51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Differentially Private Filtering

Jerome Le Ny and George J. Pappas

Abstract— **Emerging systems such as smart grids or intelligent transportation systems often require end-user applications to continuously send information to external data aggregators performing monitoring or control tasks. This can result in an undesirable loss of privacy for the users in exchange of the benefits provided by the application. Motivated by this trend, we introduce privacy concerns in a system theoretic context, and address here the problem of releasing filtered signals that respect the privacy of the input data stream. We rely on a formal notion of privacy introduced in the database literature, called *differential privacy*, which provides strong privacy guarantees against adversaries with arbitrary side information, and extend this notion to dynamic systems. We then describe methods to approximate a given filter by a differentially private version, so that the distortion introduced by the privacy mechanism is minimized. Two specific scenarios are considered, where users either provide independent input signals or contribute events to a single integer-valued stream.**

I. Introduction

A rapidly growing number of applications require users to release private data streams to third-party applications for signal processing and decision-making purposes. Examples include smart grids, health monitoring, traffic monitoring, fuel consumption optimization, and cloud computing for industrial control systems. For privacy or security reasons, the participants benefiting from the services provided by these systems generally do not want to release more information than strictly necessary. In a smart grid for example, a customer could receive better rates in exchange of continuously sending to the utility company her instantaneous power consumption, helping to improve the demand forecast mechanism. In doing so however, she is also informing the utility or a potential eavesdropper about the type of appliances she owns as well as her daily activities [1]. Hence the development of rigorous privacy preserving mechanisms is crucial to increase the level of user participation, which can in turn greatly improve the efficiency of these large-scale systems.

J. Le Ny is with the department of Electrical Engineering, Ecole Polytechnique de Montreal, QC H3T 1J4, Canada. G. Pappas is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. jerome.le-ny@polymtl.ca, pappasg@seas.upenn.edu.

Precisely defining what constitutes a breach of privacy is a delicate task. A particularly successful recent definition of privacy used in the database literature is that of *differential privacy* [2], which is motivated by the fact that any useful information provided by a dataset about a group of people can compromise the privacy of specific individuals due to the existence of side information. Differentially private mechanisms randomize their responses to dataset analysis requests and guarantee that whether or not any individual chooses to contribute her data only marginally changes the distribution over the published outputs. As a result, even an adversary cross-correlating these outputs with other sources of information cannot infer much more about specific individuals after publication than before [3].

Most work related to privacy is concerned with the analysis of static databases [2], [4] whereas cyber-physical systems clearly emphasize the need for mechanisms working with dynamic, time-varying data streams. In this context, a differentially private version of the iterative averaging algorithm for consensus is developped in [5]. Recently, information-theoretic approaches have also been proposed to guarantee some level of privacy when releasing time series [6], [7]. However, the resulting privacy guarantees only hold if the statistics of the participants' data streams obey the assumptions made (typically stationarity, dependence and distributional assumptions), and require the explicit statistical modeling of all available side information.

The main contribution of this paper is to introduce privacy concerns in the context of systems theory. Section II provides some technical background on differential privacy. We then formulate in Section III the problem of releasing the output of a dynamical system while preserving differential privacy of the driving inputs, assumed to originate from different participants. It is shown that accurate results can be published for systems with small incremental gains with respect to the individual input channels. Section IV is motivated by the recent work on "differential privacy under continual observation" [8], [9], and considers systems processing a single integer-valued signal describing events originating from many individual participants. Differentially private approximations of the systems are proposed with the

978-1-4673-2066-5/12/$31.00 ©2012 IEEE

goal of minimizing the mean squared error introduced by the privacy preserving mechanism. Proofs that are omitted due to space constraints can be found in [10].

Notation. All signals are discrete-time signals, start at time 0, and all systems are assumed to be causal. Let P_T be the truncation operator, i.e., $(P_T x)_t$ equals x_t if $t \leq T$ and 0 otherwise. We denote by $\ell^m_{p,e}$ the space of sequences with values in $\mathbb{R}^m$ and such that $x \in \ell^m_{p,e}$ if and only if $P_T x$ has finite p-norm for all integers T. The $\mathcal{H}_2$ norm and $\mathcal{H}_\infty$ norm of a stable transfer function $\mathcal{G}$ are denoted $\|\mathcal{G}\|_2$ and $\|\mathcal{G}\|_\infty$ respectively.

II. Differential Privacy

In this section we review the notion of differential privacy [2] and some basic mechanisms that can be used to achieve it. Most of the results in this section are known, but in some cases we provide more precise or slightly different versions of some statements made in previous work. We refer the reader to the surveys by Dwork, e.g., [11], for additional background on differential privacy.

A. Definition

Let us fix some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let D be a space of datasets of interest (e.g., a space of databases, or a signal space). A *mechanism* is just a measurable map $M : \mathsf{D} \times \Omega \to \mathsf{R}$, for some measurable output space $(\mathsf{R}, \mathcal{M})$, where $\mathcal{M}$ denotes a σ-algebra. In particular, for any element $d \in \mathsf{D}$, $M(d, \cdot)$ is a random variable, and we typically write simply $M(d)$. A mechanism can be viewed as a probabilistic algorithm to answer a query q, which is a map $q : \mathsf{D} \to \mathsf{R}$. In some cases, we index the mechanism by the query q of interest, writing M_q.

Example 2.1: Let $\mathsf{D} = \mathbb{R}^n$, with each entry of $d \in \mathsf{D}$ corresponding to some sensitive information for an individual contributing her data, e.g., her salary. A data analyst would like to know the average of the entries of d, i.e., $q : \mathsf{D} \to \mathbb{R}$ is defined by $q(d) = \frac{1}{n}\sum_{i=1}^n d_i$. As detailed in Section II-B, a typical mechanism M_q to answer this query in a differentially private way computes $q(d)$ and blurs the result by adding a random variable $Y : \Omega \to \mathbb{R}$, so that $M_q(d) = \frac{1}{n}\sum_{i=1}^n d_i + Y$. Note that in the absence of perturbation Y, an adversary who knows n and $d_j, j \geq 2$ can recover the remaining entry d_1 exactly if he learns $q(d)$. This can deter people from contributing their data, even though broad-based participation improves the accuracy of the analysis and can be beneficial to the population as a whole.

Next, we introduce the definition of differential privacy [2]. Intuitively, in the following definition, D is a space of datasets of interest, and we have a symmetric binary relation Adj on D, called adjacency, such that $\mathrm{Adj}(d, d')$ if and only if d and d' differ by the data of a single participant.

Definition 1: Let D be a space equipped with a symmetric binary relation denoted Adj, and let $(\mathsf{R}, \mathcal{M})$ be a measurable space. Let $\epsilon, \delta \geq 0$. A mechanism $M : \mathsf{D} \times \Omega \to \mathsf{R}$ is (ϵ, δ)-differentially private if for all $d, d' \in \mathsf{D}$ such that $\mathrm{Adj}(d, d')$, we have

$$\mathbb{P}(M(d) \in S) \leq e^\epsilon \mathbb{P}(M(d') \in S) + \delta, \quad \forall S \in \mathcal{M}. \quad (1)$$

If $\delta = 0$, the mechanism is said to be ϵ-differentially private.

Intuitively, this definition says that for two adjacent datasets, the distributions over the outputs of the mechanism should be close. The choice of the parameters ϵ, δ is set by the privacy policy. Typically ϵ is taken to be a small constant, e.g., $\epsilon \approx 0.1$ or perhaps even $\ln 2$ or $\ln 3$. The parameter δ should be kept small as it controls the probability of certain significant losses of privacy, e.g., when a zero probability event for input d' becomes an event with positive probability for input d in (1).

A fundamental property of the notion of differential privacy is that no additional privacy loss can occur by simply manipulating an output that is differentially private. To state it, recall that a probability kernel between two measurable spaces $(\mathsf{R}_1, \mathcal{M}_1)$ and $(\mathsf{R}_2, \mathcal{M}_2)$ is a function $k : \mathsf{R}_1 \times \mathcal{M}_2 \to [0, 1]$ such that $k(\cdot, S)$ is measurable for each $S \in \mathcal{M}_2$ and $k(r, \cdot)$ is a probability measure for each $r \in \mathsf{R}_1$.

Theorem 1 (Resilience to post-processing): Let $M_1 : \mathsf{D} \times \Omega \to (\mathsf{R}_1, \mathcal{M}_1)$ be an (ϵ, δ)-differentially private mechanism. Let $M_2 : \mathsf{D} \times \Omega \to (\mathsf{R}_2, \mathcal{M}_2)$ be another mechanism, such that there exists a probability kernel $k : \mathsf{R}_1 \times \mathcal{M}_2 \to [0, 1]$ verifying

$$\mathbb{P}(M_2(d) \in S | M_1(d)) = k(M_1(d), S), \text{ a.s.}, \quad (2)$$

for all $S \in \mathcal{M}_2$ and all $d \in \mathsf{D}$. Then M_2 is (ϵ, δ)-differentially private.

Note that in (2), the kernel k is not allowed to depend on the dataset d. In other words, this condition says that once $M_1(d)$ is known, the distribution of $M_2(d)$ does not further depend on d. The theorem shows that a mechanism M_2 accessing a dataset only indirectly via the output of a differentially private mechanism M_1 cannot weaken the privacy guarantee. Hence post-processing can be used freely to improve the *accuracy* of an output, as in Section IV for example, without sacrificing privacy.

B. Basic Differentially Private Mechanisms

A mechanism that throws away all the information in a dataset is obviously private, but not useful, and in general one has to trade off privacy for utility when

answering specific queries. We recall below two basic mechanisms that can be used to answer queries in a differentially private way. We are only concerned in this section with queries that return numerical answers, i.e., here a query is a map $q : \mathsf{D} \to \mathsf{R}$, where the output space R equals $\mathbb{R}^k$ for some $1 \leq k < \infty$, is equipped with a norm denoted $\|\cdot\|_{\mathsf{R}}$, and the σ-algebra $\mathcal{M}$ on R is taken to be the standard Borel σ-algebra, denoted $\mathcal{R}^k$. The following quantity plays an important role in the design of differentially private mechanisms [2].

Definition 2: Let D be a space equipped with an adjacency relation Adj. The sensitivity of a query $q : \mathsf{D} \to \mathsf{R}$ is defined as $\Delta_{\mathsf{R}} q := \max_{d,d':\mathrm{Adj}(d,d')} \|q(d) - q(d')\|_{\mathsf{R}}$. In particular, for $\mathsf{R} = \mathbb{R}^k$ equipped with the p-norm $\|x\|_p = \left(\sum_{i=1}^k |x_i|^p\right)^{1/p}$ for $p \in [1, \infty]$, we denote the ℓ_p sensitivity by $\Delta_p q$.

1) The Laplace Mechanism: This mechanism, proposed in [2], modifies an answer to a numerical query by adding i.i.d. zero-mean noise distributed according to a Laplace distribution. Recall that the Laplace distribution with mean zero and scale parameter b, denoted $\mathrm{Lap}(b)$, has density $p(x; b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$ and variance $2b^2$. Moreover, for $w \in \mathbb{R}^k$ with w_i iid and $w_i \sim \mathrm{Lap}(b)$, denoted $w \sim \mathrm{Lap}(b)^k$, we have $p(w; b) = (\frac{1}{2b})^k \exp\left(-\frac{\|w\|_1}{b}\right)$ and $\mathbb{E}[\|w\|_1] = b$.

Theorem 2: Let $q : \mathsf{D} \to \mathbb{R}^k$ be a query. Then the Laplace mechanism $M_q : \mathsf{D} \times \Omega \to \mathbb{R}^k$ defined by $M_q(d) = q(d) + w$, with $w \sim \mathrm{Lap}\,(b)^k$ and $b \geq \frac{\Delta_1 q}{\epsilon}$ is ϵ-differentially private.

Note that the mechanism requires *each* coordinate of w to have standard deviation proportional to $\Delta_1 q$, as well as inversely proportional to the privacy parameter ϵ (here $\delta = 0$). For example, if q simply consists of k repetitions of the same scalar query, then $\Delta_1 q$ increases linearly with k, and the quadratically growing variance of the noise added to each coordinate prevents an adversary from averaging out the noise.

2) The Gaussian Mechanism: This mechanism, proposed in [4], is similar to the Laplace mechanism but adds i.i.d. Gaussian noise to obtain (ϵ, δ)-differential privacy, with $\delta > 0$ but typically a smaller ϵ for the same utility. Recall the definition of the $\mathcal{Q}$-function $\mathcal{Q}(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$. The following theorem tightens the analysis from [4].

Theorem 3: Let $q : \mathsf{D} \to \mathbb{R}^k$ be a query. Then the Gaussian mechanism $M_q : \mathsf{D} \times \Omega \to \mathbb{R}^k$ defined by $M_q(d) = q(d) + w$, with $w \sim \mathcal{N}\left(0, \sigma^2 I_k\right)$, where $\sigma \geq \frac{\Delta_2 q}{2\epsilon}(K + \sqrt{K^2 + 2\epsilon})$ and $K = \mathcal{Q}^{-1}(\delta)$, is (ϵ, δ)-differentially private.

For the rest of the paper, we define $\kappa(\delta, \epsilon) = \frac{1}{2\epsilon}(K + \sqrt{K^2 + 2\epsilon})$, so that the standard deviation σ in Theorem 3 can be written $\sigma(\delta, \epsilon) = \kappa(\epsilon, \delta)\Delta_2 q$. It can be shown that $\kappa(\delta, \epsilon)$ can be bounded by $O(\ln(1/\delta))^{1/2}/\epsilon$.

III. Differentially Private Dynamic Systems

We now consider situations in which private participants contribute input signals driving a dynamic system and the queries consist of output signals of this system. First, in this section, we assume that the input of a system consists of n signals, one for each participant. An input signal is denoted $u = (u_1, \ldots, u_n)$, with $u_i \in \ell^{m_i}_{r_i,e}$ for some $m_i \in \mathbb{N}$ and $r_i \in [1, \infty]$. A simple example is that of a dynamic system releasing at each period the average over the past l periods of the sum of the input values of the participants, i.e., with output $\frac{1}{l}\sum_{k=t-l+1}^{t}\sum_{i=1}^{n} u_{i,k}$ at time t. For $r = (r_1, \ldots, r_n)$ and $m = (m_1, \ldots, m_n)$, an adjacency relation can be defined on $l^m_{r,e} = \ell^{m_1}_{r_1,e} \times \ldots \times \ell^{m_n}_{r_n,e}$ for example by $\mathrm{Adj}(u, u')$ if and only if u and u' differ by exactly one component signal, and moreover this deviation is bounded. That is, let us fix a set of nonnegative numbers $b = (b_1, \ldots, b_n)$, $b_i \geq 0$, and define

$$\mathrm{Adj}^b(u, u') \text{ iff for some } i, \|u_i - u'_i\|_{r_i} \leq b_i, \text{ and } u_j = u'_j \text{ for all } j \neq i. \quad (3)$$

A. Finite-Time Criterion for Differential Privacy

To approximate dynamic systems by versions respecting the differential privacy of the individual participants, we consider mechanisms of the form $M : \ell^m_{r,e} \times \Omega \to \ell^{m'}_{s,e}$, i.e., producing for any input signal $u \in \ell^m_{r,e}$ a stochastic process Mu with sample paths in $\ell^{m'}_{s,e}$. As in the previous section, this requires that we first specify the measurable sets of $\ell^{m'}_{s,e}$. We start by defining in a standard way the measurable sets of $(\mathbb{R}^{m'})^{\mathbb{N}}$, the space of sequences with values in $\mathbb{R}^{m'}$, to be the σ-algebra denoted $\mathcal{M}^{m'}$ generated by the so-called finite-dimensional cylinder sets of the form $\{y \in (\mathbb{R}^{m'})^{\mathbb{N}} : y_{0:T} \in H_T\}$, for $T \geq 0$ and $H_T \in \mathcal{R}^{(T+1)m'}$, where $y_{0:T}$ denotes the vector $[y_0^T, \ldots, y_T^T]^T$ (see, e.g., [12, chapter 2]). The measurable sets considered for the output of M are then obtained by intersection of $\ell^{m'}_{s,e}$ with the sets of $\mathcal{M}^{m'}$. The resulting σ-algebra is denoted $\mathcal{M}^{m'}_{s,e}$ and is generated by the sets of the form $\tilde{H}_T = \{y \in \ell^{m'}_{s,e} : y_{0:T} \in H_T\}$, for $T \geq 0, H_T \in \mathcal{R}^{(T+1)m'}$. The following technical lemma is useful to show that a mechanism on signal spaces is (ϵ, δ)-differentially private by considering only finite dimensional problems.

Lemma 4: Consider an adjacency relation Adj on $\ell^m_{r,e}$. For a mechanism $M : \ell^m_{r,e} \times \Omega \to \ell^{m'}_{s,e}$, the following are equivalent

(a) M is (ϵ, δ)-differentially private.
(b) For all u, u' in $\ell^m_{r,e}$ such that $\mathrm{Adj}(u, u')$, we have

$$\mathbb{P}((Mu)_{0:T} \in A) \leq e^{\epsilon}\, \mathbb{P}((Mu')_{0:T} \in A) + \delta, \quad (4)$$

for all $T \geq 0$ and all $A \in \mathcal{R}^{(T+1)m'}$.

B. Basic Dynamic Mechanisms

Recall (see, e.g., [13]) that for a system $\mathcal{G}$ with inputs in $\ell^m_{r,e}$ and output in $\ell^{m'}_{s,e}$, its ℓ_r-to-ℓ_s incremental gain $\gamma^{inc}_{r,s}(\mathcal{G})$ is defined as the smallest number γ such that

$$\|P_T\mathcal{G}u - P_T\mathcal{G}u'\|_s \leq \gamma \|P_T u - P_T u'\|_r, \forall u, u' \in \ell^m_{r,e}, \forall T.$$

Now consider, for $r = (r_1, \ldots, r_n)$ and $m = (m_1, \ldots, m_n)$, a system $\mathcal{G} : l^m_{r,e} \to \ell^{m'}_{s,e}$ defined by

$$\mathcal{G}(u_1, \ldots, u_n) = \sum_{i=1}^{n} \mathcal{G}_i u_i, \quad (5)$$

where $\mathcal{G}_i : \ell^{m_i}_{r_i,e} \to \ell^{m'}_{s,e}$, for all $1 \leq i \leq n$. The next theorem generalizes the Laplace and Gaussian mechanisms of Theorems 2 and 3 to causal dynamic systems.

Theorem 5: Let $\mathcal{G}$ be defined as in (5) and consider the adjacency relation (3). Then the mechanism $Mu = \mathcal{G}u + w$, where w is a white noise with $w_t \sim \mathrm{Lap}(B/\epsilon)^{m'}$ and $B \geq \max_{1\leq i\leq n}\{\gamma^{inc}_{r_i,1}(\mathcal{G}_i)\, b_i\}$, is ϵ-differentially private. The mechanism is (ϵ, δ)-differentially private if $w_t \sim \mathcal{N}(0, \sigma^2 I_{m'})$, with $\sigma \geq \kappa(\delta, \epsilon) \max_{1\leq i\leq n}\{\gamma^{inc}_{r_i,2}(\mathcal{G}_i)\, b_i\}$.

Proof: Consider two adjacent signals u, u', differing say in their i^{th} component. Then, for $\alpha \in \{1, 2\}$, we have

$$\begin{aligned}\|P_T\mathcal{G}u - P_T\mathcal{G}u'\|_\alpha &= \|P_T\mathcal{G}_i u_i - P_T\mathcal{G}_i u'_i\|_\alpha \\ &\leq \gamma_{r_i,\alpha}\|P_T u_i - P_T u'_i\|_{r_i} \\ &\leq \gamma_{r_i,\alpha}\|u_i - u'_i\|_{r_i} \\ &\leq \gamma_{r_i,\alpha} b_i.\end{aligned}$$

This leads to a bound on the ℓ_1 and ℓ_2 sensitivity of $P_T\mathcal{G}$, valid for all T. The result is then an application of Theorems 2 and 3 and Lemma 4, since (4) is satisfied for all T. ∎

Corollary 1: Let $\mathcal{G}$ be defined as in (5) with each system $\mathcal{G}_i$ linear, and $r_i = 2$ for all $1 \leq i \leq n$. Then the mechanism $Mu = \mathcal{G}u + w$, where w is a white Gaussian noise with $w_t \sim \mathcal{N}(0, \sigma^2 I_{m'})$ and $\sigma \geq \kappa(\delta, \epsilon) \max_{1\leq i\leq n}\{\|\mathcal{G}_i\|_\infty\, b_i\}$, is (ϵ, δ)-differentially private for (3).

C. Filter Approximation Set-ups for Differential Privacy

Let $r_i = 2$ for all i and $\mathcal{G}$ be linear as in the Corollary 1, and assume for simplicity the same bound $b_1^2 = \ldots = b_n^2 = B$ for the allowed variations in energy of each input signal. We have then two simple mechanisms producing a differentially private version of $\mathcal{G}$. The first one directly perturbs each input signal u_i by adding to it a white Gaussian noise w_i with $w_{i,t} \sim \mathcal{N}(0, \sigma^2 I_{m_i})$ and $\sigma^2 = \kappa(\delta, \epsilon)^2 B$. These perturbations on each input channel are then passed through $\mathcal{G}$, leading to a mean squared error (MSE) for the output equal to $\kappa(\delta, \epsilon)^2 B \|\mathcal{G}\|_2^2 = \kappa(\delta, \epsilon)^2 B \sum_{i=1}^n \|\mathcal{G}_i\|_2^2$. Alternatively, we can add a single source of noise at the output of $\mathcal{G}$ according to Corollary 1, in which case the MSE is $\kappa(\delta, \epsilon)^2 B \max_{1\leq i\leq n}\{\|\mathcal{G}_i\|_\infty^2\}$. Both of these schemes should be evaluated depending on the system $\mathcal{G}$ and the number n of participants, as none of the error bound is better than the other in all circumstances.

Example 3.1: Consider again the problem of releasing the average over the past l periods of the sum of the input signals, i.e., $\mathcal{G} = \sum_{i=1}^n \mathcal{G}_i$ with $(\mathcal{G}_i u_i)_t = \frac{1}{l}\sum_{k=t-l+1}^{t} u_{i,k}$, for all i. Then $\|\mathcal{G}_i\|_2^2 = 1/l$, whereas $\|\mathcal{G}_i\|_\infty = 1$, for all i. The MSE for the scheme with the noise at the input is then $\kappa(\delta, \epsilon)^2 Bn/l$. With the noise at the output, the MSE is $\kappa(\delta, \epsilon)^2 B$, which is better exactly when $n > l$, i.e., the number of users is larger than the averaging window.

IV. Filtering Event Streams

This section considers an application scenario motivated by the work of [8], [9] Assume now that an input signal is integer valued, i.e., $u_t \in \mathbb{Z}$ for all $t \geq 0$. Such a signal can record the occurrences of events of interest over time, e.g., the number of transactions on a commercial website, or the number of people newly infected with a virus. As in [8], [9], two signals u and u' are adjacent if and only if they differ by one at a single time, or equivalently

$$\mathrm{Adj}(u, u') \text{ iff } \|u - u'\|_1 = 1. \quad (6)$$

The motivation for this adjacency relation is that a given individual contributes a single event to the stream, and we want to preserve *event-level privacy* [8], that is, hide to some extent the presence or absence of an event at a particular time. Even though individual events should be hidden, we are still interested in producing approximate filtered versions of the original signal, e.g., a privacy-preserving moving average of the input tracking the frequency of events. The papers [8], [9] consider specifically the design of a private counter or accumulator, i.e., a system producing an output signal

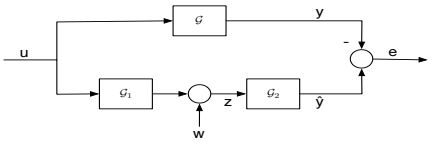

Fig. 1. Differentially private filter approximation set-up.

y with $y_t = y_{t-1} + u_t$, where u is binary valued. Note that this system is unstable. A number of other filters with slowly and monotonically decreasing impulse responses are considered in [14], using a technique similar to [9] based on binary trees. Here we show certain approximations of a general linear stable filter $\mathcal{G}$ that preserve event-level privacy. We first make the following remark.

Lemma 6: Let $\mathcal{G}$ be a single-input single-output linear system with impulse response g. Then for the adjacency relation (6) on integer-valued input signals, the ℓ_p sensitivity of $\mathcal{G}$ is $\Delta_p\mathcal{G} = \|g\|_p$. In particular for $p = 2$, we have $\Delta_2\mathcal{G} = \|\mathcal{G}\|_2$, the $\mathcal{H}_2$ norm of $\mathcal{G}$.

Proof: For two adjacent binary-valued signals u, u', we have that $u-u'$ is a positive or negative impulse signal δ, and hence $\|\mathcal{G}u - \mathcal{G}u'\|_p = \|\mathcal{G}(u-u')\|_p = \|\mathcal{G}\delta\|_p = \|g * \delta\|_p = \|g\|_p$. ∎

We measure the utility of specific schemes throughout this section by the MSE between the published and desired outputs. Similarly to our discussion in Section III-C, there are two straightforward mechanisms that provide differential privacy. One can add white noise w directly to the input signal, with $w_t \sim \text{Lap}(1/\epsilon)$ for the Laplace mechanism and $w_t \sim \mathcal{N}(0, \kappa(\delta,\epsilon))$ for the Gaussian mechanism. Or one can add noise at the output of the filter $\mathcal{G}$, with $w_t \sim \text{Lap}(\|g\|_1/\epsilon)$ for the Laplace mechanism and $w_t \sim \mathcal{N}(0, \|g\|_2\kappa(\delta,\epsilon))$ for the Gaussian mechanism. For the Gaussian mechanism, one obtains in both cases an MSE equal to $\|\mathcal{G}\|_2^2\ \kappa(\delta,\epsilon)^2$. For the Laplace mechanism, it is always better to add the noise at the input. Indeed, we obtain in this case an MSE of $2\|g\|_2^2/\epsilon^2$ instead of the greater $2\|g\|_1^2/\epsilon^2$ if the noise is added at the output.

We now generalize these mechanisms to the approximation set-up shown on Fig. 1. The previous mechanisms are recovered when $\mathcal{G}_1$ or $\mathcal{G}_2$ is the identity operator. To show that one can improve the utility of the mechanism with this set-up, consider the following choice of filters $\mathcal{G}_1$ and $\mathcal{G}_2$. Let $\mathcal{G}_1$ be a stable, minimum phase filter (hence invertible). Let $\mathcal{G}_2 = \mathcal{G}\mathcal{G}_1^{-1}$. We call this particular choice the *zero forcing equalization* (ZFE) mechanism. To guarantee (ϵ,δ)-differential privacy, the noise w is chosen to be white Gaussian with $\sigma = \kappa(\delta,\epsilon)\|\mathcal{G}_1\|_2$. The MSE for the ZFE mechanism is

$$e_{mse}^{ZFE} := \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{\infty}\mathbb{E}[\|(\mathcal{G}u)_t - (\mathcal{G}u + \mathcal{G}\mathcal{G}_1^{-1}w)_t\|_2^2] = \kappa(\epsilon,\delta)^2\|\mathcal{G}_1\|_2^2\|\mathcal{G}\mathcal{G}_1^{-1}\|_2^2.$$

Hence we are led to consider the following problem

$$\min_{\mathcal{G}_1}\frac{1}{4\pi^2}\int_{-\pi}^{\pi}|\mathcal{G}_1(e^{j\omega})|^2 d\omega\int_{-\pi}^{\pi}\left|\frac{\mathcal{G}(e^{j\omega})}{\mathcal{G}_1(e^{j\omega})}\right|^2 d\omega,$$

where the minimization is over the stable, minimum phase transfer functions $\mathcal{G}_1$.

Theorem 7: We have, for any stable, minimum phase system $\mathcal{G}_1$, $e_{mse}^{ZFE} \geq \kappa(\epsilon,\delta)^2\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}|\mathcal{G}(e^{j\omega})|d\omega\right)^2$. This lower bound on the mean-squared error of the ZFE mechanism is attained by letting $|\mathcal{G}_1(e^{j\omega})|^2 = \lambda|\mathcal{G}(e^{j\omega})|$ for all $\omega \in [-\pi,\pi)$, where λ is some arbitrary positive number. It can be approached arbitrarily closely by stable, rational, minimum phase transfer functions $\mathcal{G}_1$.

Proof: By the Cauchy-Schwarz inequality, we have

$$\left(\int_{-\pi}^{\pi}|\mathcal{G}(e^{j\omega})|d\omega\right)^2 = \left(\int_{-\pi}^{\pi}|\mathcal{G}_1(e^{j\omega})|\left|\frac{\mathcal{G}(e^{j\omega})}{\mathcal{G}_1(e^{j\omega})}\right|d\omega\right)^2 \leq \int_{-\pi}^{\pi}|\mathcal{G}_1(e^{j\omega})|^2 d\omega\int_{-\pi}^{\pi}\left|\frac{\mathcal{G}(e^{j\omega})}{\mathcal{G}_1(e^{j\omega})}\right|^2 d\omega,$$

hence the bound. Moreover, equality is attained if and only if there exists $\lambda \in \mathbb{R}$ such that $|\mathcal{G}_1(e^{j\omega})| = \lambda\left|\frac{\mathcal{G}(e^{j\omega})}{\mathcal{G}_1(e^{j\omega})}\right|$, i.e., $|\mathcal{G}_1(e^{j\omega})|^2 = \lambda|\mathcal{G}(e^{j\omega})|, \quad \forall\omega \in \mathbb{R}$. To see that the bound can be approached using finite-dimensional filters, by Weierstrass theorem we can first approximate $|\mathcal{G}(e^{j\omega})|$ arbitrarily closely by a rational positive function $\hat{\mathcal{G}}$. We then set $\mathcal{G}_1$ to be the minimum-phase spectral factor of $\hat{\mathcal{G}}$. ∎

The MSE obtained for the best ZFE mechanism in Theorem 7 cannot be worse than the MSE for the scheme adding noise at the input, and is generally strictly smaller, since by Jensen's inequality we have

$$\left(\int_{-\pi}^{\pi}|\mathcal{G}(e^{j\omega})|\frac{d\omega}{2\pi}\right)^2 \leq \int_{-\pi}^{\pi}|\mathcal{G}(e^{j\omega})|^2\frac{d\omega}{2\pi} = \|\mathcal{G}\|_2^2.$$

In addition, the MSE of the ZFE mechanism is independent of the input signal u. However, a smaller error could be obtained with other schemes, in particular schemes that exploit some knowledge about the input signal. Note that once $\mathcal{G}_1$ is chosen, designing $\mathcal{G}_2$ is a standard equalization problem [15]. The name of the ZFE mechanism is motivated by the choice of trying to cancel the effect of $\mathcal{G}_1$ by using its inverse (zero forcing equalizer). Nonlinear components can be very useful as well. In particular if we add the hypothesis

that the input signal is binary valued, as in [8], [9], we can modify the simple scheme adding noise at the input by including a detector H in front of the system $\mathcal{G}$, namely, for $\hat{u}_t = u_t + w_t$, $H\hat{u}_t = 1$ if $\hat{u}_t \geq 1/2$ and 0 if $\hat{u}_t < 1/2$. This exploits the knowledge that the input signal is binary valued, preserves differential privacy by Theorem 1, and sometimes significantly improves the MSE, depending on other characteristics of the signal.

A. Exploiting Additional Public Knowledge

To further illustrate the idea of exploiting potentially available additional knowledge about the input signal, assume now that it is publicly known that u is wide-sense stationary with known mean and autocorrelation. Then one should design a minimum mean squared error (MMSE) estimator for $\mathcal{G}_2$ rather than employing $\mathcal{G}\mathcal{G}_1^{-1}$, since the latter can significantly amplify the noise at frequencies where $\mathcal{G}_1$ is small [15]. We could still choose $\mathcal{G}_1$ according to Theorem 7, although now this choice is not optimal any more if $\mathcal{G}_2$ is not $\mathcal{G}\mathcal{G}_1^{-1}$. According to Theorem 1, differential privacy is preserved since the filter $\mathcal{G}_2$ only processes the already differentially private signal z. Even if the statistical assumptions turn out not to be satisfied by u, the privacy guarantee still holds and only performance is impacted.

B. Related Work

Some papers closely related to the event filtering problem considered in this section are [8], [9], [14], [16]. As previously mentioned, [8], [9] consider an unstable filter, the accumulator. The techniques employed there are quite different, relying essentially on binary trees to keep track of intermediate calculations and reduce the amount of noise introduced by the privacy mechanism. Bolot et al. [14] extend this technique to the differentially private approximation of certain filters with monotonic, slowly decaying impulse response. In fact, this technique can be extended to general linear systems by using a state-space realization and keeping track of the system state at carefully chosen times in a binary tree. However, the usefulness of this approach seems to be limited for most practical stable filters, the resulting MSE being typically too large and the implementation of the scheme significantly more complex than for a simple recursive filter.

V. Conclusion

We have discussed mechanisms for preserving the differential privacy of individual users transmitting time-varying signals to a trusted central server releasing sanitized filtered outputs based on these inputs. Decentralized versions of the mechanism of Section III can in fact be implemented in the absence of trusted server by means of cryptographic techniques [16]. We believe that research on privacy issues is critical to encourage the development of future cyber-physical systems, which typically rely on the participants data to improve their efficiency. Numerous directions of study are open for dynamical systems, including designing better filtering mechanisms, and understanding design trade-offs between privacy or security and performance in large-scale control systems.

References

[1] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, December 1992.

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

[3] S. P. Kasiviswanathan and A. Smith, "A note on differential privacy: Defining resistance to arbitrary side information," March 2008. [Online]. Available: http://arxiv.org/abs/0803.3946

[4] C. Dwork, K. Kenthapadi, F. McSherry, I. M. M. Naor, and Naor, "Our data, ourselves: Privacy via distributed noise generation," *Advances in Cryptology-EUROCRYPT 2006*, pp. 486–503, 2006.

[5] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proceedings of the CCS Workshop on Privacy in the Electronic Society (WPES)*, Raleigh, North Carolina, October 2012, to appear.

[6] D. Varodayan and A. Khisti, "Smart meter privacy using a rechargeable battery: minimizing the rate of information leakage," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prag, Czech Republic, 2011.

[7] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "A theory of privacy and utility in databases," Princeton University, Tech. Rep., February 2011.

[8] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observations," in *STOC'10*, Cambridge, MA, June 2010.

[9] T.-H. H. Chan, E. Shi, and D. Song, "Private and continual release of statistics," *ACM Transactions on Information and System Security*, vol. 14, no. 3, pp. 26:1–26:24, November 2011.

[10] J. Le Ny and G. J. Pappas, "Differentially private filtering," July 2012. [Online]. Available: http://arxiv.org/abs/1207.4305

[11] C. Dwork, "Differential privacy," in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*, ser. Lecture Notes in Computer Science, vol. 4052. Springer-Verlag, 2006.

[12] L. Breiman, *Probability*, ser. Classics in Applied Mathematics. SIAM, 1992.

[13] A. van der Schaft, *L2-gain and passivity techniques in nonlinear control*. Springer Verlag, 2000.

[14] J. Bolot, N. Fawaz, S. Muthukrishnan, A. Nikolov, and N. Taft, "Private decayed sum estimation under continual observation," September 2011, http://arxiv.org/abs/1108.6123.

[15] J. Proakis, *Digital Communications*. McGraw-Hill, 2000.

[16] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proceedings of the ACM Conference on Management of Data (SIGMOD)*, Indianapolis, IN, June 2010.