq \begin{cases} A_n, & \text{if } t+1 = S_n, \\ \text{AoI}(t) + 1, & \text{otherwise.} \end{cases} \quad (24)$$

In other words, age of information represents the time elapsed since the time stamp of the input database for the most recently published output. It has been used to estimate the value (e.g., accuracy) of the output of $M_n(X_{S_n - A_n})$ [2].

Before we present the formal characterization of how the age-dependent privacy guarantee evolves over time for any multi-query mechanisms, we present an illustrative example of what the age of information in (24) and the privacy risk guarantees may look like. As shown in Fig. 2, the privacy risk $\epsilon(t)$ decreases within each epoch (for any $\Delta(t)$ diminishing in $t$) and spikes at the beginning of the next epoch. Therefore, we define $\epsilon(S_n)$ as the *in-epoch peak privacy risk* for the $n$-th epoch. On the other hand, the age of information increases within each epoch.

*Theorem 3 (Composition):* If the process $\{X_t\}_{t\in\mathcal{N}}$ is Markovian, for any multi-query mechanism $\mathcal{M}_t$ given in (22), in which each single-query mechanism $M_n$ is $\epsilon_{C,n}$-DP for all $n \in \mathbb{N}$, then mechanism $\mathcal{M}_t$ is $(\epsilon(t), t, \mathcal{S}, \mathcal{A})$-age-dependent DP, where

$$\epsilon(t) = \ln\left(1 + \Delta\left(t - S_n + A_n\right)\right.$$
$$\left. \times \left(\exp(\epsilon_{C,n} + \epsilon(S_n - A_n)) - 1\right)\right), \quad (25)$$

for all $t \in [S_n, S_{n+1}), n \in \mathbb{N}$, where $\Delta(t)$ is given in (12).

We present the proof of Theorem 3 in Appendix E. The result in Theorem 3 has a recurrent form: the privacy risk $\epsilon(t)$ in the $n$-th epoch (i.e., the time interval $[S_n, S_{n+1})$ depends on some previous in-epoch privacy risk, $\epsilon(S_n - A_n)$), as well as the privacy risk of the $n$-th (single-query) mechanism, $\epsilon_{C,n}$. In addition, if $\Delta(t)$ decreases in $t$ (as for aperiodic and irreducible Markov chains), then $\epsilon(t)$ decreases in $t$ during each epoch $[S_n, S_{n+1})$, as shown in Fig. 2.

Intuitively, similar to Theorem 1, the maximal total variation distance $\Delta(t - S_n + A_n)$ characterizes the impact of data timeliness on the privacy guarantee. Theorem 1 generalizes the result in Proposition 2, as (25) becomes $\epsilon(t) = \epsilon_{C,n} + \epsilon(S_n - A_n)$ when $\Delta(t - S_n + A_n) = 1$ (as in a static database). This result in fact coincides with the basic composition result (for static databases). On the other hand, several existing studies have provided state-of-the-art improvements on composition (e.g., [25], [63]) by considering more sophisticated composition techniques (e.g. adaptive composition as in [25]). Hence, it is also possible to obtain a tighter bound than (25) by combining sophisticated composition techniques and the impact of data timeliness, which will be left for future work.

### B. Tradeoff Between Privacy and Utility

Theorem 3 only characterizes a privacy risk for each specific time. To facilitate our analysis of tradeoffs, we need another metric to capture the overall privacy protection performance over the entire time horizon. We thus consider the (overall) peak privacy risk:[5]

$$\epsilon^* \triangleq \sup_n \epsilon(S_n). \quad (26)$$

To make the optimal tradeoff between the privacy and utility, we let $\bar{S}_n \triangleq S_{n+1} - S_n$ denote the inter-publish time between the $n$-th and the $(n+1)$-th updates, for all $n \in \mathbb{N}$.

[5]We note that the literature of age of information has extensively used *peak age of information* as their overall performance metric as well. [2].

We further define a (noise-aware) peak age penalty, denoted by

$$f(\max_t \mathrm{AoI}(t), \epsilon_C) = f(A + \bar{S}, \epsilon_C), \quad (27)$$

where $f(t, \epsilon)$ is increasing in $t$ and decreasing in $\epsilon$. We note that the peak age satisfies $\max_t \mathrm{AoI}(t) = A + \bar{S}$. Specifically, from [2, Sec. II-B], the peak age is given by the sum of the expected interarrival time ($\bar{S}$ in our case) and the expected system time ($A$ in our case). Function $f$ measures the accuracy loss of $M_t(\mathcal{H}_t(\mathcal{S}^E, \mathcal{A}^E))$ due to data staleness (characterized by the peak age of information $A + \bar{S}$) and the injected noise (characterized by $\epsilon_C$).

We consider the following optimization problem:

$$\min \quad \sup_{n\in\mathbb{N}} \epsilon(S_n) \quad (28a)$$
$$\text{s.t.} \quad f(A_n + \bar{S}_n, \epsilon_{C,n}) \leq \bar{f}, \forall n \in \mathbb{N}, \quad (28b)$$
$$\text{var.} \quad \epsilon_n \geq 0, A_n \geq 0, \bar{S}_n \geq 0, \forall n \in \mathbb{N}, \quad (28c)$$

where $\bar{f}$ represents the maximum allowable penalty defined in (28b). We introduce (28b) to facilitate the trade-offs between privacy and utility. The challenge in solving Problem (28) mainly lies in the difficulty of dealing with the recurrent form of $\epsilon(S_n)$. To optimally solve Problem (28), we will show in the following that a specific class of solutions $(\mathcal{S}, \mathcal{A}, \epsilon)$ are optimal. We can then express $\sup \epsilon(S_n)$ in a closed form.

Specifically, we consider the following simplified mechanism, defined as:

*Definition 12 (Simplified Multi-Query Mechanism):* A simplified multi-query mechanism $(\mathcal{S}^E = \{S_n^E\}_{n\in\mathbb{N}}, \mathcal{A}^E = \{A_n^E\}_{n\in\mathbb{N}})$ is a mechanism in which

$$A_n^E = A, \quad S_{n+1}^E - S_n^E = \bar{S}, \text{ and } \epsilon_{C,n} = \epsilon_C, \ \forall n \in \mathbb{N}, \quad (29)$$

for some $(A, \bar{S}, \epsilon_C)$.

In other words, a simplified mechanism publishes updates periodically, ages updates equally, and uses the same single-query DP mechanism for each update. By exploiting the monotonicity of $\epsilon(S_n)$ in $n$ and the existence of an optimal solution that have identical values of $(A_n^*, \bar{S}_n^*, \epsilon_{C,n}^*)$, we can prove the existence of an optimal solution satisfying (29) in the following theorem:

*Theorem 4:* If the process $\{X_t\}_{t\in\mathcal{N}}$ is Markovian, and $\Delta(t)$ decreases in $t$, then there exists an optimal solution to (28) that is a simplified multi-query mechanism, i.e., $A_n^* = A^*, \bar{S}_n^* = \bar{S}^*$ and $\epsilon_{C,n}^* = \epsilon_C^*$ for some $(A^*, \bar{S}^*, \epsilon_C^*)$.

We present the complete proof of Theorem 4 in Appendix G. Theorem 4 indicates that there is no need to vary polices across different epochs, and hence we only need to focus on optimizing three variables $(A, \bar{S}, \epsilon_C)$.

In light of Theorem 4, the following shows the peak privacy risk in (26) for a simplified multi-query mechanism:

*Proposition 3 (Peak Privacy Risk):* If the process $\{X_t\}_{t\in\mathbb{N}}$ is Markovian and $\Delta(t)$ decreases in $t$, for a simplified multi-query mechanism $\mathcal{M}_t(\mathcal{H}_t(\mathcal{S}^E, \mathcal{A}^E))$ with an equal-spacing policy $(\mathcal{S}^E, \mathcal{A}^E)$ and each single-query mechanism $M_n$ being $\epsilon_C$-DP for all $n \in \mathcal{N}$, the peak privacy risk
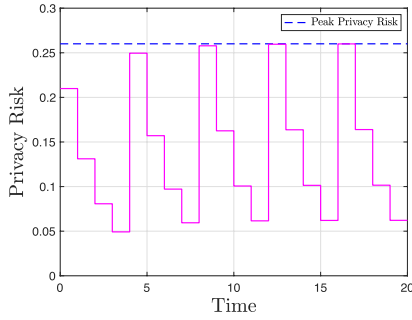
Fig. 3. An illustrative example of the privacy risk $\epsilon(t)$ and the peak privacy risk for equal-spacing policies. We set $A = 2$, $\bar{S} = 4$, and $\epsilon_C = 0.5$.

is given by the unique fixed point satisfying

$$\epsilon^* = \ln\left(1 + \frac{\Delta(A) \cdot (\exp(\epsilon_C) - 1)}{1 - \Delta(\bar{S}) \exp(\epsilon_C)}\right), \quad (30)$$

whenever

$$\Delta(\bar{S}) \exp(\epsilon_C) < 1. \quad (31)$$

Furthermore, the following fixed-point iteration constitutes a contraction mapping:

$$\epsilon(S_{n+1} - A)$$
$$= \ln\left(1 + \Delta(\bar{S}) \cdot (\exp(\epsilon_C) \exp(\epsilon(S_n - A)) - 1)\right). \quad (32)$$

From (30), we note that $\Delta(\bar{S}) \exp(\epsilon_C) < 1$ is the condition for the existence of a finite value of $\epsilon^*$. That is, when the inter-publishing time $\bar{S}$ is sufficiently long and the privacy leakage for each query $\epsilon_C$ is sufficiently small, then such a unique fixed point in (30) exists. Otherwise, the privacy risk $\epsilon(S_n)$ diverges to infinity.

*Proof:* First, define

$$F(\epsilon) \triangleq \ln\left(1 + \Delta(\bar{S}) \cdot (\exp(\epsilon_C) \exp(\epsilon) - 1)\right). \quad (33)$$

It follows that

$$0 < \frac{dF(\epsilon)}{d\epsilon} = \frac{\Delta(\bar{S}) \cdot \exp(\epsilon_C) \exp(\epsilon)}{\Delta(\bar{S}) \cdot (\exp(\epsilon_C) \exp(\epsilon) - 1) + 1} < 1, \quad (34)$$

which implies that $F(\epsilon)$ has a Lipschitz constant $0 < L < 1$. Therefore, (32) is a contraction mapping. This implies that, for a simplified multi-query mechanism $\mathcal{M}_t(\mathcal{H}_t(\mathcal{S}^E, \mathcal{A}^E))$, $\lim_{\to\infty} \epsilon(S_n - A)$ converges and so does $\lim_{\to\infty} \epsilon(S_n)$. Let $\epsilon^* = \lim_{n\to\infty} \epsilon(S_n)$ and $\epsilon' \triangleq \lim_{n\to\infty} \epsilon(S_n - A)$. Substituting $S_n$ and $S_{n+1} - A$ into $t$ in (25), we have

$$\frac{\exp(\epsilon^*) - 1}{\exp(\epsilon') - 1} = \frac{\Delta(A)}{\Delta(\bar{S})}. \quad (35)$$

Following Theorem 3, we have

$$\exp(\epsilon^*) - 1$$
$$= \Delta(A)(\exp(\epsilon_C) \exp(\epsilon') - 1)$$
$$= \Delta(A)\left(\exp(\epsilon_C)\left(1 + \frac{\Delta(\bar{S})}{\Delta(A)}(\exp(\epsilon^*) - 1)\right) - 1\right), \quad (36)$$

which implies that

$$(\exp(\epsilon^*) - 1)(1 - \Delta(\bar{S}) \exp(\epsilon_C)) = \Delta(A)(\exp(\epsilon_C) - 1)$$

$$\implies \exp(\epsilon^*) = 1 + \frac{\Delta(A)(\exp(\epsilon_C) - 1)}{1 - \Delta(\bar{S}) \exp(\epsilon_C)}. \quad (37)$$

We note that $\epsilon(t) \geq 0$ for all $t \geq 0$ from Definition 10. Therefore, from (33), we must have $\frac{\Delta(A)(\exp(\epsilon_C)-1)}{1-\Delta(\bar{S})\exp(\epsilon_C)} \geq 0$, implying (31).

We completed the proof. ∎

We present an illustration in Figure 3. We observe that the in-epoch peak privacy risk $\epsilon(S_n)$ increases in $n$ and converges to the fixed point in (30). It only takes 5 epoches for the privacy risk to approximately attain the fixed point.

### C. Optimization Algorithms

In the following, we consider an optimization problem that accounts for tradeoffs between privacy and utility. Since many practical stochastic processes may have a geometrically decaying $\Delta(t)$ as we have shown in Appendix A and Section IV, we are motivated to focus on the following specific form:

$$\Delta(t) = c \cdot \rho^t, \forall t \in \mathbb{N}, \quad (38)$$

for coefficients $\rho \in (0, 1)$ and $c \geq 1$.

Instead of (28), we consider the following peak age risk minimization problem:

$$\min \quad \ln\left(1 + \frac{c\rho^A(\exp(\epsilon_C) - 1)}{1 - c\rho^{\bar{S}} \exp(\epsilon_C)}\right) \quad (39a)$$

$$\text{s.t.} \quad c\rho^{\bar{S}} \exp(\epsilon_C) < 1, \quad (39b)$$

$$f(A + \bar{S}, \epsilon_C) \leq \bar{f}, \quad (39c)$$

$$\text{var.} \quad \epsilon_C \geq 0, A \geq 0, \bar{S} \geq 0, \quad (39d)$$

where the constraint in (39b) is from (31) in order to make sure that the denominator of the term in (39a) remains positive and the peak age-of-information penalty constraint in (39c) serves the purpose of making tradeoffs. Note that we can drop constraint $\bar{S} \geq 0$ because (39b) implies that $\bar{S}$ must be positive. For trackability in (39), we relax the integer constraints on $A$ and $\bar{S}$. After obtaining the optimal solution $(\epsilon_C^*, A^*, \bar{S}^*)$ to (39), we can round $A^*$ and $\bar{S}^*$ to their respective nearest integers to obtain an approximate solution.

*Proposition 4:* When the constraint, $A \geq 0$, is not binding, the optimal solution to Problem (39) satisfies

$$\rho^{\bar{S}} \exp(\epsilon_C) = \frac{1}{2c}, \quad (40)$$

or equivalently, $\ln(\rho^{-1})\bar{S} = \ln(2c) + \epsilon_C$.

This shows that when aging is used ($A > 0$), $\bar{S}$ should decrease as $\epsilon_C$ decreases. We prove Proposition 4 by exploiting the Karush–Kuhn–Tucker conditions of Problem (39), and present the proof in Appendix F.

In light of Proposition 4, we solve the reduced optimization problems of (39) in the following two cases, depending on whether constraint $A \geq 0$ in (39) is binding or not:

*1) When the Constraint $A \geq 0$ Is Not Binding:* Substituting (40) into Problem (39), we have the following equivalent reformulated problem:

$$\min \quad \rho^A \cdot (\exp(\epsilon_C) - 1) \quad (41a)$$

$$\text{s.t.} \quad \ln(\rho^{-1})\bar{S} = \ln(2c) + \epsilon_C, \quad (41b)$$

---

**Algorithm 1** Heuristic Solution to (41)

---

**1** Initialize $\phi$, $K$, and $\bar{\epsilon}$; **for** $k = \{1, 2, \ldots, K\}$ **do**

**2**    Set $\epsilon_C(k) = \frac{\bar{\epsilon}k}{K}$ and
$\bar{S}(k) = (\ln(2c) + \epsilon_C(k))/\ln(\rho^{-1})$;
   // Exhaustive search for $\epsilon_C$

**3**    Set $A_L = 0$ and $A_H = \tilde{A}$, where $\tilde{A}$ satisfies that $\lim_{\epsilon_C \to \infty} f(\tilde{A}, \epsilon_C(k)) = \bar{f}$;

**4**    **while** $A_H - A_L \le \phi$ **do**
     // Bisection search for $A$

**5**      **if** $f(A + \bar{S}, \epsilon_C(k)) \le \bar{f}$ **then**

**6**        Set $A_L = A(k)$;

**7**      **else**

**8**        Set $A_H = A(k)$;

**9**      **end**

**10**    **end**

**11** **end**

**12** Select $k^* = \arg\min \rho^{A(k)} \cdot (\exp(\epsilon_C(k)) - 1)$;

**13** **return** the solution $\{\epsilon_C(k^*), A(k^*), \bar{S}(k^*)\}$.

---

**Algorithm 2** Heuristic Solution to (42)

---

**1** Initialize $\phi$, $K$, and $\bar{\epsilon}$;

**2** **for** $k \in \{1, 2, \ldots, K\}$ **do**
   // Exhaustive search for $\epsilon_C$

**3**    Set $\epsilon_C(k) = k\bar{\epsilon}/K$;

**4**    Set $\bar{S}_L = 0$ and $\bar{S}_H = \bar{B}$, where $\bar{B}$ satisfies that $\lim_{\epsilon_C \to \infty} f(\bar{B}, \epsilon_C) = \bar{f}$ ;

**5**    **while** $\bar{S}_H - \bar{S}_L \le \phi$ **do**
     // Bisection search for $\bar{S}$

**6**      **if** $f(\bar{S}, \epsilon_C) \le \bar{f}$ **then**

**7**        Set $\bar{S}_L = \bar{S}(k)$;

**8**      **else**

**9**        Set $\bar{S}_H = \bar{S}(k)$;

**10**      **end**

**11**    **end**

**12** **end**

**13** Select $k^* = \arg\min \frac{\exp(\epsilon_C(k)) - 1}{1 - c\rho^{\bar{S}(k)} \exp(\epsilon_C(k))}$;

**14** **return** the solution $\{\epsilon_C(k^*), A = 0, \bar{S}(k^*)\}$;

---

$$f(A + \bar{S}, \epsilon_C) \le \bar{f}, \tag{41c}$$

$$\text{var.} \quad \epsilon_C \ge 0, A \ge 0, \bar{S} \ge 0. \tag{41d}$$

Problem (41) is a convex problem whenever $f(t, \epsilon_C)$ is convex in $(t, \epsilon_C)$, in which case Problem (41) can be readily solved by standard solvers (e.g., CVX [70]). When $f(t, \epsilon_C)$ is non-convex, we present a heuristic algorithm to solve Problem (41) as shown in Algorithm 1. The computational complexity of Algorithm 1 is $\mathcal{O}(K \log(1/\phi))$, where $K$ comes from the exhaustive search for $\epsilon_C$ and $\log(1/\phi)$ comes from the bisection search for $A$.

*2) When the Constraint $A \ge 0$ Is Binding:* Substituting $A = 0$ into Problem (39), we have the following equivalent reformulated problem:

$$\min \quad \frac{\exp(\epsilon_C) - 1}{1 - c\rho^{\bar{S}} \exp(\epsilon_C)} \tag{42a}$$

$$\text{s.t.} \quad c\rho^{\bar{S}} \exp(\epsilon_C) < 1, \tag{42b}$$

$$f(\bar{S}, \epsilon_C) \le \bar{f}, \tag{42c}$$

$$\text{var.} \quad \epsilon_C \ge 0, \bar{S} \ge 0. \tag{42d}$$

We solve Problem (42) in Algorithm 2, in which we set $A = 0$ and search for the optimal $\epsilon_C$ exhaustively and use a bisection search for the optimal $\bar{S}$. Specifically, for each $\epsilon_C$, we use a bisection method to search for the corresponding $\bar{S}$ such that $f(\bar{S}, \epsilon_C) = \bar{f}$. We then select $\epsilon_C$ and the corresponding optimal $\bar{S}$ to attain the minimal value of the objective in (42a). Similarly, the complexity of Algorithm 2 is also $\mathcal{O}(K \log(1/\phi))$.

### D. Possibility of No Tradeoffs

Intuitively, increasing privacy should lead to a decrease in data utility (measured by an increase in the peak age penalty function $f$). However, unlike the single-query mechanism, this tradeoff between privacy and utility in the multi-query scenario may not exist, i.e., there may be cases where increasing the bound on the penalty ($\bar{f}$) does not lead to greater peak privacy risk. We provide a sufficient condition for the non-existence of a tradeoff in the following corollary:

*Corollary 3:* If $f\left(\frac{\ln(\exp(\epsilon_C) - 1) + \epsilon_C}{\ln(\rho^{-1})} + a, \epsilon_C\right)$ decreases in $\epsilon_C$ for any coefficient $a$, then an increase in $\bar{f}$ does not decrease the minimal objective value in Problem (39).

We present the proof of Corollary 3 in Appendix H. The main idea of the proof involves showing the possibility of constructing a new solution $f(A' + \bar{S}', \epsilon_C') < f(A + \bar{S}, \epsilon_C)$ while maintaining the peak privacy risk under the condition in Corollary 3. Corollary 3 implies that sacrificing one of the metrics (privacy or utility) does not necessarily improve the performance of another. In Section VI-A, we will present concrete examples that satisfy and violate the condition in Corollary 3.

## VI. NUMERICAL AND EXPERIMENTAL EXAMPLES

In this section, we consider two concrete examples of modeling the process $\{X_t\}$. We first consider a two-state Markov chain, to study the tradeoffs between age-dependent DP and accuracy loss achieved by our proposed scheme (by aging and noise injection) and the classical scheme. We further consider an experiment of our proposed scheme applied to electricity consumption data. We also consider an autoregressive model and show how our analysis can also be extended to continuous-valued processes in Appendix I.

### A. Two-State Markov Chain

In this subsection, we first consider the following example with two states [18]:

*Example 1:* Consider a system of $I$ agents and two locations. We use $x_{i,t} \in \{-1, 1\}$ to denote the location at time $t$ of agent $i$ and let $X_t = \{x_{i,t}\}_{i \in \mathcal{I}}$. Each agent $i$'s time-varying database is given by an identical discrete-time Markov chain,

with the following transition probability matrix:[6]

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}, \quad (43)$$

which can be verified to be reversible. Assuming $|1-p-q| < 1$,[7] it follows that the stationary distribution is $\pi(0) = \frac{q}{p+q}$ and $\pi(1) = \frac{p}{p+q}$, and the $t$-step transition probability matrix is:

$$P_t = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1-p-q)^t}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}. \quad (44)$$

It follows that

$$\|P_{t,-1} - \pi\| = \frac{p}{p+q}|1-p-q|^t, \quad (45)$$

$$\|P_{t,1} - \pi\| = \frac{q}{p+q}|1-p-q|^t, \quad (46)$$

where $P_{t,x}(y) = P_t(x,y)$ for all $x, y \in \{-1, 1\}$. Therefore, the maximal total variation distance satisfies

$$\Delta(t) = |1-p-q|^t, \forall t \in \mathbb{N}. \quad (47)$$

Such a two-state Markov chain model may capture several practical scenarios, e.g., i) the status of a home being empty or not, ii) a user being on a road at one of two locations. In the latter example, an analyst aims at designing a mechanism to estimate the current aggregate traffic congestion levels of two locations (e.g., to estimate the traffic conditions), whereas users may try to preserve their own location privacy.

We consider a common single-query aggregation mechanism $M$ that estimates the average value of $x_{i,t}$ across all users, e.g., to analyze the overall traffic conditions:

$$M(X_0) = \frac{1}{I}\sum_{i \in \mathcal{I}} x_{i,0} + n, \quad (48)$$

where $n$ is a Laplace$(\frac{1}{\epsilon_C I})$ random variable (which ensures $M$ to be $\epsilon_C$-DP by Proposition 1). Based on (48), we consider two utility metrics:[8]

- *Mean-Square Error*: We assume no prior information is available and simply use the given released data $M(X_0)$ as the estimate. The mean-square error (MSE) is given by [69]:

$$f_{\mathrm{MSE}}(t, \epsilon_C) = \mathbb{E}\left[\left(M(X_0) - \frac{1}{I}\sum_i x_{i,t}\right)^2\right]. \quad (49)$$

Such a mean-squared (estimation) accuracy loss is a commonly used age penalty function (e.g., [2], [35]).

- *Failure Rate*: Motivated by the age of incorrect information proposed in [58], we consider a binary estimator.

---

[6]We drop the index $i$ in the matrices for Example 1, as agents' time-varying databases are identical.

[7]If $p = q = 0$, the Markov chain is reducible. If $p = q = 1$, the Markov chain is periodic.

[8]It is worth noting that analytically quantifying the benefits of our proposed mechanism compared to the noise only benchmark in the general setting is highly challenging. This challenge arises due to the significant disparity between the metrics used to quantify "data utility" and "accuracy," resulting in diverse observations across different settings. Consequently, studying utility-privacy tradeoffs and their corresponding benefits usually requires an approach tailored to each specific case.

Specifically, based on the outcome of the mechanism $M(X_0)$, we seek to determine whether $\sum_i x_{i,t}/I \geq 0$ is true (e.g., whether a the traffic at a specific location is congested). Let $z_t$ be an indicator random variable for the event that $\sum_i x_{i,t}/I \geq 0$. The maximum likelihood estimate of $z_t$ is given by:

$$g_{\mathrm{MLE}}(M(X_0)) \triangleq \arg\max_{x \in \{0,1\}} \Pr[z_t = x | M(X_0)]. \quad (50)$$

We use the following definition of the failure rate (conditional on the initial state $X_0$) as our noise-aware age penalty:

$$f_{\mathrm{MLE}}(t, \epsilon_c) = \Pr[g_{\mathrm{MLE}}(M(X_0)) \neq z_t | X_0]. \quad (51)$$

*1) The Single-Query Case:* Combining (47) and Theorem 2, we can characterize the achievable privacy guarantee in the single-query case as follows:

*Corollary 4:* Any $\epsilon_C$-DP single-query $M$ under (48) is $(\epsilon(t), t)$-age-dependent DP, where $\epsilon(t)$ satisfies

$$\epsilon(t) = \ln\left(1 + |1-p-q|^t(\exp(\epsilon_C) - 1)\right), \forall t \in \mathbb{N}. \quad (52)$$

*2) The Multi-Query Case:* We further consider a multi-query mechanism with an equal-spacing policy described in (29). The privacy guarantee result follows from Proposition 5:

*Corollary 5:* Any $\epsilon_C$-DP multi-query $\mathcal{M}_t$ based on (48) is $(\epsilon(t), t, \mathcal{S}, \mathcal{A})$-age-dependent DP with a peak privacy risk given by

$$\epsilon^* = \ln\left(1 + \frac{|1-p-q|^A \cdot (\exp(\epsilon_C) - 1)}{1 - |1-p-q|^{\bar{S}}\exp(\epsilon_C)}\right). \quad (53)$$

*3) Numerical Results:* To understand the tradeoffs between privacy and utility for Example 1, we present numerical results for the mean-squared error and the failure rate scenarios in Figs. 4 and 5, respectively.

In the mean-square error scenario, we present the accuracy loss and privacy risk at different ages $t$ in Figs. 4(a) and 4(b), respectively, and the tradeoffs between privacy risk and accuracy loss in Figure 4(c). In Figure 4(a), we show that the accuracy loss increases in $t$, and for a given $\epsilon_C$, it converges to an upper bound as $t \to \infty$. This implies that aging does not necessarily incur significant accuracy loss in terms of the mean-square error. On the other hand, Figure 4(b) shows that privacy risks decrease in $t$ and converge to 0, which is mainly because $\Delta(t)$ converges to 0. In Figure 4(c), we compare our proposed mechanisms against a classical noise only benchmark that only injects Laplace noise, whose accuracy loss is given by $2/(\epsilon_C^2 I^2)$. We show that, to achieve an arbitrary small privacy risk, the accuracy loss incurred by the benchmark grows unbounded. The large loss is partially because of the lack of prior information in the considered mean-square error scenario. Our proposed scheme combines both aging and noise injection and is able to achieve a finite accuracy loss, as the privacy risk approaches 0. We note that this is partially because our age-dependent DP protects privacy specifically from adversaries seeking to infer $X_t$, whereas DP does not assume any type of data distributions or adversaries.

In the failure rate scenario, we present the accuracy loss and privacy risk for a single query mechanism at different
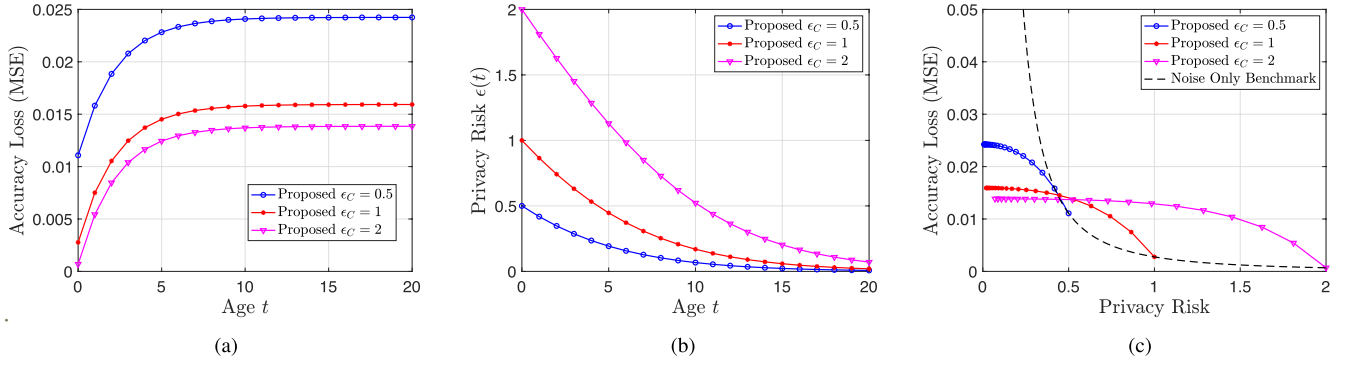
Fig. 4.   Numerical results for Example 1 with mean-square errors for single-query mechanisms. We set $p = q = 0.1$ and $I = 20$.
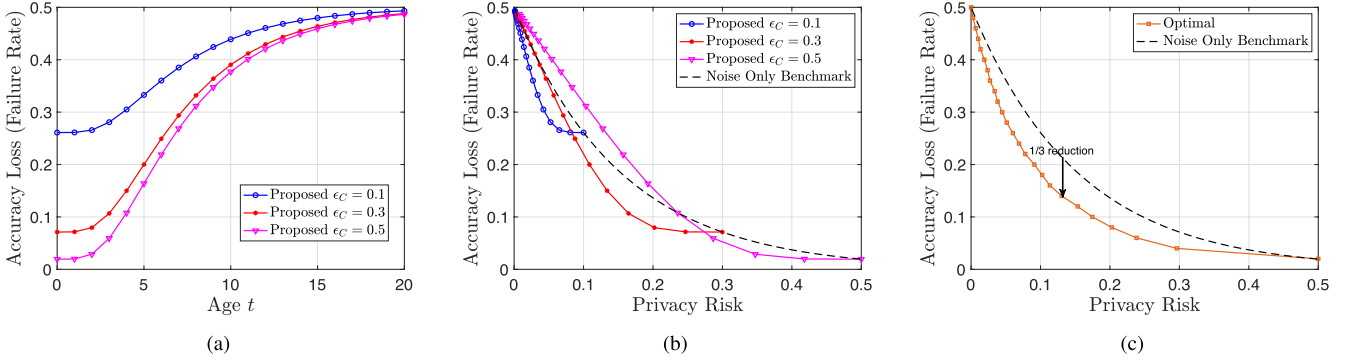


Fig. 5.   Numerical results for Example 1 with the failure rate for single-query mechanisms. Here $p = q = 0.1$ and $I = 20$.

ages $t$ in Figure 5(a), and the tradeoffs between privacy risk and accuracy loss in Figures 5(b)(c). Figure 5(a) shows that the accuracy loss increases in $t$ as it does in Figure 4(a), but at a different rate. The accuracy loss eventually converges to 0.5 as $t \to \infty$. We note that the privacy risks for a given $t$ and $\epsilon_C$ are the same as in the mean-squared error scenario (Figure 4(b)). In terms of the tradeoffs between privacy and accuracy loss, we show that our proposed scheme can achieve less accuracy loss compared to the noise only benchmark. In addition, Figure 5(b) also shows that different utility-privacy pairs require different $t$ and $\epsilon_C$. In Figure 5(c), we present the optimal tradeoffs between privacy and accuracy loss. Compared to the noise only benchmark, this shows that our proposed scheme can achieve a reduction up to $1/3$ in accuracy loss given the same privacy risk.

Finally, Figure 6 presents the tradeoffs between privacy and utility in the multi-query scenario. Figure 6(a) show that there exists a pair that minimizes both the peak accuracy loss and the peak privacy risk. This is mainly because the condition in Corollary 3 holds in this case and hence the tradeoff between privacy and utility does not exist. Figure 6(b) demonstrates the existence of the optimal tradeoffs between privacy and utility for multi-query mechanisms in the failure rate scenario. That is, a smaller achievable peak privacy risk leads to a larger achievable peak accuracy loss.

### B. Electricity Consumption Forecast

In this experiment, we use electricity consumption readings of households in London provided in [75] to evaluate our proposed scheme. UK Power Networks recorded Power consumption (in kWh) every 30 minutes between November 2011 and February 2014. We selected 40 households, with $28,000$ readings per household on average. We quantized the power values into 12 intervals, resulting in a Markov chain with 12 states for each household. Related studies also considered such a Markov chain formulation of electricity consumption readings [18]. Our goal is to publish a privacy-preserving approximation to forecast households' average electricity consumption. In particular, to evaluate the privacy risks, we use the following result to estimate an upper bound for the total variation distance between the transition probability and the stationary probability:

*Proposition 5 (Bounds on Variation Distance [66]):* Let $\{X_t\}_{t \in \mathbb{N}}$ be an irreducible, aperiodic, and reversible Markov chain on a finite set $\mathcal{X}$. Then for all $X \in \mathcal{X}$, $t \in \mathbb{N}$,

$$4 \cdot \delta(P_t(X, \cdot), \pi)^2 \leq \frac{1 - \pi(X)}{\pi(X)} \lambda_*^{2t}. \tag{54}$$

Based on Proposition 5, we can bound the maximal total variation distance $\Delta(t)$, i.e.,

$$\Delta(t) \leq \min \left\{ 1, \max_{i \in \mathcal{I}} \max_{x_i \in \mathcal{X}_i} \sqrt{\frac{1 - \pi_i(x_i)}{\pi_i(x_i)}} \lambda_{i,*}^t \right\}, \tag{55}$$

where $1 - \lambda_{i,*}$ represents the spectral gap of household $i$'s time-varying database $\{x_{i,t}\}_{t \in \mathbb{N}}$.

In Figure 7, we study the accuracy loss (measured by the MSE between the released data mean usage and the true mean usage) and the privacy risks of our single-query mechanism applied to the electricity power dataset for different $\epsilon_C$ values.
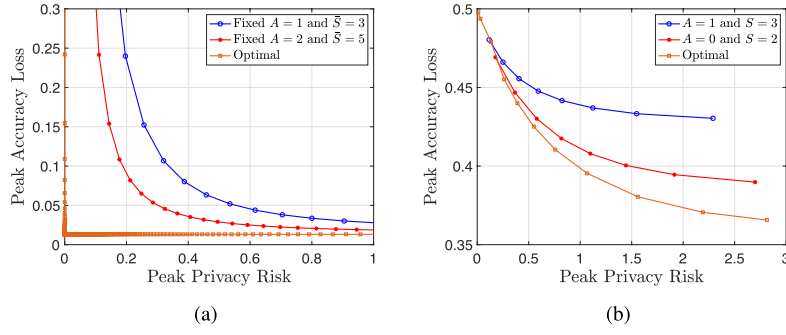
Fig. 6.  Numerical results for Example 1 with for multi-query mechanisms with (a) mean-square errors and (b) the failure rate. We set $p = q = 0.1$ and $I = 20$.
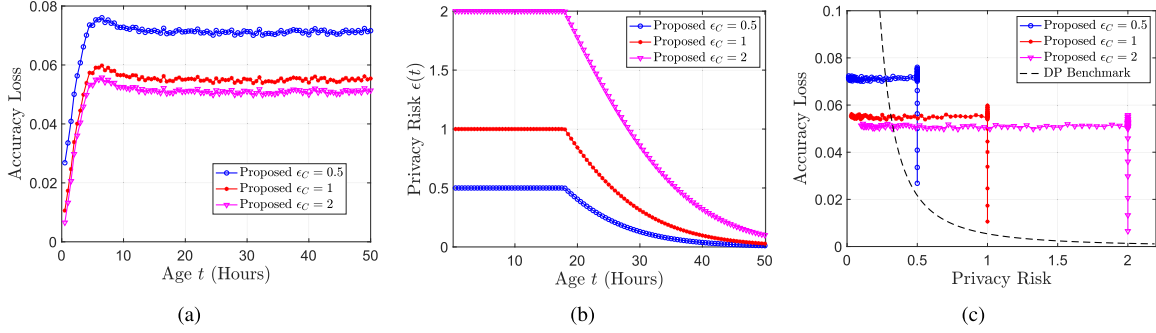


Fig. 7.  Experimental results for electricity consumption data of London households.

As shown in Figure 7(a), the accuracy loss first increases rapidly in $t$ and then remains a relatively constant level, for all choices of $\epsilon_C$ values. In Figure 7(b), we observe that the upper bound on the privacy risks in (55) are first the same as their corresponding classical DP risks and then decrease when the age is greater than 18 hours. This mainly results from (55), in which the second term $\max_{i \in \mathcal{I}} \max_{x_i \in \mathcal{X}_i} \sqrt{\frac{1 - \pi_i(x_i)}{\pi_i(x_i)}}$ only becomes less than 1 when $t \geq 18$ hours.[9] Finally, Figure 7(c) depicts the tradeoffs between privacy and accuracy loss. Similarly, the accuracy loss incurred by the DP benchmark grows unbounded as the privacy risk approaches zero, while a finite accuracy loss is always achievable under our proposed scheme that combines both noise injection and aging.

## VII. CONCLUSION

In this paper, we proposed an age-dependent generalization of differential privacy. We characterized the impact of data staleness on data privacy guarantees and showed that aging provides a new direction to protect data privacy (in addition to noise injection). Motivated by frequent data updates in real-time applications, we further characterized how age-dependent privacy risks compose, given any publishing and aging policies. Finally, our case studies showed that mechanisms combining aging and noise injection may significantly outperform the classical mechanisms (that inject noise only), when making tradeoffs between privacy and utility.

As the first study on understanding the impact of data freshness on privacy, there are many future research directions.

First, our work assumes that the adversary is only interested in keeping track on the most current state of some user. An important potential direction is to extend our results to different adversary models, e.g., adversaries that aim to infer a subset of the entire history of users' databases. Second, important future research directions include the extensions of our results to the $(\epsilon, \delta)$-DP setting[10] and the non-stationary setting. Third, it is interesting to study adaptive composition to enhance the privacy guarantees for multi-query mechanisms.

## APPENDIX

### A. Preliminaries: Mixing, Ergodicity, and Markov Chains

In order to characterize how privacy guarantees change over time under the new framework of age-dependent DP, we introduce several related concepts and preliminary results from the literature in this subsection.

*1) Geometric Ergodicity:* To understand under what conditions a Markov chain satisfies the geometric ergodicity, we first introduce the following important result in analyzing the transition probability matrix $P(\cdot, \cdot)$:

*Lemma 3 (Bounds on Eigenvalues [66]):* Let $P(\cdot, \cdot)$ be the transition matrix of an irreducible Markov chain $\{X_t\}_{t \in \mathbb{N}}$.

---

[9]It is possible to derive a tighter bound than (55), which is, however, out of the scope of this paper.

[10]We would like to emphasize that extending our analysis to the new setting is not straightforward. One of the main challenges is that the proof of Theorem 1 explicitly requires $f(\underline{z}_i)$ to be non-zero. While this condition can be automatically satisfied under $\epsilon$-differential privacy, it may not hold under $(\epsilon, \delta)$-differential privacy.

Nevertheless, a potential approach for analyzing the two-parameter version of age-dependent differential privacy is through a closely related privacy notion called zero concentrated differential privacy (zCDP) [76]. By establishing an age-dependent version of zCDP, it may be possible to conduct the analysis of the two-parameter version of age-dependent differential privacy.

Then the eigenvalues of $P$, $\{\lambda_k\}$, satisfy:

$$1 = \lambda_0 > \lambda_1 \geq \ldots \geq \lambda_{m-1} \geq -1, \qquad (56)$$

where $m = |\mathcal{X}|$ denotes the total number of states.

The largest eigenvalue $\lambda_0$ being 1 results from irreducibility. It is known that the convergence rate largely depends on the value $\lambda_* = \max(\lambda_1, |\lambda_{m-1}|)$.

There are a few studies in the literature of applied probability on analysis of the value of $\lambda_*$ (e.g., [66], [67]). As an example, Landau and Odlyzko in [67] provided a bound for a random walk on a connected graph and showed the following result:

*Proposition 6   [67]:* Let $\{X_t\}_{t \in \mathbb{N}}$ be a random walk on a connected graph, it follows that

$$\lambda_* \leq 1 - \frac{1}{|\mathcal{X}| d_* (1 + \gamma_*)} < 1 - \frac{1}{|\mathcal{X}|^3}, \qquad (57)$$

where $|\mathcal{X}|$ is the number of vertices in the graph, $d_*$ is the maximum degree of the graph, and $\gamma_*$ is the diameter of the graph.

## B. Basic Properties

*Proposition 7 (Post-Processing):* Let $M : \mathcal{X} \to \mathcal{Y}$ be a single-query $(\epsilon, t)$-age-dependent differentially private mechanism. Then, for any mapping $f : \mathcal{Y} \to \mathcal{Y}'$, $f(M(\mathcal{X}_0))$ is $(\epsilon, t)$-age-dependent differentially private.

*Proof:* By the definition of age-dependent DP in Definition 4, we have:

$$\Pr[M(X_0) \in \mathcal{W}|X_t = \boldsymbol{x}] \leq \exp(\epsilon)\Pr[M(X_0) \in \mathcal{W}|X_t = \boldsymbol{x}'], \qquad (58)$$

for any set of outcomes $\mathcal{W} \subset \mathcal{Y}$.

For a mapping $f(\cdot)$, it follows that

$$\begin{aligned}
&\Pr[f(M(X_0)) \in \mathcal{T}|X_t = \boldsymbol{x}] \\
&= \Pr[M(X_0) \in \mathcal{W}|X_t = \boldsymbol{x}] \\
&\leq \exp(\epsilon)\Pr[M(X_0) \in \mathcal{W}|X_t = \boldsymbol{x}'] \\
&= \exp(\epsilon)\Pr[f(M(X_0)) \in \mathcal{T}|X_t = \boldsymbol{x}'] \qquad (59)
\end{aligned}$$

where $\mathcal{W} = f^{-1}[\mathcal{T}]$, i.e. $\mathcal{S}$ is the preimage of $\mathcal{T}$ under mapping $f$. $\blacksquare$

## C. Proof of Lemma 1

It follows from the definition of $\epsilon_C$-DP that

$$\hat{P}_{-i,t}(\boldsymbol{x}_{-i}, \boldsymbol{z}_{-i})\Pr(\mathcal{M}(z_i, \boldsymbol{z}_{-i}) \in \mathcal{W})$$
$$\leq \exp(\epsilon_C)\hat{P}_{-i,t}(\boldsymbol{x}_{-i}, \boldsymbol{z}_{-i})\Pr(\mathcal{M}(z_i', \boldsymbol{z}_{-i}) \in \mathcal{W}) \qquad (60)$$

for all $\boldsymbol{x}_{-i}$, $\boldsymbol{z}_{-i} \in \mathcal{X}_{-i} \triangleq \prod_{j \neq i} \mathcal{X}_j$, and any pair of $z_i$ and $z_i' \in \mathcal{X}_i$. Summing (60) over all $\boldsymbol{z}_{-i} \in \mathcal{X}_{-i}$ yields Lemma 1.

## D. Proof of Theorem 2

In Lemma 2, we bound $\sum_{i \in \mathcal{X}_i} (p(x_i) + \delta(x_i)) x_i$ based on both the constraint on $\delta(x_i)$ and the bound of $x_i$. The key difference in this proof from that of Theorem 1 is that we consider a lemma which does not require the bound of $x_i$, based on which we can prove Theorem 2. We start with such a lemma:

*Lemma 4:* For any part of vectors with lengths $K$ denoted by $\{a_1, a_2, \ldots, a_k\}$ and $\{b_1, b_2, \ldots, b_k\}$, if $g \leq \frac{a_k}{b_k} \leq G$, then $g \leq \frac{\sum_{k=1}^{K} a_k}{\sum_{k=1}^{K} b_k} \leq G$.

*Proof:* We can prove Lemma 4 by summing $b_k g \leq a_k \leq b_k G$ all over $1 \leq k \leq K$, and the dividing the all sides of the resultant inequality by $\sum_{k=1}^{K} b_k$. $\blacksquare$

Without loss of generality, we assume that $X$ and $X'$ only differs in the $i$-th user's data, i.e., we have $X = \{x_i\}_{i \in \mathcal{I}}$ and $X' = \{x_1, \ldots, x_{i-1}, x_i', x_i, \ldots, x_I\}$. It follows that

$$\begin{aligned}
&\ln\left(\frac{\Pr[M(X_0) \in \mathcal{W}|X_t = X]}{\Pr[M(X_0) \in \mathcal{W}|X_t = X']}\right) \\
&= \ln\left(\frac{\sum_{\boldsymbol{z} \in \mathcal{X}} \Pr[M(\boldsymbol{z}) \in \mathcal{W}]\Pr[X_0 = \boldsymbol{z}|X_t = X]}{\sum_{\boldsymbol{z} \in \mathcal{X}} \Pr[M(\boldsymbol{z}) \in \mathcal{W}]\Pr[X_0 = \boldsymbol{z}|X_t = X']}\right) \\
&= \ln\left(\frac{\mathbb{E}_{z_i \sim \hat{P}_{i,t}(x_i, \cdot)}\{\mathbb{E}_{\boldsymbol{z}_{-i} \sim \hat{P}_{i,t}(\boldsymbol{x}_{-i}, \cdot)}[\Pr(M(\boldsymbol{z}) \in \mathcal{W})]\}}{\mathbb{E}_{z_i \sim \hat{P}_{i,t}(x_i', \cdot)}\{\mathbb{E}_{\boldsymbol{z}_{-i} \sim \hat{P}_{i,t}(\boldsymbol{x}_{-i}, \cdot)}[\Pr(M(\boldsymbol{z}) \in \mathcal{W})]\}}\right) \\
&\overset{(a)}{\leq} \ln\left(\max_{x_i, x_i', z_i} \frac{\hat{P}_{i,t}(x_i, z_i)}{\hat{P}_{i,t}(x_i', z_i)}\right) \\
&\overset{(b)}{\leq} \max_{x_i, x_i', z_i} \frac{P_{i,t}(z_i, x_i) - P_{i,t}(z_i, x_i')}{P_{i,t}(z_i, x_i')} \\
&\overset{(c)}{\leq} \max_{z_i} \frac{\sqrt{\frac{1-\pi(z_i)}{\pi(z_i)}}\lambda_*^t}{\pi(z_i) - \sqrt{\frac{1-\pi(z_i)}{4\pi(z_i)}}\lambda_*^t} = \mathcal{O}(\lambda_*^t), \qquad (61)
\end{aligned}$$

where (a) is due to Lemma 4 and the fact that $X$ and $X'$ only differs in the $i$-th user's data; (b) is because $\ln(1 + x) \leq x$; (c) is from Proposition 1.

This completes the proof of Theorem 2.

## E. Proof of Theorem 3

We first rewrite $\mathcal{M}_t(\mathcal{H}_t(\mathcal{S}, \mathcal{A}))$ into $[\mathcal{M}_{S_{n-1}}(\mathcal{H}_{S_{n-1}}(\mathcal{S}, \mathcal{A}))\ M_n(X_{S_n - A_n})]$ for $t \in (S_{n-1}, S_n]$. For notational simplicity, in this proof, we denote $\mathcal{M}_{S_{n-1}}$ by $\mathcal{M}'$ and $M_n$ by $M$, respectively. We further denote $\mathcal{H}_t(\mathcal{S}, \mathcal{A})$ and $X_{S_n - A_n}$ by $\boldsymbol{z}^{(1)}$ and $\boldsymbol{z}^{(2)}$, respectively.

We present the analysis of the privacy risk relation in (62) shown at the bottom of the next page, where $t_2 = t - S_n + A_n$, $h_1(X) = \Pr(M(X) \in \mathcal{W}_2)$ and $h_2(X) = \Pr[\mathcal{M}'(X_{S_n - A_n}) \in \mathcal{W}_1|X_{S_n - A_n} = X]$. Note that

$$h_1(X) \leq \exp(\epsilon_{C,n})h_1(X'), \qquad (63a)$$
$$h_2(X) \leq \exp(\epsilon(S_n - A_n))h_2(X') \qquad (63b)$$

for all $X, X' \in \mathcal{X}$ that differ only in one entry. Substituting (63) into (19) yields the last inequality of (62). This completes the proof.

### F. Proof of Proposition 4

Since $\ln(1+x)$ is monotonically increasing in $x$, we can rewrite (39) as

$$\min \quad \frac{c\rho^A \cdot (\exp(\epsilon_C) - 1)}{1 - c\rho^{\bar{S}}\exp(\epsilon_C)} \tag{64a}$$

$$\text{s.t.} \quad c\rho^{\bar{S}}\exp(\epsilon_C) < 1, \tag{64b}$$

$$f(A + \bar{S}, \epsilon_C) \leq \bar{f}, \tag{64c}$$

$$\text{var.} \quad \epsilon_C \geq 0, A \geq 0, \bar{S} \geq 0, \tag{64d}$$

We first drop the constraint in (64b) and formulate the corresponding Lagrangian:

$$\mathcal{L}(A, \bar{S}, \epsilon_C) = \frac{c\rho^A \cdot (\exp(\epsilon_C) - 1)}{1 - c\rho^{\bar{S}}\exp(\epsilon_C)} - \lambda(f(A + \bar{S}, \epsilon_C) - \bar{f}), \tag{65}$$

where $\lambda$ is the dual variable corresponding to the constraint in (64c). The Karush–Kuhn–Tucker conditions (necessary for all local optima) are [71]

$$\frac{-c\rho^A \ln(\rho)(\exp(\epsilon_C)-1)}{c\rho^{\bar{S}}\exp(\epsilon_C) - 1} - \lambda \frac{\partial f(A + \bar{S}, \epsilon_C)}{\partial A} = 0, \tag{66a}$$

$$\frac{c^2 \rho^{A+\bar{S}}\ln(\rho)\exp(\epsilon_C)(\exp(\epsilon_C)-1)}{(c\rho^{\bar{S}}\exp(\epsilon_C) - 1)^2} - \lambda \frac{\partial f(A + \bar{S}, \epsilon_C)}{\partial \bar{S}} = 0. \tag{66b}$$

Combining (66) and the fact that $\frac{\partial f(A+\bar{S},\epsilon_C)}{\partial \bar{S}} = \frac{\partial f(A+\bar{S},\epsilon_C)}{\partial A}$ leads to (40). We note that i) (40) is necessary even if the problem in (64) is not convex; ii) (40) automatically satisfies (64b), which implies that the relaxation of (64b) does not lose any feasibility.

### G. Proof of Theorem 4

To prove Theorem 4, we will first prove that the monotonicity of $\epsilon(S_n)$ in $n$ (Lemma 5). We will then present the intermediate result of the existence of an optimal solution that have identical $(A_n^*, \bar{S}_n^*, \epsilon_{C,n}^*)$ after $n \geq m$ in Lemma 6, based on which we complete the whole proof.

Define

$$F(A_n, \bar{S}_n, \epsilon_{C,n}, \epsilon)$$
$$\triangleq \ln\left(1 + \Delta\left(\bar{S}_n + A_n\right)\left(\exp(\epsilon_{C,n})[\Delta(A) \cdot (\epsilon - 1) + 1] - 1\right)\right), \tag{67}$$

for all $t \in [S_n, S_{n+1}), n \in \mathbb{N}$. We have $\epsilon_{n+1} = F(A_n, \bar{S}_n, \epsilon_{C,n}, \epsilon_n)$, which is strictly increasing in $\epsilon$ and $\epsilon_{C,n}$. For any $m \in \mathbb{N}$, if we fix $A_n, \bar{S}_n, \epsilon_{C,n}$ for all $n \geq m$, it follows that $\epsilon_m + 1$ strictly increases in $\epsilon_m$.

Let the maximal objective value of (28) be $\epsilon^*$. Let $m \in \mathbb{N} \cup \{+\infty\}$ be the first epoch that the peak age-dependent privacy risk such that $\epsilon(S_m) = \epsilon^*$.

We first introduce the following lemma:

*Lemma 5:* For any optimal solution $\mathcal{Z}^* = (\{A_n^*\}, \{\bar{S}_n^*\}, \{\epsilon_{C,n}^*\})$, the sequence $\{\epsilon_n\}_n$ is non-decreasing in $n$.

To prove Lemma 5, suppose that there exists an optimal solution $\mathcal{Z}^* = (\{A_n^*\}, \{\bar{S}_n^*\}, \{\epsilon_{C,n}^*\})$ such that $\epsilon_{k+1} < \epsilon_k$ for some $k$. In this case, we replace $(A_n^*, \bar{S}_n^*, \epsilon_{C,n}^*)$ for all $n \geq k$ by $(A_k^*, \bar{S}_k^*, \epsilon_{C,k}^*)$, i.e., we construct the following new solution $\mathcal{Z}' = (\{A_n'\}, \{\bar{S}_n'\}, \{\epsilon_{C,n}'\})$ satisfying:

$$(A_n', \bar{S}_n', \epsilon_{C,n}') = \begin{cases} (A_k^*, \bar{S}_k^*, \epsilon_{C,k}^*), & \text{if } n \geq k, \\ (A_n^*, \bar{S}_n^*, \epsilon_{C,n}^*), & \text{otherwise.} \end{cases} \tag{68}$$

For the new sequence $\{\epsilon_n'\}$ generated by $\mathcal{Z}'$, it follows that

$$F(A_k^*, \bar{S}_k^*, \epsilon_{C,k}^*, \epsilon_n') < F(A_k^*, \bar{S}_k^*, \epsilon_{C,k}^*, \epsilon_n), \tag{69}$$

---

$$\ln\left(\frac{\Pr[\mathcal{M}' \in \mathcal{W}_t, M(X_{S_n - A_n}) \in \mathcal{W}_{S_n - A_n} | X_t = X]}{\Pr[\mathcal{M}' \in \mathcal{W}_t, M(X_{S_n - A_n}) \in \mathcal{W}_{S_n - A_n} | X_t = X']}\right)$$

$$= \ln\left(\frac{\sum_{z^{(1)}}\sum_{z^{(2)}} \Pr[\mathcal{M}'(z^{(1)}) \in \mathcal{W}_1]\Pr[M(z^{(2)}) \in \mathcal{W}_2]\Pr[X_{S_{n-1}} = z^{(1)}, X_{S_n - A_n} = z^{(2)}|X_t = X]}{\sum_{z^{(1)}}\sum_{z^{(2)}} \Pr[\mathcal{M}'(z^{(1)}) \in \mathcal{W}_1]\Pr[M(z^{(2)}) \in \mathcal{W}_2]\Pr[X_{S_{n-1}} = z^{(1)}, X_{S_n - A_n} = z^{(2)}|X_t = X']}\right)$$

$$= \ln\left(\frac{\sum_{z^{(2)}} \Pr[X_{S_n - A_n} = z^{(2)}|X_t = x]\Pr[M(z^{(2)}) \in \mathcal{W}_2]\Pr[\mathcal{M}'(X_{S_n - A_n}) \in \mathcal{W}_1|X_{S_n - A_n} = z^{(2)}]}{\sum_{z^{(2)}} \Pr[X_{S_n - A_n} = z^{(2)}|X_t = x']\Pr[M(z^{(2)}) \in \mathcal{W}_2]\Pr[\mathcal{M}'(X_{S_n - A_n}) \in \mathcal{W}_1|X_{S_n - A_n} = z^{(2)}]}\right)$$

$$= \ln\left(\frac{\mathbb{E}_{z^{(2)} \sim \hat{P}_{t_2}(x, \cdot)}\left[\Pr(M(z^{(2)}) \in \mathcal{W}_2) \cdot \Pr[\mathcal{M}'(X_{S_n - A_n}) \in \mathcal{W}_1|X_{S_n - A_n} = z^{(2)}]\right]}{\mathbb{E}_{z^{(2)} \sim \hat{P}_{t_2}(x', \cdot)}\left[\Pr(M(z^{(2)}) \in \mathcal{W}_2) \cdot \Pr[\mathcal{M}'(X_{S_n - A_n}) \in \mathcal{W}_1|X_{S_n - A_n} = z^{(2)}]\right]}\right)$$

$$= \ln\left(\frac{\mathbb{E}_{z_i^{(2)} \sim \hat{P}_{i,t_2}(x_i, \cdot)}\left\{\mathbb{E}_{z_{-i}^{(2)} \sim \hat{P}_{-i,t_2}(x_{-i}, \cdot)}\left[\Pr(M(z^{(2)}) \in \mathcal{W}_2) \cdot \Pr[\mathcal{M}'(X_{S_n - A_n}) \in \mathcal{W}_1|X_{S_n - A_n} = z^{(2)}]\right]\right\}}{\mathbb{E}_{z_i^{(2)} \sim \hat{P}_{i,t_2}(x_i', \cdot)}\left\{\mathbb{E}_{z_{-i}^{(2)} \sim \hat{P}_{-i,t_2}(x_{-i}, \cdot)}\left[\Pr(M(z^{(2)}) \in \mathcal{W}_2) \cdot \Pr[\mathcal{M}'(X_{S_n - A_n}) \in \mathcal{W}_1|X_{S_n - A_n} = z^{(2)}]\right]\right\}}\right)$$

$$= \ln\left(\frac{\mathbb{E}_{z_i^{(2)} \sim \hat{P}_{i,t_2}(x_i, \cdot)}\left\{\mathbb{E}_{z_{-i}^{(2)} \sim \hat{P}_{-i,t_2}(x_{-i}, \cdot)}\left[h_1(z^{(2)}) \cdot h_2(z^{(2)})\right]\right\}}{\mathbb{E}_{z_i^{(2)} \sim \hat{P}_{i,t_2}(x_i', \cdot)}\left\{\mathbb{E}_{z_{-i}^{(2)} \sim \hat{P}_{-i,t_2}(x_{-i}, \cdot)}\left[h_1(z^{(2)}) \cdot h_2(z^{(2)})\right]\right\}}\right)$$

$$\leq \ln\left(1 + \Delta\left(t - S_n + A_n\right) \cdot (\exp(\epsilon_{C,n} + \epsilon(S_n - A_n)) - 1)\right), \forall t \in [S_n, S_{n+1}), n \in \mathbb{N}, \tag{62}$$

i.e., $\epsilon'_{n+1} < \epsilon_{n+1}$. Similarly, we have

$$\epsilon'_{n+1} < \epsilon'_n, \forall n \geq k. \tag{70}$$

Therefore, we have $\epsilon^* > \epsilon_k > \epsilon'_k > \epsilon_n$ for all $n \in \mathbb{N}$, which is a contradiction to the fact that $\mathcal{Z}^*$ is optimal.

We have the new solution $\mathcal{Z}'$ is also feasible (as $f(A_m + \bar{S}_m, \epsilon_{C,m}) \leq \bar{f}$).

$$F(A_k^*, \bar{S}_k^*, \epsilon_{C,n}^*) = \ln\left(1 + \Delta\left(\bar{S}_k^* + A_k^*\right)\right.$$
$$\left. \cdot \left(\exp(\epsilon_{C,n}^*)[\Delta(A_k^*) \cdot (\epsilon_n - 1) + 1] - 1\right)\right), n \in \mathbb{N}. \tag{71}$$

In addition, we must have

*Lemma 6:* For any optimal solution $\mathcal{Z}^* = (\{A_n^*\}, \{\bar{S}_n^*\}, \{\epsilon_{C,n}^*\})$, the sequence $\{\epsilon_n\}_n$, let $\epsilon_m = \epsilon_n^*$ for $m$, then there exists another new optimal solution such that $A_n^* = A^*$, $S_n^* = S^*$, and $\epsilon_{C,n}^* = \epsilon_C^*$ for all $n \geq m$.

For any optimal solution $\mathcal{Z}^* = (\{A_n^*\}, \{\bar{S}_n^*\}, \{\epsilon_{C,n}^*\})$, we construct a new solution such that

$$(A'_n, \bar{S}'_n, \epsilon'_{C,n}) = \begin{cases} (A_m^*, \bar{S}_m^*, \epsilon_{C,m}^*), & \text{if } n \geq m, \\ (A_n^*, \bar{S}_n^*, \epsilon_{C,n}^*), & \text{otherwise.} \end{cases} \tag{72}$$

It follows that

$$F(A_m^*, \bar{S}_m^*, \epsilon_{C,m}^*, \epsilon_m) = \epsilon_{m+1} = \epsilon^*. \tag{73}$$

Now, we focus on the optimal solution $\mathbf{Z}^*$ such that $A_n^* = A_{n+1}^* = A^*$, and $\bar{S}_n^* = \bar{S}_{n+1}^* \bar{S}^*$, $\epsilon_{C,n}^* = \epsilon_{C,n+1}^* = \epsilon_C^*$ for all $n \geq m$. Now, we consider a new solution such that $\mathcal{Z}' = (\{A'_n\}, \{\bar{S}'_n\}, \{\epsilon'_{C,n}\})$ such that $A'_n = A^*$, $\bar{S}'_n = \bar{S}^*$, and $\epsilon'_{C,n} = \epsilon_C^*$ for all all $n \in \mathbb{N}$.

Note that, since

$$F(A^*, \bar{S}^*, \epsilon_C^*, \epsilon^*) = \epsilon^*, \tag{74}$$

and $F(A^*, \bar{S}^*, \epsilon_C^*, \epsilon)$ in strictly increasing in $\epsilon$. We have that, for the sequence $\{\epsilon_n\}$ generated according to $\epsilon_{n+1} = F(A^*, \bar{S}^*, \epsilon_C^*, \epsilon_n)$ for all $n \in \mathbb{N}$ with $\epsilon_0 = 0$, $\epsilon_n \leq \epsilon^*$ for all $n \in \mathbb{N}$. Therefore, the new solution $\mathcal{Z}'$ must be optimal as well.

### H. Proof of Corollary 3

Let $(A, \bar{S}, \epsilon_C)$ be any feasible solution, we construct another solution $(A', \bar{S}', \epsilon'_C)$ such that

$$A' > A, \quad \bar{S}' > S, \epsilon'_C > \epsilon_C, \tag{75}$$

$$\rho^A(\exp(\epsilon_C) - 1) = \rho^{A'}(\exp(\epsilon'_C) - 1), \tag{76}$$

$$\rho^{\bar{S}}\exp(\epsilon_C) = \rho^{\bar{S}'}\exp(\epsilon'_C). \tag{77}$$

From (39a), we can show that $(A', \bar{S}', \epsilon'_C)$ leads to the same objective value as $(A, \bar{S}, \epsilon_C)$ does.

On the other hand, we have that

$$A' = A + \frac{\ln(\exp(\epsilon'_C) - 1)}{\ln(\rho^{-1})} - \frac{\ln(\exp(\epsilon_C) - 1)}{\ln(\rho^{-1})}, \tag{78}$$

$$\bar{S}' = \bar{S} + \frac{\epsilon'_C}{\ln(\rho^{-1})} - \frac{\epsilon_C}{\ln(\rho^{-1})}. \tag{79}$$

Therefore, the new value of the peak age penalty is given by

$$f\left(A' + \bar{S}', \epsilon'_C\right) = f\left(\frac{\ln(\exp(\epsilon'_C) - 1)}{\ln(\rho^{-1})} + \frac{\epsilon'_C}{\ln(\rho^{-1})} + a, \epsilon'_C\right), \tag{80}$$

where $a = A - \frac{\ln(\exp(\epsilon_C) - 1)}{\ln(\rho^{-1})} + \bar{S} - \frac{\epsilon_C}{\ln(\rho^{-1})}$.

The assumption that (80) is decreasing in $\epsilon'_C$ means that the new solution will satisfy $f(A' + \bar{S}', \epsilon'_C) < f(A + \bar{S}, \epsilon_C)$ while achieving the same peak privacy guarantees.

### I. Autoregressive Model

We further consider a continuous-state autoregressive (AR) models, expressed as

$$x_{i,t+1} = \rho_1 x_{i,t} + \rho_2 x_{i,t-1} + \rho_3 x_{i,t-2} \ldots + \epsilon_{i,t}, \tag{81}$$

where $\epsilon_{i,t}$ follows an i.i.d. Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

We start with the following time reversibility result:

*Lemma 7 [68]:* When $\epsilon_{i,t}$ follows i.i.d. Gaussian distributions, the process in (81) is time reversible.

In the following, we focus on an AR(1) model for trackability, i.e., among all $\{\rho_t\}$ only the first coefficient $\rho_1$ takes a non-zero value. Therefore, we drop the index 1 in $\rho_1$ in the following.

When $\epsilon_{i,t}$ has mean zero and variance $\sigma^2$, it follows that

$$\mathbb{E}[x_{i,t+n}|x_{i,n}] = \rho^t x_{i,n}, \forall t, n \in \mathbb{N}_+, \tag{82}$$

and

$$\text{Var}[x_{i,t}|x_{i,0}] = \frac{(1 - \rho^{2t})\sigma^2}{1 - \rho^2}, \forall i \in \mathcal{I}, \forall n \in \mathbb{N}_+. \tag{83}$$

*Lemma 8:* When $\epsilon$ follows $\mathcal{N}(0, \sigma^2)$ for all $i \in \mathcal{I}$ and $t \in \mathbb{N}$, the probability distribution of $X_{i,t}$ conditional on $X_{i,0}$ is expressed as

$$x_{i,t}|x_{i,0} \sim \mathcal{N}\left(\rho^t x_{i,0}, \frac{(1 - \rho^{2t})\sigma^2}{1 - \rho^2}\right), \quad \forall i \in \mathcal{I}, t \in \mathbb{N}_+. \tag{84}$$

In addition, as in (15) the total variation distance is closely related to the KL *divergence*:

$$D_{\text{KL}}(\mu \| \pi) \triangleq \sum_{x \in \mathcal{X}} \pi(x) \log\left(\frac{\mu(x)}{\pi(x)}\right), \tag{85}$$

and its relation to the total variation distance can be expressed in the following:

*Lemma 9:* The total variation distance is related to the Kullback–Leibler divergence by the Pinsker's inequality:

$$\delta(\pi, \mu) \leq \sqrt{\frac{1}{2}D_{\text{KL}}(\pi \| \mu)}, \forall \pi, \mu. \tag{86}$$

It follows that the KL divergence is

$$D_{\text{KL}}(P_t(x_{i,0}, \cdot) \| P_t(x'_{i,0}, \cdot))$$
$$= \frac{1}{2}\left\{\frac{(\mu_1 - \mu_0)^2}{\sigma^2}\right\} = \frac{(1 - \rho^2)\rho^{2t}(x_{i,t} - x'_{i,t})^2}{2\sigma^2(1 - \rho^{2n})}$$
$$\leq \frac{\rho^{2t}(x_{i,0} - x'_{i,0})^2}{2\sigma^2} = \mathcal{O}(\rho^{2t}). \tag{87}$$
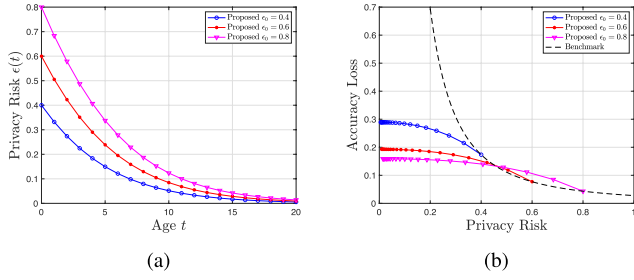
Fig. 8. Numerical results for the autoregressive model.

Based on the relation between the KL divergence and the total variation distance in Lemma 9, it follows that

$$
\begin{aligned}
\Delta(t) &= \delta(P_t(x_{i,0}, \cdot), P_t(x'_{i,0}, \cdot)) \\
&\le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P_t(x_{i,0}, \cdot) \parallel P_t(x'_{i,0}, \cdot))} \le \frac{\rho^t |x_{i,0} - x'_{i,0}|}{2\sigma}.
\end{aligned}
\tag{88}
$$

The convergence rate of $\Delta(t)$ again is $\mathcal{O}(\rho^t)$.

From Theorem 2, we conclude that any $\epsilon_C$-DP mechanism $M$ is also $(\epsilon(t), t)$-age-dependent DP, satisfying

$$
\epsilon(t) = \ln\left(1 + \frac{\rho^t |x_{i,0} - x'_{i,0}|}{2\sigma} \cdot (\exp(\epsilon_C) - 1)\right).
\tag{89}
$$

The single-query mechanism $M^*$ that estimates the average value of $x_{i,t}$ over all users, i.e.,

$$
M(X_0) = \left[\frac{\sum_i x_{i,0}}{I}\right] + n,
\tag{90}
$$

where $n$ is a Laplace $(1/\epsilon_C)$ random variable. Similarly, we use the mean-squared estimation error as the accuracy loss metric, given by

$$
f_{\mathrm{MMSE}}(t, \epsilon_C) = \mathbb{E}\left[\left(M(X_0) - \frac{1}{I}\sum_i x_{i,t}\right)^2\right],
\tag{91}
$$

$$
= \frac{(1 - \rho^{2t})\sigma^2}{I(1 - \rho^2)} + \frac{1}{\epsilon_C^2 I^2}.
\tag{92}
$$

We present numerical results in Figure 8 which are similar to those in Figure 4. In Figure 8(a), we show that the accuracy loss increases in $t$, and for a given $\epsilon_C$, it converges to an upper bound as $t \to \infty$. In Figure 8(b), we compare our proposed mechanisms against a classical DP benchmark that only injects Laplace noise. To achieve an arbitrary small privacy risk, the accuracy loss incurred by the benchmark grows unbounded. Compared to the benchmark, combining both aging and noise injection does not incur meaningful accuracy loss, which is upper bounded as the privacy risk approaches 0.

## REFERENCES

[1] M. Zhang, E. Wei, R. Berry, and J. Huang, "Age-dependent differential privacy," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 50, no. 1, pp. 115–116, Jun. 2022.

[2] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.

[3] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Colloquium Automata, Lang. Programm.*, vol. 4052, Venice, Italy, Jul. 2006, pp. 1–12.

[4] *Differential Privacy Team Apple, Learning with Privacy at Scale*, Apple, Cupertino, CA, USA, 2017.

[5] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2014, pp. 1054–1067.

[6] J. M. Abowd, "The challenge of scientific reproducibility and privacy protection for statistical agencies," Census Scientific Advisory Committee, Suitland, MD, USA, Sep. 2016. [Online]. Available: https://www2.census.gov/cac/sac/meetings/2016-09/2016-abowd.pdf

[7] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 1–10.

[8] P. Samarati and L. Sweeney, "Generalizing data to provide anonymity when disclosing information," in *Proc. 17th ACM SIGACT-SIGMOD-SIGART Symp. Princ. Database Syst. (PODS)*, Seattle, WA, USA, 1998, p. 188.

[9] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. IEEE 23rd Int. Conf. Data Eng.*, Apr. 2007, pp. 106–115.

[10] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. IEEE Symp. Secur. Privacy*, May 2011, pp. 247–262.

[11] I. Rouf, H. Mustafa, M. Xu, W. Xu, R. Miller, and M. Gruteser, "Neighborhood watch: Security and privacy analysis of automatic meter reading systems," in *Proc. ACM Conf. Comput. Commun. Secur.*, Oct. 2012, pp. 462–473.

[12] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Comput. Surveys*, vol. 51, no. 3, pp. 1–38, May 2019.

[13] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Jun. 2011, pp. 193–204.

[14] R. Hall, A. Rinaldo, and L. Wasserman, "Random differential privacy," 2011, *arXiv:1112.2680*.

[15] B. Yang, I. Sato, and H. Nakagawa, "Bayesian differential privacy on correlated data," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, May 2015, pp. 747–762.

[16] A. Triastcyn and B. Faltings, "Bayesian differential privacy for machine learning," in *Proc. Int. Conf. Mach. Learn.*, Nov. 2020, pp. 9583–9592.

[17] D. Kifer and A. Machanavajjhala, "Pufferfish: A framework for mathematical privacy definitions," *ACM Trans. Database Syst.*, vol. 39, no. 1, pp. 3–36, Jan. 2014.

[18] S. Song, Y. Wang, and K. Chaudhuri, "Pufferfish privacy mechanisms for correlated data," in *Proc. ACM Int. Conf. Manage. Data*, New York, NY, USA, May 2017, pp. 1291–1306.

[19] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," in *Proc. 42nd ACM Symp. Theory Comput.*, Jun. 2010, pp. 715–724.

[20] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to non-interactive database privacy," in *Proc. 40th Annu. ACM Symp. Theory Comput.*, May 2008, pp. 609–618.

[21] M. Hardt and G. N. Rothblum, "A multiplicative weights mechanism for privacy-preserving data analysis," in *Proc. IEEE 51st Annu. Symp. Found. Comput. Sci.*, Oct. 2010, pp. 61–70.

[22] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth, "The reusable holdout: Preserving validity in adaptive data analysis," *Science*, vol. 349, no. 6248, pp. 636–638, Aug. 2015.

[23] A. Smith and A. G. Thakurta, "(Nearly) optimal algorithms for private online learning in full-information and bandit settings," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2013, pp. 2733–2741.

[24] T.-H.-H. Chan, E. Shi, and D. Song, "Private and continual release of statistics," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 3, pp. 1–24, Nov. 2011.

[25] R. Cummings, S. Krehbiel, K. A. Lai, and U. Tantipongpipat, "Differential privacy for growing databases," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8864–8873.

[26] F. Farokhi, "Temporally discounted differential privacy for evolving datasets on an infinite horizon," in *Proc. ACM/IEEE 11th Int. Conf. Cyber-Physical Syst. (ICCPS)*, Apr. 2020, pp. 1–8.

[27] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2015, pp. 1298–1309.

[28] Y. Xiao, X. Li, S. Zhang, and Y. Cao, "LocLok: Location cloaking with differential privacy via hidden Markov model," in *Proc. VLDB Endowment*, 2017, pp. 1901–1904.

[29] H. Wang et al., "PrivTrace: Differentially private trajectory synthesis by adaptive Markov model," in *Proc. USENIX Secur. Symp.*, 2023, pp. 1–18.

[30] Y. Cao, M. Yoshikawa, Y. Xiao, and L. Xiong, "Quantifying differential privacy in continuous data release under temporal correlations," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 7, pp. 1281–1295, Jul. 2019.

[31] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in GPS traces via uncertainty-aware path cloaking," in *Proc. 14th ACM Conf. Comput. Commun. Secur.*, New York, NY, USA, Oct. 2007, pp. 161–171.

[32] M. Wright, M. Adler, B. N. Levine, and C. Shields, "An analysis of the degradation of anonymous protocols," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, vol. 2, 2002, pp. 39–50.

[33] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki, "ARAVAN: Providing location privacy for VANET," in *Embedded Security in Cars*, 2005, pp. 29–37.

[34] R. Talak and E. H. Modiano, "Age-delay tradeoffs in queueing systems," *IEEE Trans. Inf. Theory*, vol. 67, no. 3, pp. 1743–1758, Mar. 2021.

[35] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, Mar. 2019.

[36] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5215–5232, Aug. 2019.

[37] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1681–1685.

[38] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The age of information in a discrete time queue: Stationary distribution and non-linear age mean analysis," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1352–1364, May 2021.

[39] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, Jul. 2018.

[40] C. Kam, S. Kompella, and A. Ephremides, "Age of information under random updates," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 66–70.

[41] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[42] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Optimal sampling and scheduling for timely status updates in multi-source networks," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 4019–4034, Jun. 2021.

[43] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Trans. Netw.*, vol. 28, no. 1, pp. 15–28, Feb. 2020.

[44] I. Kadota, A. Sinha, and E. Modiano, "Scheduling algorithms for optimizing age of information in wireless networks with throughput constraints," *IEEE/ACM Trans. Netw.*, vol. 27, no. 4, pp. 1359–1372, Aug. 2019.

[45] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *J. Commun. Netw.*, vol. 21, no. 3, pp. 204–219, Jun. 2019.

[46] M. Zhang, A. Arafa, J. Huang, and H. V. Poor, "Pricing fresh data," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1211–1225, May 2021.

[47] M. Zhang, A. Arafa, E. Wei, and R. Berry, "Optimal and quantized mechanism design for fresh data acquisition," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1226–1239, May 2021.

[48] Z. Wang, Q. Meng, S. Zhang, and H. Luo, "The power of age-based reward in fresh information acquisition," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2023, pp. 1–10.

[49] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The age of information in multihop networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1248–1257, Jun. 2019.

[50] R. Jin, X. He, and H. Dai, "Minimizing the age of information in the presence of location privacy-aware mobile agents," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1053–1067, Feb. 2021.

[51] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 534–556, Jan. 2020.

[52] E. T. Ceran, D. Gündüz, and A. György, "Average age of information with hybrid ARQ under a resource constraint," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1900–1913, Mar. 2019.

[53] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broad-cast wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, Dec. 2018.

[54] J. Sun, L. Wang, Z. Jiang, S. Zhou, and Z. Niu, "Age-optimal scheduling for heterogeneous traffic with timely throughput constraints," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1485–1498, May 2021.

[55] H. H. Yang, A. Arafa, T. Q. S. Quek, and H. V. Poor, "Optimizing information freshness in wireless networks: A stochastic geometry approach," *IEEE Trans. Mobile Comput.*, vol. 20, no. 6, pp. 2269–2280, Jun. 2021.

[56] X. Chen, K. Gatsis, H. Hassani, and S. S. Bidokhti, "Age of information in random access channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 10, pp. 6548–6568, Oct. 2022.

[57] B. Buyukates, A. Soysal, and S. Ulukus, "Scaling laws for age of information in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2413–2427, Apr. 2021.

[58] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, Oct. 2020.

[59] X. Wang, Z. Ning, S. Guo, M. Wen, and H. V. Poor, "Minimizing the age-of-critical-information: An imitation learning-based scheduling approach under partial observations," *IEEE Trans. Mobile Comput.*, vol. 21, no. 9, pp. 3225–3238, Sep. 2022.

[60] S. Banerjee, R. Bhattacharjee, and A. Sinha, "Fundamental limits of age-of-information in stationary and non-stationary environments," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1741–1746.

[61] R. D. Yates, "The age of gossip in networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Melbourne, VIC, Australia, Jul. 2021, pp. 2984–2989.

[62] B. Buyukates, M. Bastopcu, and S. Ulukus, "Version age of information in clustered gossip networks," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 1, pp. 85–97, Mar. 2022.

[63] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1376–1385.

[64] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, vol. 107. Providence, RI, USA: American Mathematical Society, 2017.

[65] R. C. Bradley, "Basic properties of strong mixing conditions. A survey and some open questions," 2005, *arXiv:math/0511078*.

[66] P. Diaconis and D. Stroock, "Geometric bounds for eigenvalues of Markov chains," *Ann. Appl. Probab.*, vol. 1, no. 1, pp. 36–61, Feb. 1991.

[67] H. J. Landau and A. M. Odlyzko, "Bounds for eigenvalues of certain stochastic matrices," *Linear Algebra Appl.*, vol. 38, pp. 5–15, Jun. 1981.

[68] S. Cambanis and I. Fakhre-Zakeri, "Forward and reversed time prediction of autoregressive sequences," *J. Appl. Probab.*, vol. 33, no. 4, pp. 1053–1060, Dec. 1996.

[69] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[70] M. Grant, S. Boyd, and Y. Ye. (2009). *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: http://cvxr.com/cvx/

[71] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[72] L. Michel, J. Näf, and N. Meinshausen, "High probability lower bounds for the total variation distance," 2020, *arXiv:2005.06006*.

[73] A. A. Fedotov, P. Harremoës, and F. Topsøe, "Refinements of Pinsker's inequality," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1491–1498, Jun. 2003.

[74] F. Perez-Cruz, "Kullback–Leibler divergence estimation of continuous distributions," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1666–1670.

[75] UK Power Networks. (2013). *SmartMeter Energy Consumption Data in London Households*. [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

[76] C. Dwork and G. N. Rothblum, "Concentrated differential privacy," 2016, *arXiv:1603.01887*.

**Meng Zhang** (Member, IEEE) received the Ph.D. degree in information engineering from The Chinese University of Hong Kong in 2019. He was a Visiting Student Research Collaborator with the Department of Electrical Engineering, Princeton University, from 2018 to 2019. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, Northwestern University, from 2020 to 2021. He is currently an Assistant Professor with the Zhejiang University—University of Illinois at Urbana–Champaign Institute, Zhejiang University. He also holds an Adjunct Assistant Professor position with the Electrical and Computer Engineering Department, University of Illinois at Urbana–Champaign. His primary research interests include network economics and wireless networks, with a current emphasis on mechanism design and optimization for age of information and federated learning. He was a recipient of WiOpt 2021 Best Paper Award.

**Randall Berry** (Fellow, IEEE) received the B.S. degree from the University of Missouri Rolla in 1993 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology in 1996 and 2000, respectively. He is currently the John A. Dever Professor and the Chair of Electrical and Computer Engineering with Northwestern University. He is also a Principle Engineer with Roberson and Associates and has been on the Technical Staff of the MIT Lincoln Laboratory. He was a recipient of the NSF Career Award. He has served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2006 to 2009 and an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY from 2009 to 2011. He is currently a Division Editor for the *Journal of Communications and Networks* and an Area Editor of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. He has also been a Guest Editor of special issues of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING, IEEE TRANSACTIONS ON INFORMATION THEORY, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He has served on the program and organizing committees for numerous conferences, including serving as the Chair for the 2012 IEEE Communication Theory Workshop, the TPC Chair for 2010 IEEE ICC Wireless Networking Symposium, and the TPC Chair for the 2018 ACM Mobihoc Conference.

**Ermin Wei** (Member, IEEE) received the bachelor's degree in computer engineering, in finance, and in mathematics, with a minor in German from the University of Maryland, College Park, and the Ph.D. degree in electrical engineering and computer science and the M.S. degree from MIT in 2014, advised by Prof. Asu Ozdaglar. She is currently an Associate Professor with the Electrical and Computer Engineering Department and the Industrial Engineering and Management Sciences Department, Northwestern University. Her team won the second place in the GO-competition challenge 1, an electricity grid optimization competition organized by the Department of Energy. Her research interests include distributed optimization methods, convex optimization and analysis, smart grids, communication systems and energy networks, and market economic analysis.

**Jianwei Huang** (Fellow, IEEE) received the Ph.D. degree from Northwestern University in 2005. He is currently a Presidential Chair Professor and the Associate Vice President (Institutional Development) of The Chinese University of Hong Kong, Shenzhen, and the Associate Director of the Shenzhen Institute of Artificial Intelligence and Robotics for Society. He was a Post-Doctoral Research Associate with Princeton University from 2005 to 2007. His research interests are network optimization and economics, with applications in communication networks, energy networks, data markets, and crowd intelligence. He has published more than 350 papers in leading international venues, with a Google Scholar citation of more than 16,000. He has coauthored 11 Best Paper Awards, including the 2011 IEEE Marconi Prize Paper Award in Wireless Communications. He has coauthored seven books, including the textbook *Wireless Network Pricing*. He has been an IEEE ComSoc Distinguished Lecturer, a Clarivate Web of Science Highly Cited Researcher, and an Elsevier Most Cited Chinese Researcher. He is the Editor-in-Chief of IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING and the Associate Editor-in-Chief of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.