Author/s:
Farokhi, F

Title:
Temporally discounted differential privacy for evolving datasets on an infinite horizon

Date:
2020-05-19

Citation:
Farokhi, F. (2020). Temporally discounted differential privacy for evolving datasets on an infinite horizon. Proceedings - 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems, ICCPS 2020, 00, pp.1-8. IEEE. https://doi.org/10.1109/ICCPS48487.2020.00008.

Temporally Discounted Differential Privacy for Evolving Datasets on an Infinite Horizon

Farhad Farokhi

***Abstract*—We define *discounted differential privacy*, as an alternative to (conventional) differential privacy, to investigate privacy of *evolving datasets, containing time series* over an unbounded horizon. We use *privacy loss* as a measure of the amount of information leaked by the reports at a certain fixed time. We observe that privacy losses are weighted equally across time in the definition of differential privacy, and therefore the magnitude of privacy-preserving additive noise must grow without bound to ensure differential privacy over an infinite horizon. Motivated by the discounted utility theory within the economics literature, we use *exponential and hyperbolic discounting of privacy losses* across time to relax the definition of differential privacy under continual observations. This implies that privacy losses in distant past are less important than the current ones to an individual. We use discounted differential privacy to investigate privacy of evolving datasets using additive Laplace noise and show that the magnitude of the additive noise can remain bounded under discounted differential privacy. We illustrate the quality of privacy-preserving mechanisms satisfying discounted differential privacy on smart-meter measurement time-series of real households, made publicly available by Ausgrid (an Australian electricity distribution company).**

***Index Terms*—privacy; differential privacy; evolving datasets; temporal discounting.**

I. Introduction

Real-time analytics of customer data can benefit decision making of businesses in sectors, such as energy (e.g., real-time smart-meter measurements for demand and load forecasting), intelligent transportation (e.g., real-time traffic estimation by monitoring of movements of individuals), and retail industry (e.g., real-time analysis of customer interactions and purchases with online retail services to maximize profits). Privacy concerns, however, may restrict the availability of customer data or its use in decision making. For instance, smart-meter time-series can leak private information about household occupancy, entertainment habits, and air conditioning decisions [1]–[3]. The extent of privacy concerns have sometimes proved to even hinder the roll out of smart meters [4].

Differential privacy [5]–[7] is a natural candidate to alleviate privacy concerns in general. However, differential privacy literature most often deals with providing privacy-preserving responses to queries based on large, yet static datasets that are kept securely by a data curator while, in real-time analytics, the underlying data in possession of the curator changes over time. The composition rule of differential privacy (see, e.g., [5]) implies that the magnitude of the additive noise that ensures differential privacy must grow rapidly, or that the privacy budget of each response must decrease correspondingly, to ensure that the entire privacy budget remains bounded.

Recently, better performance bounds have been derived for differentially-private responses to queries on evolving datasets [8]–[12]. These studies however consider certain sets of queries, such as counting queries [9]. In these studies, the magnitude of the additive noise still remains unbounded; the best bound is $\mathcal{O}(\log(t)^{1.5})$ with t denoting the number of observations, scaling with time. In fact, having a finite magnitude for the additive noise might not be even possible for some queries [13].

Interesting approaches to differential privacy have stemmed from the study of dynamical systems in, e.g., [14]–[19]. However, they differ from the context of this paper in multiple directions. For instance, in [14], [15], the notion of neighbouring datasets requires that at most one component signal to differ and this deviation must be bounded (in norm or count) while in this paper the deviation does not need to be bounded as the horizon over which we consider the problem can tend to infinity[1]. In [17], [20], finite horizon frameworks are considered. Although potentially infinite horizon algorithms were considered in [21]–[23], only privacy of the initial condition of the algorithm is investigated and thus the notion of neighbouring datasets only requires single deviations.

Motivated by these observations, in this paper, we relax the definition of the differential privacy by discounting privacy losses in distant past to be able to ensure privacy of evolving datasets over a possibly infinite[2] horizon. We use the notion of *privacy loss* from [5] to break down the definition of differential privacy across time. Privacy loss can be seen as a measure of the amount of information leaked by the reports at a certain fixed time. To ensure differential privacy, the summation of all privacy losses across time must be bounded by the privacy budget. We show that, because in the definition of differential privacy, the privacy losses are weighted equally across time, the magnitude of the privacy-preserving additive noise must grow unbounded to ensure differential privacy over an infinite horizon. Therefore, the reports become meaningless after a while. This motivates discounting privacy losses across time

F. Farokhi is with the CSIRO's Data61 and the University of Melbourne. e-mails: farhad.farokhi@data61.csiro.au; farhad.farokhi@unimelb.edu.au

The work of F. Farokhi was funded, in part, by the Australian Government Department of the Environment and Energy through National Energy Analytics Research (NEAR) Program.

[1]Note that, here, we assume that each deviation is bounded but the norm of the signal can be unbounded as we do not consider finite horizon deviations.

[2]Note that, in practice, an infinite horizon does not exists. However, studying infinite horizons helps us to understand cases in which an upper bound on the horizon is not known. For instance, in the smart-metering example, we may not know, in advance, the duration for which a dataset of measurements is curated and made available for real-time analytic.

to generalize, or better-said relax, the definition of differential privacy.

Discounting losses or gains across time is common place in the economics literature [24], [25]. A common practice is to exponentially discount losses or gains across time, i.e., scaling them by α^k, where α is the discount factor and k is the delay (time to or since the observation of the loss or gain). This discounting regime dates back to the early 20th century and is motivated by interest/cash rates [26], [27]. More recently, it has been shown that humans and animals follow a hyperbolic discounting regime [24], [28]–[31]. In hyperbolic discounting, losses or gains are scaled by $1/(1+\beta k)$, where β is the discounting coefficient and k is again the delay. Hyperbolic discounting has been found to relate to real-world examples of self-control [32], which makes it more interesting within the context of privacy preservation as it has been observed that privacy and self-control are heavily related in personal decision making [33].

The idea that privacy loss in a distant past is less important to an individual, and is thus discounted exponentially or hyperbolically, could be motivated by that people change their habits and addresses across time. For instance, according to the [34] 2007-08 Survey of Income and Housing, 43% of people in Australia have moved house within the last five years. Therefore, private information regarding their previous location can be deemed less sensitive. Temporal discounting is conjectured to be one of the reasons behind why few individuals take no action to protect their personal information, even when doing so involves limited costs [35]. Note that privacy behaviour (what people do) and privacy attitudes (what people think) should not be mistaken with each other as most individuals express that they are concerned about their information privacy and are willing to act to protect it [35], [36], yet in experiments they do not [37], [38]. Discounting privacy is more compatible with privacy behaviour in comparison to privacy attitudes. Finally, note that the discounting factor or coefficient can be chosen based on the preferences of an individual such that its effect is negligible over a long enough, yet finite, horizon, e.g., a person's life expectancy or active life span, albeit at the risk of reducing the quality of reports. This points to bigger issue of privacy and utility trade-off.

Discounting or decaying privacy was first studied in [39]. However, the discounting in that paper was not motivated from a human's perception of loss in the economics theory. Therefore, in that paper, only exponential discounting was considered. As stated earlier, exponential discounting does not fully capture a human's response to loss across time. In this paper, in addition to exponential discounting, we consider hyperbolic discounting of privacy losses that is a better match for a human's perception of loss in distant past. An alternative to discounting or decaying privacy losses is to only consider a recent window, which was studied in [40], [41]. These approaches are of interest if we know the underlying population that generates the evolving dataset changes over time in regular intervals (e.g., WiFi users in a pubic domain over twenty-four hours window of time) and do not match the perception of a single individual from privacy loss over an entire horizon, which can be better matched by continuous discounting rather than windowing.

In summary, in this paper, we make the following contributions:

- We use exponential and hyperbolic discounting of privacy losses across time to relax the definition of differential privacy for use with evolving datasets;
- We use discounted differential privacy to investigate privacy of evolving datasets using an additive Laplace noise and show that exponentially discounted differential privacy can be achieved with bounded magnitude of additive noise in contrast with differential privacy;
- We demonstrate the applicability of this paper's privacy-preserving mechanisms, which meet the standards of discounted differential privacy, on smart-meter measurement time series of private households, available from [42].

The rest of the paper is organized as follows. In Section II, we define privacy loss and illustrate its relationship with differential privacy. We relax the notion of differential privacy to take into account temporal discounting of privacy losses in Section III. We illustrate the results of the paper on smart-meter measurements of households in Section IV. We conclude the paper and present avenues for future research in Section V.

II. Differential Privacy and Privacy Loss

In this section, we introduce differential privacy for evolving datasets. We breakdown the definition of differential privacy across time by using privacy loss as a measure of the amount of information leaked by the reports at a certain fixed time.

A. Differential Privacy

Consider an evolving, longitudinal dataset, containing time series, of the form

$$X(t) := \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(t) \\ x_2(1) & x_2(2) & \cdots & x_2(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(1) & x_n(2) & \cdots & x_n(t) \end{bmatrix} \in \mathcal{X}^{n\times t} \subseteq \mathbb{R}^{n\times t}, \quad (1)$$

where $x_i(t) \in \mathcal{X} \subseteq \mathbb{R}$ denotes the entry for individual $i \in \{1,\dots,n\}$ at time instant $t \in \mathbb{N}$. In this paper, we assume that t is not bounded from above and can potentially approach infinity, i.e., it can grow unbounded. Note that, for any $1 \leq k \leq t$, $X(k)$ denotes a matrix extracted by eliminating the last $t-k$ columns of $X(t)$. An example of such a longitudinal dataset is a dataset containing regular smart meter reading of n fixed households.

Remark 1 (Addition/Removal of Individuals)**.** *In the dataset model in* (1)*, we can consider addition and removal of individuals to the dataset across time. In this case, for all time instants in which a measurement for an individual is not available because it has not yet arrived or has left the dataset, we can use a special characters, such as* $\emptyset$*. Hence, we must have* $\mathcal{X} = \mathbb{R} \cup \{\emptyset\}$.

We assume that, as the dataset evolves, the custodian reports

$$Y(t) = \begin{bmatrix} y(1) & y(2) & \cdots & y(t) \end{bmatrix} \in \mathcal{Y}^t \subseteq \mathbb{R}^t. \quad (2)$$

Similarly, for any $1 \leq k \leq t$, $Y(k)$ denotes a row vector extracted by eliminating the last $t-k$ entries of $Y(t)$. At time instance t, to generate the entry $y(t)$, the custodian uses conditional probability density function $p_{y|X}(\cdot|\cdot)$. From now on, we refer to this as the mechanism of the curator. The mechanisms are causal by construct, that is, at any time instant t, the report $Y(t)$ can only be a function of the entries of the longitudinal dataset up to time t, $X(t)$. In what follows, when it is evident from the context, we use $p(\cdot|\cdot)$ instead of $p_{y|X}(\cdot|\cdot)$. In this paper, we are interested in the differential privacy as a notion of privacy.

Definition 2 (Neighbouring Datasets)**.** *Two datasets $X(t)$ and $X'(t)$ are neighbouring datasets, shown by $X(t) \sim X'(t)$, if they differ from each other in at most one row.*

Definition 3 (Differential Privacy)**.** *A reporting mechanism is ϵ-differentially private for $\epsilon > 0$ if for any pair of neighbouring datasets $X(t), X'(t)$ and any output $Y(t)$,*

$$p(Y(t)|X(t)) \leq \exp(\epsilon) p(Y(t)|X'(t)). \quad (3)$$

Now, we are ready to introduce the notion of privacy loss from [5] by de-constructing the ratio of probability density functions for use within the definition of differential privacy.

B. From Differential Privacy to Privacy Loss

In this subsection, we dig deeper in the notion of differential privacy to define privacy loss. Assuming conditional independence of $Y(k)$ and $Y(k-1)$ given $X(t)$ for all $2 \leq k \leq t$, we get

$$\begin{aligned} p(Y(t)|X(t)) =& p(y(t)|X(t), Y(t-1)) p(Y(t-1)|X(t)) \\ =& p(y(t)|X(t)) p(Y(t-1)|X(t)) \\ =& p(y(t)|X(t)) p(y(t-1)|X(t)) \\ & \times p(Y(t-2)|X(t)), \end{aligned}$$

where the equalities follow from the definition of conditional probability density function. Following this line of reasoning, we get

$$\begin{aligned} p(Y(t)|X(t)) &= \prod_{k=1}^{t} p(y(k)|X(t)) \\ &= \prod_{k=1}^{t} p(y(k)|X(k)), \end{aligned}$$

where the last equality follows from the causality of the reports. This results in

$$\frac{p(Y(t)|X(t))}{p(Y(t)|X'(t))} = \prod_{k=1}^{t} \frac{p(y(k)|X(k))}{p(y(k)|X'(k))}.$$

These derivations motivate the use of $p(y(k)|X(k))/p(y(k)|X'(k))$, or in fact its logarithm, as a measure of privacy loss because if the ratio $p(y(k)|X(k))/p(y(k)|X'(k))$ is large, $y(k)$ leaks more information in terms of increasing the required differential-privacy budget for time instant k.

Definition 4 (Privacy Loss)**.** *Privacy loss due to entry $y(k)$ is*

$$\rho(k) = \sup_{y(k)} \sup_{X(k), X'(k): X(k) \sim X'(k)} \log \left(\frac{p(y(k)|X(k))}{p(y(k)|X'(k))} \right).$$

We can relate the notion of differential privacy and privacy loss together. This is explored in the next theorem.

Theorem 5 (Privacy Loss and Differential Privacy)**.** *A reporting mechanism is ϵ-differentially private for $\epsilon > 0$ if*

$$\sum_{k=1}^{t} \rho(k) \leq \epsilon. \quad (4)$$

Proof. We have

$$\begin{aligned} \frac{p(Y(t)|X(t))}{p(Y(t)|X'(t))} &= \prod_{k=1}^{t} \frac{p(y(k)|X(k))}{p(y(k)|X'(k))} \\ &\leq \prod_{k=1}^{t} \exp(\rho(k)) \\ &= \exp\left(\sum_{k=1}^{t} \rho(k) \right). \end{aligned}$$

The rest of the proof follows from the definition of differential privacy. □

Theorem 5 states that if the summation of all privacy losses is bounded from above by the total privacy budget ϵ, ϵ-differential privacy can be established.

Now, consider the case where the curator is given a family of queries $f_t : \mathcal{X}^{n \times t} \to \mathbb{R}$, $\forall t \in \mathbb{N}$, to compute on the evolving dataset. In return, the curator provides noisy reports of the form:

$$y(t) = f_t(X(t)) + w(t), \quad (5)$$

where $(w(t))_{t \in \mathbb{N}}$ is a sequence of i.i.d.[3] Laplace random variables with zero mean and scale $b_t > 0$.

Theorem 6. *The reporting mechanism* (5) *is ϵ-differentially private for $\epsilon > 0$ if*

$$\sum_{k=1}^{t} \frac{\Delta f_k}{b_k} \leq \epsilon, \quad (6)$$

where Δf_k is the sensitivity of the query defined as

$$\Delta f_k := \sup_{X(k), X'(k): X(k) \sim X'(k)} |f_k(X'(k)) - f_k(X(k))|. \quad (7)$$

[3] i.i.d. stands for independently and identically distributed.

Proof. Note that

$$\begin{aligned}\exp(\rho(k)) &= \sup_{y(k)} \sup_{X(k)\sim X'(k)} \frac{\exp(-|y(k)-f_k(X(k))|/b_k)}{\exp(-|y(k)-f_k(X'(k))|/b_k)} \\ &= \sup_{y(k)} \sup_{X(k)\sim X'(k)} \exp\left(\frac{1}{b_k}|y(k)-f_k(X'(k))| \right. \\ &\qquad \left. -\frac{1}{b_k}|y(k)-f_k(X(k))|\right) \\ &\leq \sup_{X(k)\sim X'(k)} \exp\left(\frac{1}{b_k}|f_k(X'(k))-f_k(X(k))|\right) \\ &\leq \exp\left(\frac{\Delta f_k}{b_k}\right),\end{aligned}$$

where the third equality follows from that $|a|-|b| \leq |a-b|$ because $|a| = |b-(b-a)| \leq |b|+|a-b|$. Hence, $\sum_{k=1}^{t}\rho(k) = \sum_{k=1}^{t}\Delta f_k/b_k$. The rest follows from the application of Theorem 5. □

Theorem 6 implies, with a reporting mechanism in the form of (5), we might not be able to ensure ϵ-differential privacy over an unbounded horizon unless the magnitude of the additive noise grows unbounded or the queries become gradually less intrusive by decreasing Δf_k rapidly enough. This is because, if b_t is kept constant and Δf_k does not decrease, $\sum_{k=1}^{t}\Delta f_k/b_k = +\infty$, which makes the satisfaction of the condition of Theorem 6 impossible. However, such a result might not be necessary as Theorem 6 is only sufficient. Nonetheless this seems to be line with the differential privacy literature [9], [13].

Corollary 7. *Assume that $\Delta f_k = \Delta f$ for all $k \in \mathbb{N}$. Then $\sum_{k=1}^{t}\Delta f_k/b_k \leq \epsilon$ for $\epsilon > 0$ only if $\lim_{k\to} b_k = +\infty$.*

Proof. Assume $\lim_{k\to} b_k = +\infty$ does not hold. Therefore, there exists a subsequence $(k_\ell)_{\ell\in\mathbb{N}}$ such that k_ℓ are increasing and $b_{k_\ell} \leq B$. Therefore, $\sum_{k=1}^{\infty}\Delta f_k/b_k \geq \sum_{\ell=1}^{\infty}\Delta f/b_{k_\ell} \geq \sum_{\ell=1}^{\infty}\Delta f/B = \infty$. Therefore, there exists a large enough t_0 for which $\sum_{k=1}^{t_0}\Delta f_k/b_k \leq \epsilon$ cannot be satisfied for any $t \geq t_0$. □

Corollary 8. *Assume that $\Delta f_k = \Delta f$ for all $k \in \mathbb{N}$. The reporting mechanism* (5) *is ϵ-differentially private for $\epsilon > 0$ if*

$$b_k = \frac{\Delta f \pi^2 k^2}{6\epsilon}, \quad \forall k \in \mathbb{N}. \tag{8}$$

Proof. Let $b_k = bk^2$. Hence, $\sum_{k=1}^{t}\Delta f_k/b_k \leq \sum_{k=1}^{\infty}\Delta f_k/b_k = (\Delta f/b)\pi^2/6$. Therefore, we can satisfy the condition of Theorem 6 for any t if $(\Delta f/b)\pi^2/6 \leq \epsilon$. □

These corollaries show that it might be necessary for the magnitude of the privacy-preserving additive noise to grow unbounded if we want to ensure differential privacy over an infinite horizon. This negative result could be caused by that the privacy loss at time instants k and t are weighted equally even if $k \ll t$ and, at t, the information leakage at time k is no longer relevant. This motivates discounting privacy losses across time to relax the definition of differential privacy.

III. Discounted Differential Privacy

Humans perceive the severity (i.e., magnitude) of losses and gains differently across time. This fact is captured in the economics literature, especially, within expected utility theory, by discounting losses and gains that occur a long time from now.

A. Exponentially Discounted Differential Privacy

We start with exponentially discounted privacy loss and differential privacy.

Definition 9 (Exponentially Discounted Privacy Loss)**.** *At time instant t, privacy loss due to entry $y(k)$ is*

$$\varrho(k,t) = \alpha^{t-k}\rho(k),$$

where $\alpha \in (0,1]$ is the discount factor.

Instead of privacy loss, we can use discounted privacy loss to ensure a certain level of privacy. Note that, at any time instant t, we have $\sum_{k=1}^{t}\varrho(k,t) = \sum_{k=1}^{t}\alpha^{t-k}\rho(k)$. Using this, we can define exponentially discounted differential privacy.

Definition 10 (Exponentially Discounted Differential Privacy)**.** *A reporting mechanism is (ϵ,α)-exponentially discounted differentially private for $\epsilon > 0$ and $\alpha \in (0,1]$ if*

$$\sum_{k=1}^{t}\alpha^{t-k}\rho(k) \leq \epsilon. \tag{9}$$

Note that $(\epsilon,1)$-exponentially discounted differential privacy is equivalent to ϵ-differential privacy.

Theorem 11. *The reporting mechanism* (5) *is (ϵ,α)-exponentially discounted differentially private for $\epsilon > 0$ and $\alpha \in (0,1]$ if*

$$\sum_{k=1}^{t}\alpha^{t-k}\left(\frac{\Delta f_k}{b_k}\right) \leq \epsilon. \tag{10}$$

Proof. The proof follows the same line of reasoning as in Theorem 6 while substituting the definition of ϵ-differential privacy with the definition of (ϵ,α)-exponentially discounted differential privacy. □

Corollary 12. *Assume that $\Delta f_k = \Delta f$ for all $k \in \mathbb{N}$. The reporting mechanism* (5) *is (ϵ,α)-exponentially discounted differentially private for $\epsilon > 0$ and $\alpha \in (0,1)$ if*

$$b_k = \frac{\Delta f}{\epsilon(1-\alpha)}, \quad \forall k \in \mathbb{N}. \tag{11}$$

Proof. Let $b_k = b$. Hence, $\sum_{k=1}^{t}\alpha^{t-k}\Delta f_k/b_k = (\Delta f/b)\sum_{k=0}^{t-1}\alpha^k \leq (\Delta f/b)\sum_{k=0}^{\infty}\alpha^k = (\Delta f/b)/(1-\alpha)$. Therefore, we can satisfy the condition of Theorem 11 for any $t \in \mathbb{N}$ if $(\Delta f/b)/(1-\alpha) \leq \epsilon$. □

Note that, in Corollary 12, the magnitude of the additive noise remains bounded. Therefore, the quality of the reports do not degrade with time, which is a drawback of adopting most notions of privacy for evolving datasets.

B. Hyperbolic Discounted Differential Privacy

As stated in the introduction, it has been shown that humans and animals follow a hyperbolic discounting regime. This motivate defining hyperbolic discounted differential privacy.

Definition 13 (Hyperbolic Discounted Privacy Loss)**.** *At time instant t, privacy loss due to entry $y(k)$ is*

$$\varrho'(k,t) = \frac{\rho(k)}{1+\beta(t-k)}$$

where $\beta \geq 0$ is the discounting coefficient.

In this case, we have $\sum_{k=1}^{t} \varrho'(k,t) = \sum_{k=1}^{t} \rho(k)/(1+\beta k)$. Using this, we can define hyperbolic discounted differential privacy.

Definition 14 (hyperbolic Discounted Differential Privacy)**.** *A reporting mechanism is (ϵ,β)-hyperbolic discounted differentially private for $\epsilon > 0$ and $\beta > 0$ if*

$$\sum_{k=1}^{t} \frac{1}{1+\beta(t-k)}\rho(k) \leq \epsilon. \tag{12}$$

Note that $(\epsilon, 0)$-hyperbolic discounted differential privacy is equivalent to ϵ-differential privacy.

Theorem 15. *The reporting mechanism in* (5) *is (ϵ,β)-hyperbolic discounted differentially private for $\epsilon > 0$ and $\beta \geq 0$ if*

$$\sum_{k=1}^{t} \frac{1}{1+\beta(t-k)}\frac{\Delta f_k}{b_k} \leq \epsilon. \tag{13}$$

Proof. The proof follows the same line of reasoning as in Theorem 6 while substituting the definition of ϵ-differential privacy with the definition of (ϵ,β)-hyperbolic discounted differential privacy. □

Corollary 16. *Assume that $\Delta f_k = \Delta f$. The reporting mechanism in* (5) *is (ϵ,β)-hyperbolic discounted differentially private for $\epsilon > 0$ and $\beta \geq 0$ if*

$$b_k = \frac{2\Delta f\left(\operatorname{atanh}\left(\frac{1}{\sqrt{3}}\right)+\operatorname{atanh}\left(\sqrt{\frac{\beta}{1+\beta}}\right)\right)\sqrt{k}}{\epsilon\sqrt{\beta(\beta+1)}}, \quad \forall k \in \mathbb{N}. \tag{14}$$

Proof. Let $b_k = b\sqrt{k}$. By computing the derivatives, we can check that $\Delta f_k/((1+\beta(t-k))(b\sqrt{k}))$ is decreasing up to $(\beta t+1)/3\beta$ and is increasing afterwards. Let us define $t_0 = \lfloor(\beta t+1)/(3\beta)\rfloor$. We have

$$\begin{aligned}\sum_{k=1}^{t_0} \frac{1}{1+\beta(t-k)}\frac{\Delta f_k}{b_k} \leq& \frac{\Delta f}{b}\int_0^{t_0-1} \frac{1}{1+\beta(t-x)}\frac{1}{\sqrt{x}}\mathrm{d}x \\ =& \frac{2\Delta f}{b}\frac{\operatorname{atanh}\left(\sqrt{\frac{\beta(t_0-1)}{1+\beta t}}\right)}{\sqrt{\beta(\beta t+1)}}.\end{aligned}$$

Since $\operatorname{atanh}(\cdot)$ is an increasing function and $t_0 - 1 \leq (\beta t+1)/(3\beta)$, we have

$$\operatorname{atanh}\left(\sqrt{\frac{\beta(t_0-1)}{1+\beta t}}\right) \leq \operatorname{atanh}\left(\frac{1}{\sqrt{3}}\right)$$

We can also show that

$$\begin{aligned}&\sum_{k=t_0+1}^{t} \frac{1}{1+\beta(t-k)}\frac{\Delta f_k}{b_k} \\ &\quad\leq \frac{\Delta f}{b}\int_{t_0+1}^{t} \frac{1}{(1+\beta(t-x))\sqrt{x}}\mathrm{d}x \\ &\quad= \frac{2\Delta f}{b}\frac{\left(\operatorname{atanh}\left(\sqrt{\frac{\beta t}{1+\beta t}}\right)-\operatorname{atanh}\left(\sqrt{\frac{\beta(t_0+1)}{1+\beta t}}\right)\right)}{\sqrt{\beta(\beta t+1)}} \\ &\quad\leq \frac{2\Delta f}{b}\frac{\operatorname{atanh}\left(\sqrt{\frac{\beta t}{1+\beta t}}\right)}{\sqrt{\beta(\beta t+1)}}.\end{aligned}$$

Note that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\operatorname{atanh}\left(\sqrt{\frac{\beta t}{1+\beta t}}\right)}{\sqrt{\beta(\beta t+1)}} =& \frac{\sqrt{\beta t}}{\sqrt{t}(\beta t+1)^{3/2}}\left(\sqrt{\frac{\beta t}{1+\beta t}}\right. \\ &\left.-\operatorname{atanh}\left(\sqrt{\frac{\beta t}{1+\beta t}}\right)\right) < 0,\end{aligned}$$

where the inequality follows from that $\operatorname{atanh}(x) > x$ for all $x > 0$. Therefore,

$$\frac{\operatorname{atanh}\left(\sqrt{\frac{\beta t}{1+\beta t}}\right)}{\sqrt{\beta(\beta t+1)}} \leq \frac{\operatorname{atanh}(\sqrt{\beta/(1+\beta)})}{\sqrt{\beta(\beta+1)}}.$$

This implies that

$$\begin{aligned}&\sum_{k=1}^{t} \frac{1}{1+\beta(t-k)}\frac{\Delta f_k}{b_k} \\ &\quad\leq \frac{2\Delta f}{b}\frac{\operatorname{atanh}(\sqrt{1/3})+\operatorname{atanh}(\sqrt{\beta/(1+\beta)})}{\sqrt{\beta(\beta+1)}}.\end{aligned}$$

The rest of the proof follows from the application of Theorem 15. □

In Corollary 16, the magnitude of the additive noise grows unbounded but at a much slower rate than Corollary 8. Therefore, the quality of the reports, although degrading with time, remain better than differential privacy.

IV. Numerical Results

In this section, we illustrate the results of the paper on regular smart-meter measurements from real households. We numerically investigate the quality of privacy-preserving reports ensuring discounted differential privacy using the expected difference between noisy privacy-preserving daily averages

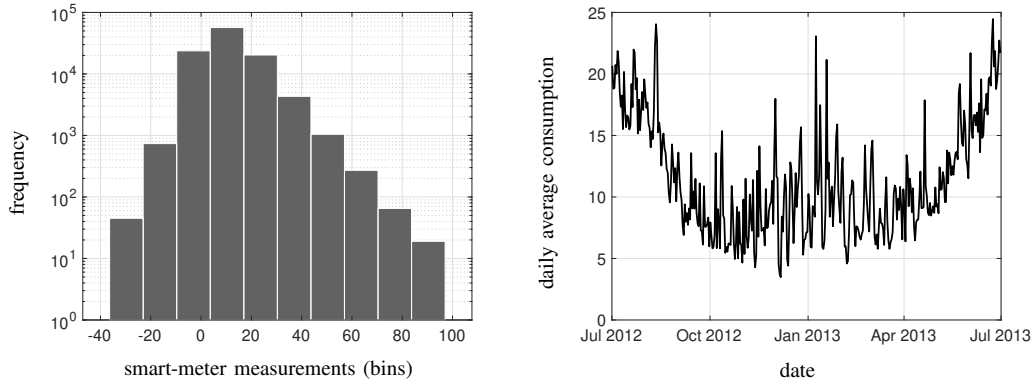

Fig. 1: Statistics of the Ausgrid data: [left] histogram of the smart-meter measurements of the households across the year and [right] average daily smart-meter measurements of the households.

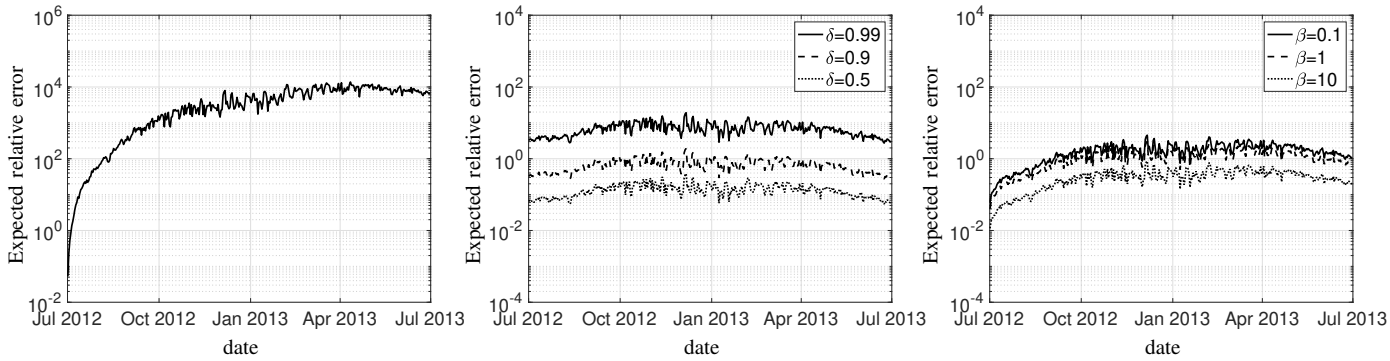

Fig. 2: Quality of reports for the Ausgrid data under various notions of privacy: differential privacy [left], exponentially discounted differential privacy [middle], and hyperbolic discounted differential privacy [right].

and true potentially privacy-intrusive average consumption of the households.

A. Data

We use smart meter measurements of of households, made available by the [42] to illustrate the results of this paper. The individuals in both these datasets have been de-identified. The dataset in [42] contains electricity data for 300 homes with rooftop solar systems that are measured by a smart meter that, in addition to measuring the usage from the grid, records the total amount of solar power generated. The measurements are obtained every 30 minutes over 2010-2013. In this paper, we use the data over July 2012 to June 2013. Figure 1 illustrates the statistics of the Ausgrid data. Figure 1 [left] shows the histogram of the smart-meter measurements of the households across the year. Figure 1 [right] illustrates the daily smart-meter measurements of the households, averaged across the individuals. This is the intended report that must to be released using the privacy-preserving mechanism in (5). Therefore,

$$f_t(X(t)) = \frac{1}{n}\sum_{i=1}^{n} x_i(t).$$

It can be seen that, for this example, $\Delta f = 200/300$ in which 200 kWh is maximum changes in consumption across a household, according to Figure 1 [left], and 300 is the number of households.

B. Setup

We present daily consumption of the households across the year, averaged for the individuals. We use the reporting mechanism (5) to ensure the privacy of the households. We consider three setups of differential privacy, exponentially discounted differential privacy, and hyperbolic discounted differential privacy. To this aim, we use the results of Corollaries 8, 12, and 16 to ensure that the reports are privacy preserving in their corresponding notions. We are interested in investigating the quality of the privacy-preserving reports, i.e., the difference between the reports and the average consumption of the households. Particularly, we use the expected relative error, defined as

$$\text{expected relative error } (t) := \frac{\mathbb{E}\left\{\left|y(t) - \frac{1}{n}\sum_{i=1}^{n} x_i(t)\right|\right\}}{\left|\frac{1}{n}\sum_{i=1}^{n} x_i(t)\right|},$$

as a measure of quality. We also use average expected relative error, which is defined as

$$\text{avrage expected relative error} = \frac{1}{T}\sum_{t=1}^{T} \text{expected relative error } (t),$$

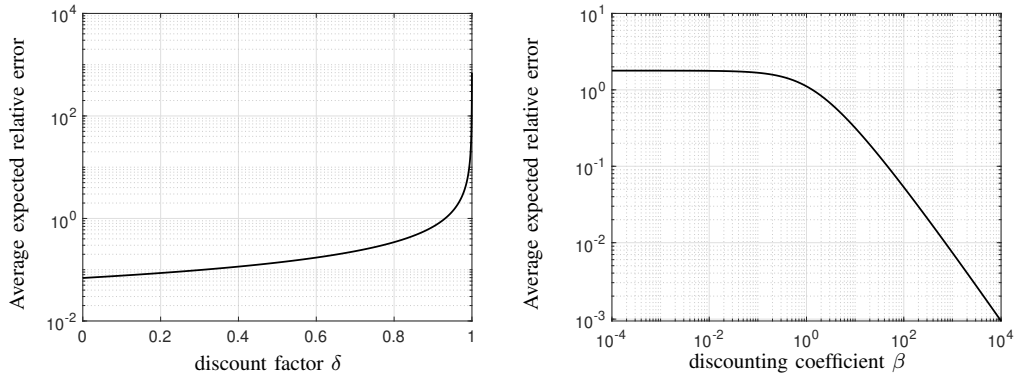

Fig. 3: Quality of reports for the Ausgrid data versus [left] discount factor α and [right] discounting coefficient β.

where T denotes the horizon of the experiment, one year in this paper.

C. Results

Figure 2 shows the quality of reports in (5) under various notions of privacy: differential privacy [left], exponentially discounted differential privacy [middle], and hyperbolic discounted differential privacy [right]. As expected, quality of reports for ϵ-differential private reporting policy is bad and degrades with time, as the magnitude of the noise needs to be increased to ensure differential privacy. It is interesting to note that, in the hyperbolic discounted differential privacy, because the increase in the magnitude of the privacy-preserving noise is so low that we cannot observe its disruptive effect within 1 year.

To quantify the effects of the discount factor and discounting coefficient in discounted differential privacy, we can study the average expected relative error in Figure 3. As expected, by increasing the discounting across time, which can be achieved by decreasing discount factor or increasing discounting coefficient, the performance improves. However, the privacy guarantee also weakens as privacy losses from the past are dismissed faster, which might not be desirable.

V. Conclusions and Future Work

We defined discounted differential privacy to investigate privacy in the context of evolving datasets. We used exponential and hyperbolic discounting of privacy losses across time to relax the definition of differential privacy under continual observations. We used discounted differential privacy to investigate privacy of evolving datasets using an additive Laplace noise. We illustrate the quality of privacy-preserving mechanisms satisfying discounted differential privacy on smart-meter measurement. Future work can focus on capturing the effect of temporal discounting on the ability of an adversary to observe private information of an individual household.

Acknowledgements

The author is thankful to the anonymous reviewers for improving the presentation of the paper and for spotting some flaws in the preliminary proofs that was fixed in the final version of the paper.

References

[1] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security & Privacy*, vol. 7, no. 3, pp. 75–77, 2009.

[2] U. Greveler, B. Justus, and D. Loehr, "Multimedia content identification through smart meter power usage profiles," *Computers, Privacy and Data Protection*, vol. 1, no. 10, 2012.

[3] S. S. Hosseini, K. Agbossou, S. Kelouwani, and A. Cardenas, "Non-intrusive load monitoring through home energy management systems: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1266–1274, 2017.

[4] C. Cuijpers and B.-J. Koops, "Smart metering and privacy in Europe: lessons from the Dutch case," in *European data protection: Coming of age*, pp. 269–293, Springer, 2013.

[5] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference*, pp. 265–284, Springer, 2006.

[7] C. Dwork, "Differential privacy," in *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II*, ICALP'06, (Berlin, Heidelberg), pp. 1–12, Springer-Verlag, 2006.

[8] M. Joseph, A. Roth, J. Ullman, and B. Waggoner, "Local differential privacy for evolving data," in *Advances in Neural Information Processing Systems*, pp. 2375–2384, 2018.

[9] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," in *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pp. 715–724, ACM, 2010.

[10] T.-H. H. Chan, E. Shi, and D. Song, "Private and continual release of statistics," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 3, p. 26, 2011.

[11] V. Perrier, H. J. Asghar, and D. Kaafar, "Private continual release of real-valued data streams," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2019.

[12] R. Cummings, S. Krehbiel, K. A. Lai, and U. Tantipongpipat, "Differential privacy for growing databases," in *Advances in Neural Information Processing Systems*, pp. 8864–8873, 2018.

[13] S. Vadhan, "The complexity of differential privacy," in *Tutorials on the Foundations of Cryptography*, pp. 347–450, Springer, 2017.

[14] J. Le Ny and G. J. Pappas, "Differentially private kalman filtering," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1618–1625, IEEE, 2012.

[15] J. Le Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, 2013.

[16] J. Cortés, G. E. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 4252–4272, IEEE, 2016.

[17] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud, "Differential privacy in linear distributed control systems: Entropy minimizing mechanisms and performance tradeoffs," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 118–130, 2017.

[18] S. Han and G. J. Pappas, "Privacy in control and dynamical systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 309–332, 2018.

[19] J. Le Ny and M. Mohammady, "Differentially private MIMO filtering for event streams," *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 145–157, 2017.

[20] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 50–64, 2016.

[21] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus with optimal noise selection," *IFAC-PapersOnLine*, vol. 48, no. 22, pp. 203–208, 2015.

[22] Y. Mo and R. M. Murray, "Privacy preserving average consensus," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 753–765, 2016.

[23] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design," *Automatica*, vol. 81, pp. 221–231, 2017.

[24] J. Myerson and L. Green, "Discounting of delayed rewards: Models of individual choice," *Journal of the Experimental Analysis of Behavior*, vol. 64, no. 3, pp. 263–276, 1995.

[25] G. S. Berns, D. Laibson, and G. Loewenstein, "Intertemporal choice–toward an integrative framework," *Trends in Cognitive Sciences*, vol. 11, no. 11, pp. 482–488, 2007.

[26] F. P. Ramsey, "A mathematical theory of saving," *The Economic Journal*, vol. 38, no. 152, pp. 543–559, 1928.

[27] P. A. Samuelson, "A note on measurement of utility," *The Review of Economic Studies*, vol. 4, pp. 155–161, 02 1937.

[28] G. Ainslie, "Specious reward: A behavioral theory of impulsiveness and impulse control," *Psychological Bulletin*, vol. 82, no. 4, p. 463, 1975.

[29] G. W. Ainslie, "Impulse control in pigeons," *Journal of the Experimental Analysis of Behavior*, vol. 21, no. 3, pp. 485–489, 1974.

[41] Y. Chen, A. Machanavajjhala, M. Hay, and G. Miklau, "Pegasus: Data-adaptive differentially private stream processing," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1375–1388, ACM, 2017.

[30] G. Ainslie and R. J. Herrnstein, "Preference reversal and delayed reinforcement," *Animal Learning & Behavior*, vol. 9, no. 4, pp. 476–482, 1981.

[31] K. N. Kirby, "Bidding on the future: Evidence against normative discounting of delayed rewards," *Journal of Experimental Psychology: General*, vol. 126, no. 1, p. 54, 1997.

[32] R. E. Vuchinich and C. A. Simpson, "Hyperbolic temporal discounting in social drinkers and problem drinkers," *Experimental and Clinical Psychopharmacology*, vol. 6, no. 3, p. 292, 1998.

[33] V. J. Derlega and A. L. Chaikin, "Privacy and self-disclosure in social relationships," *Journal of Social Issues*, vol. 33, no. 3, pp. 102–115, 1977.

[34] Australian Bureau of Statistics (ABS), "Australian social trends December 2010: Moving house," 2010. https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/LookupAttach/4102.0Publication14.12.104/$File/41020_housingmobility2010.pdf.

[35] A. Acquisti and J. Grossklags, "Privacy attitudes and privacy behavior," in *Economics of Information Security*, pp. 165–178, Springer, 2004.

[36] J. Rosen, "The right to be forgotten," *Stanford Law Review*, vol. 64, p. 88, 2011.

[37] R. K. Chellappa and R. G. Sin, "Personalization versus privacy: An empirical examination of the online consumer's dilemma," *Information Technology and Management*, vol. 6, no. 2, pp. 181–202, 2005.

[38] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior," in *Proceedings of the 3rd ACM conference on Electronic Commerce*, pp. 38–47, ACM, 2001.

[39] J. Bolot, N. Fawaz, S. Muthukrishnan, A. Nikolov, and N. Taft, "Private decayed predicate sums on streams," in *Proceedings of the 16th International Conference on Database Theory*, pp. 284–295, ACM, 2013.

[40] G. Kellaris, S. Papadopoulos, X. Xiao, and D. Papadias, "Differentially private event sequences over infinite streams," *Proceedings of the VLDB Endowment*, vol. 7, no. 12, pp. 1155–1166, 2014.

[42] Ausgrid, "Solar home electricity data," 2014. https://www.ausgrid.com.au/Industry/Innovation-and-research/Data-to-share/Solar-home-electricity-data.