Differentially Private Filtering

Jerome Le Ny, *Member, IEEE*, and George J. Pappas, *Fellow, IEEE*

***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, this paper introduces privacy concerns in a system theoretic context, and addresses the problem of releasing filtered signals that respect the privacy of the user data streams. Our approach relies on a formal notion of privacy from the database literature, called *differential privacy*, which provides strong privacy guarantees against adversaries with arbitrary side information. Methods are developed 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. First, the notion of differential privacy is extended to dynamic systems with many participants contributing independent input signals. Kalman filtering is also discussed in this context, when a released output signal must preserve differential privacy for the measured signals or state trajectories of the individual participants. Second, differentially private mechanisms are described to approximate stable filters when participants contribute to a single event stream, extending previous work on differential privacy under continual observation.**

***Index Terms*—Estimation, filtering, Kalman filtering, privacy.**

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, population health monitoring, online recommendation systems, traffic monitoring, fuel consumption optimization, and cloud computing for industrial control systems. For privacy, confidentiality 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, thereby 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 [3]. Similarly, individual private signals can be recovered from published outputs aggregated from many users, and anonymizing a dataset is not enough to guarantee privacy, due to the existence of public side information. This is demonstrated in [4], [5] for example, where private ratings and transactions from individuals on commercial websites are successfully inferred with the help of information from public recommendation systems. Emerging traffic monitoring systems using position measurements from smartphones [6] is another application area where individual position traces can be re-identified by correlating them with public information such as a person's location of residence or work [6], [7]. Hence, the development of rigorous privacy preserving mechanisms is crucial to address the justified concerns of potential users and thus encourage an increasing level of participation, which can in turn greatly improve the efficiency of these large-scale systems.

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* [8], 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 an 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 [9].

Most work related to privacy is concerned with the analysis of static databases [8], [10]–[12], whereas cyber-physical systems clearly emphasize the need for mechanisms working with dynamic, time-varying data streams. Recently, the problem of releasing differentially private statistics when the input data takes the form of a binary stream describing event occurrences aggregated from many participants has been considered in [13]–[15]. This work forms the basis for the scenario studied in Section VI, and is discussed in more details in Section VI-C. However, most of this paper is devoted to a different situation where participants individually provide real-valued signals. A differentially private version of the iterative averaging algorithm for consensus is considered in [16]. In this case, the input data to protect consists of the initial values of the participants and is

Manuscript received September 04, 2012; revised April 16, 2013; accepted August 22, 2013. Date of publication September 23, 2013; date of current version January 21, 2014. This work was supported in part by the Natural Sciences and Engineering Research Council under Grant RGPIN-435905-13 and in part by TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA. Preliminary versions of this paper appeared in [1] and [2]. Recommended by Associate Editor L. Schenato.

J. Le Ny is with the Department of Electrical Engineering, Ecole Polytechnique de Montreal, Montreal, QC H3T 1J4, Canada (e-mail: jerome.le-ny@polymtl.ca).

G. J. Pappas is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: pappasg@seas.upenn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2013.2283096

thus a single vector, but the update mechanism subject to privacy attacks is dynamic. Information-theoretic approaches have also been proposed to guarantee some level of privacy when releasing time series [17], [18]. 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. This task is very difficult in general as new side information can become available after releasing the results. In contrast, differential privacy is a worst-case notion that holds independently of any probabilistic assumption on the dataset, and provides guarantees (different from those of [17], [18]) against adversaries with arbitrary side information [9]. Once such a privacy guarantee is enforced, one can still leverage potential additional statistical information about the dataset to improve the quality of the outputs.

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 for 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. These results are extended in Section IV to the problem of designing a differentially private Kalman filter, as an example of situation where additional information about the process generating the individual signals can be leveraged to publish more accurate results. Finally, Section VI is motivated by the recent work on "differential privacy under continual observation" [13], [14], and considers systems processing a single integer-valued signal describing the occurrence of events originating from many individual participants. Throughout the paper, differentially private approximations of the systems are proposed with the goal of minimizing the mean squared error (MSE) introduced by the privacy preserving mechanisms.

II. Differential Privacy

In this section, we review the notion of differential privacy [8] as well as some basic mechanisms that can be used to achieve it when the released data belongs to a finite-dimensional vector space. In the original papers on differential privacy [8], [10], [19], a sanitizing mechanism has access to a database and provides noisy answers to queries submitted by data analysts wishing to draw inference from the data. However, the notion of differential privacy can be defined for fairly general types of datasets. 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., [20], for additional background on differential privacy.

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 data tables, or a signal space). A *mechanism* is just a map $M : \mathsf{D} \times \Omega \rightarrow \mathsf{R}$, for some measurable output space $(\mathsf{R}, \mathcal{M})$, where $\mathcal{M}$ denotes a σ-algebra, such that for any element $d \in \mathsf{D}$, $M(d, \cdot)$ is a random variable, typically written simply $M(d)$. A mechanism can be viewed as a probabilistic algorithm to answer a query q, which is a map $q : \mathsf{D} \rightarrow \mathsf{R}$. In some cases, we index the mechanism by the query q of interest, writing M_q.

1) Example 1: Let $\mathsf{D} = \mathbb{R}^n$, with each real-valued 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., the query is $q : \mathsf{D} \rightarrow \mathbb{R}$ with $q(d) = \frac{1}{n} \sum_{i=1}^n d_i$. 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 $Y : \Omega \rightarrow \mathbb{R}$, so that $M_q : \mathsf{D} \times \Omega \rightarrow \mathbb{R}$ with $M_q(d) = \frac{1}{n} \sum_{i=1}^n d_i + Y$. Note that in the absence of perturbation Y, an adversary who knows n and all d_j for $j \geq 2$ can recover the remaining entry d_1 exactly if he learns $q(d)$. This can deter people from contributing their data, even though broader participation improves the accuracy of the analysis and thus can provide useful knowledge to the population.

2) Example 2: A database could record, for n participants, a k-tuple of binary values for the presence or absence of k attributes, e.g., being less that 50 years old. For statistical analysis purposes, it is typically sufficient to consider this dataset as a vector $d \in \mathbb{N}^{2^k}$, where d_i represents the number of occurrences of the k-tuple i. While this histogram representation removes certain obvious forms of identification, e.g., names, publishing it still carries considerable privacy risks due to the possibility of linking its information to other public datasets, in order to re-identify individuals for example [21]. Small entries in the vector d, corresponding to attributes characterizing few peoples, are particularly sensitive [22]. An alternative to publishing d or a perturbed version of it, is to keep d secure and force analysts to directly send their query q to the database server, which can then provide an approximate answer and better control the leakage of private information [8], [23]. For example, a linear query of the form $q(d) = Md$ for $M \in \{0,1\}^{p \times 2^k}$ can be used to request simultaneously p marginals, partial histograms, etc. Much work has been done on answering such statistical queries in a differentially private way, see, e.g., [12] and [19].

3) Example 3: We consider in Section V a road traffic monitoring system, where individuals send their location to a data aggregator at discrete time intervals, and receive a dynamic estimate of the average traffic velocity on a road segment of interest. For n users, the dataset consists of n discrete-time position signals $(x_1, \ldots, x_n)$, with $x_{i,t} \in \mathbb{R}^2, t \geq 0$, and a query is also a signal $q_t = \frac{1}{\sum_{i=1}^n 1_{\{x_{i,t} \in RS\}}} \sum_{i=1}^n \dot{x}_{i,t} 1_{\{x_{i,t} \in RS\}}$, where $1_{\{\cdot\}}$ denotes the indicator function and RS the road segment.

Next, we introduce the notion of differential privacy [8], [10]. In the following definition, we have a symmetric binary relation Adj on a space of datasets D, called adjacency. Intuitively $\mathrm{Adj}(d, d')$ 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 $M : \mathsf{D} \times \Omega \rightarrow \mathsf{R}$ is

(ϵ, δ)-differentially private for Adj if for all $d, d' \in \mathsf{D}$ such that $\mathrm{Adj}(d, d')$, we have

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

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

In words, 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).

Remark 1: The definition of differential privacy depends on the adjacency relation Adj. However, we often omit to mention it when Adj is clear from the context. Definition 1 also depends on the choice of σ-algebra $\mathcal{M}$. When we need to state this σ-algebra explicitly, we write $M : \mathsf{D} \times \Omega \to (\mathsf{R}, \mathcal{M})$. In particular, this σ-algebra should be sufficiently large, since (1) is trivially satisfied by any mechanism if $\mathcal{M} = \{\emptyset, \mathsf{R}\}$.

The next lemma provides alternative technical characterizations of differential privacy and appears to be partially new. First, we introduce some notation. We call a signed measure ν on $(\mathsf{R}, \mathcal{M})$ δ-bounded if it satisfies $\nu(S) \leq \delta$ for all $S \in \mathcal{M}$. A measure is sometimes called positive measure for emphasis. For $(\mathsf{R}, \mathcal{M})$ a measurable space, we denote by $\mathcal{F}_b(\mathsf{R})$ the space of bounded real-valued measurable functions on R and we define $\mu g := \int g \, d\mu$ for $g \in \mathcal{F}_b(\mathsf{R})$ and μ a positive measure on $\mathcal{M}$.

Lemma 1: Let $\epsilon, \delta \geq 0$, and $M : \mathsf{D} \times \Omega \to (\mathsf{R}, \mathcal{M})$ be a mechanism, where D is a space equipped with an adjacency relation Adj. *The following are equivalent:*

(a) *M is (ϵ, δ)-differentially private, satisfying (1).*

(b) *For all d, d' in D such that* $\mathrm{Adj}(d, d')$, *there exists a δ-bounded positive measure $\mu^{d,d'}$ on $(\mathsf{R}, \mathcal{M})$ such that we have, for all S in $\mathcal{M}$*

$$\mathbb{P}(M(d) \in S) \leq e^{\epsilon} \mathbb{P}(M(d') \in S) + \mu^{d,d'}(S). \tag{2}$$

(c) *For all d, d' in D such that* $\mathrm{Adj}(d, d')$, *there exists a δ-bounded positive measure $\mu^{d,d'}$ on $(\mathsf{R}, \mathcal{M})$ such that for all $g \in \mathcal{F}_b(\mathsf{R})$, we have*

$$\mathbb{E}(g(M(d))) \leq e^{\epsilon} \mathbb{E}(g(M(d'))) + \mu^{d,d'} g. \tag{3}$$

Proof:

$(a) \Rightarrow (b)$: Suppose that M is (ϵ, δ)-differentially private. Define the signed measure $\nu^{d,d'}$ by $S \mapsto \nu^{d,d'}(S) := \mathbb{P}(M(d) \in S) - e^{\epsilon} \mathbb{P}(M(d') \in S)$. By the definition (1), $\nu^{d,d'}$ is δ-bounded. Let $\mu^{d,d'}$ be the positive variation of $\nu^{d,d'}$, i.e., $\mu^{d,d'}(S) = \sup\{\nu^{d,d'}(G) : G \in \mathcal{M}, G \subset S\}$, for all $S \in \mathcal{M}$. Then $\mu^{d,d'}$ is a positive measure [24, Th. 5.6.1], and is δ-bounded since $\nu^{d,d'}$ is. Since $\nu^{d,d'}(S) \leq \mu^{d,d'}(S)$ for all $S \in \mathcal{M}$, we have (2).

$(b) \Rightarrow (c)$: Let B be a bound on $|g|$. For any $k \geq 1$, we divide the interval $[-B, B]$ into k consecutive intervals I_i of length $2B/k$, and we let $A_i = g^{-1}(I_i)$ and c_i be the midpoint of the interval I_i. Then (c) holds for the simple function $\sum_{i=1}^{k} c_i 1_{A_i}$, and these functions approximate g. We conclude using the dominated convergence theorem.

$(c) \Rightarrow (a)$: Take $g = 1_S$ and use the fact that $\mu^{d,d'}$ is δ-bounded.

Finally, we mention that for the special case $\delta = 0$, the implication $(a) \Rightarrow (c)$ is shown in [25]. ■

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. This result is similar in spirit to the data processing inequality from information theory [26]. 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 Postprocessing): Let $\epsilon, \delta \geq 0$. 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, for all $S \in \mathcal{M}_2$ and $d \in \mathsf{D}$

$$\mathbb{P}(M_2(d) \in S | M_1(d)) = k(M_1(d), S), \text{ a.s.}. \tag{4}$$

Then M_2 is (ϵ, δ)-differentially private.

Note that in (4), 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 says 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, postprocessing can be used freely to improve the *accuracy* of an output, as in Section VI for example, without worrying about a possible loss of privacy. Similarly, an adversary processing a differentially private output without accessing the original data cannot weaken the guarantee.

Proof: To the best of our knowledge, there is no previous proof of the resilience to postprocessing theorem available for the case of randomized postprocessing and $\delta > 0$. Let M_1 be (ϵ, δ)-differentially private. We have, for two adjacent elements $d, d' \in \mathsf{D}$ and for any S in $\mathcal{M}_2$

$$\begin{aligned}\mathbb{P}(M_2(d) \in S) &= \mathbb{E}[\mathbb{P}(M_2 \in S | M_1(d))] = \mathbb{E}[k(M_1(d), S)] \\ &\leq e^{\epsilon} \mathbb{E}[k(M_1(d'), S)] + \int_{\mathsf{R}_1} k(m_1, S) \, d\mu_1^{d,d'}(m_1) \\ &= e^{\epsilon} \mathbb{P}(M_2(d') \in S) + \mu_2^{d,d'}(S).\end{aligned}$$

The first equality is just the smoothing property of conditional expectations, and the inequality comes from (3) applied to the function $k(\cdot, S)$. Since k is a probability kernel, the integral on the second line defines a measure $\mu_2^{d,d'}$ on R_2, which is δ-bounded since $\mu_2^{d,d'}(\mathsf{R}_2) = \int_{\mathsf{R}_1} k(m_1, \mathsf{R}_2) d\mu_1^{d,d'}(m_1) = \int_{\mathsf{R}_1} 1 \, d\mu_1^{d,d'}(m_1) = \mu_1^{d,d'}(\mathsf{R}_1) \leq \delta$. ■

A. 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 [8].

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 $\Delta_{\mathsf{R}} q := \sup_{d,d':\mathrm{Adj}(d,d')} \|q(d) - q(d')\|_{\mathsf{R}}$. 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) Laplace Mechanism: This mechanism, proposed in [8], modifies an answer to a numerical query by adding independent and identically distributed (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 i.i.d. and $w_i \sim \mathrm{Lap}(b)$, denoted $w \sim \mathrm{Lap}(b)^k$, we have $p(w; b) = (\frac{1}{2b})^k \exp\left(-\frac{\|w\|_1}{b}\right)$, $\mathbb{E}[\|w\|_1] = b$, and $\mathbb{P}(\|w\|_1 \geq tb) = e^{-t}$.

Theorem 2: Let $q : \mathsf{D} \to \mathbb{R}^k$ be a query, and $\epsilon > 0$. 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.

Proof: We have, for $S \subset \mathbb{R}^k$ measurable and d, d' two adjacent datasets in D,

$$\begin{aligned} P(M_q(d) \in S) &= \left(\frac{1}{2b}\right)^k \int_{\mathbb{R}^k} 1_S(q(d) + w) e^{-\frac{\|w\|_1}{b}} dw \\ &= \left(\frac{1}{2b}\right)^k \int_{\mathbb{R}^k} 1_S(u) e^{-\frac{\|u-q(d)\|_1}{b}} dw \\ \leq e^{\frac{\|q(d)-q(d')\|_1}{b}} &\left(\frac{1}{2b}\right)^k \int_{\mathbb{R}^k} 1_S(u) e^{-\frac{\|u-q(d')\|_1}{b}} dw \end{aligned}$$

since $-\|u - q(d)\|_1 \leq -\|u - q(d')\|_1 + \|q(d) - q(d')\|_1$ by the triangle inequality. With the choice of $b = \Delta_1 q/\epsilon$, we obtain the definition (1) of differential privacy (i.e., with $\delta = 0$). ■

2) Gaussian Mechanism: This mechanism, proposed in [10], is similar to the Laplace mechanism but adds i.i.d. Gaussian noise to provide (ϵ, δ)-differential privacy, with $\delta > 0$ but typically a smaller ϵ for the same utility. First, 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 [10].

Theorem 3: Let $q : \mathsf{D} \to \mathbb{R}^k$ be a query, and $\epsilon > 0, \frac{1}{2} > \delta > 0$. 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.

Proof: Let d, d' be two adjacent elements in D, and denote $v := q(d) - q(d')$. We use the notation $\|\cdot\|$ for the 2-norm in this proof. For $S \in \mathcal{R}^k$, we have

$$\begin{aligned} \mathbb{P}(M_q(d) \in S) &= \frac{1}{(2\pi\sigma^2)^{k/2}} \int_{\mathbb{R}^k} 1_S(q(d) + w) e^{-\frac{\|w\|^2}{2\sigma^2}} dw \\ &= \frac{1}{(2\pi\sigma^2)^{k/2}} \int_{\mathbb{R}^k} 1_S(u) e^{-\frac{\|u-q(d)\|^2}{2\sigma^2}} du \\ &= \frac{1}{(2\pi\sigma^2)^{k/2}} \int_S e^{-\frac{\|u-q(d')\|^2}{2\sigma^2}} e^{\frac{2(u-q(d'))^T v - \|v\|^2}{2\sigma^2}} du \\ &\leq e^\epsilon \mathbb{P}(M_q(d') \in S) + \frac{1}{(2\pi\sigma^2)^{k/2}} \int_S \Bigg[e^{-\frac{\|u-q(d)\|^2}{2\sigma^2}} \\ &\qquad 1_{\{2(u-q(d'))^T v \geq \|v\|^2 + 2\epsilon\sigma^2\}} \Bigg] du. \end{aligned}$$

The last integral term defines a measure $S \mapsto \mu^{d,d'}(S)$ on $\mathbb{R}^k$ that we wish to bound by δ. With the change of variables $y = (u - q(d))/\sigma$ and the choice $S = \mathbb{R}^k$ in the integral, we can rewrite it as $\mathbb{P}(Y^T v \geq \epsilon\sigma - \|v\|^2/2\sigma)$, with $Y \sim \mathcal{N}(0, I_k)$. In particular, $Y^T v \sim \mathcal{N}(0, \|v\|^2)$, hence is equal to $\|v\| Z$ in distribution, with $Z \sim \mathcal{N}(0, 1)$. We are then led to set σ sufficiently large so that $\mathbb{P}(Z \geq \epsilon\sigma/\|v\| - \|v\|/2\sigma) \leq \delta$, i.e., $\mathcal{Q}(\epsilon\sigma/\|v\| - \|v\|/2\sigma) \leq \delta$. The result then follows by straightforward calculation. ■

As an illustration of the theorem, to guarantee (ϵ, δ)-differential privacy with $\epsilon = \ln 2$ and $\delta = 0.05$, the standard deviation of the Gaussian noise should be about 2.65 times the ℓ_2-sensitivity of q. 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

In this section, we introduce the notion of differential privacy for dynamic systems. We start with some notations and technical prerequisites. All signals are discrete-time signals, start at time 0, and all systems are assumed to be causal. "Linear and time-invariant" is abbreviated by LTI, and "single-input single-output" by SISO. For each time T, let P_T be the truncation operator, so that for any signal x we have

$$(P_T x)_t = \begin{cases} x_t, & t \leq T \\ 0, & t > T. \end{cases}$$

Hence, a deterministic system $\mathcal{G}$ is causal if and only if $P_T \mathcal{G} = P_T \mathcal{G} P_T$. We denote by ℓ_{pe}^m the space of sequences with values in $\mathbb{R}^m$ and such that $x \in \ell_{pe}^m$ 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 defined, respectively, as $\|\mathcal{G}\|_2 = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{Tr}(\mathcal{G}^*(e^{i\omega})\mathcal{G}(e^{i\omega})) d\omega\right)^{1/2}$, $\|\mathcal{G}\|_\infty = \operatorname{ess\,sup}_{\omega \in [-\pi,\pi)} \sigma_{\max}(\mathcal{G}(e^{i\omega}))$, where $\sigma_{\max}(A)$ denotes the maximum singular value of a matrix A.

We 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

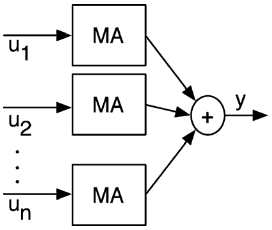

Fig. 1. Illustrative example of a system computing the sum of the moving averages (MA) of input signals contributed by n individual participants. A differentially private version of this system, for the adjacency relation (5), guarantees to user i that the distribution of the output signal does not vary significantly when her input varies in r_i-norm by at most b_i. In particular, the distribution of the output signal will not change significantly if user i's input is zero ($u_i \equiv 0$, e.g., because the user is not present), or is not zero but satisfies $\|u_i\|_{r_i} \leq b_i$.

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, see Fig. 1. For $r = (r_1, \ldots, r_n)$ and $m = (m_1, \ldots, m_n)$, an adjacency relation can be defined on $l^m_{re} := \ell^{m_1}_{r_1 e} \times \ldots \times \ell^{m_n}_{r_n e}$ 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

$$\begin{aligned} \mathrm{Adj}^b(u, u') \text{ iff for some } i, \|u_i - u_i'\|_{r_i} \leq b_i, \\ \text{and } u_j = u_j' \text{ for all } j \neq i. \end{aligned} \tag{5}$$

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 $M : \ell^m_{re} \times \Omega \to \ell^{m'}_{se}$, i.e., producing for any input signal $u \in \ell^m_{re}$ a stochastic process Mu with sample paths in $\ell^{m'}_{se}$. As in Section II, this requires that we first specify the measurable sets of $\ell^{m'}_{se}$. We start by defining in a standard way the measurable sets of $\ell^{m'} := (\mathbb{R}^{m'})^{\mathbb{N}}$, 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 $\{y \in \ell^{m'} : y_{0:T} \in H_T\}$, 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., [27, Ch. 2]). The measurable sets considered for the output of M are then obtained by intersection of $\ell^{m'}_{se}$ with the sets of $\mathcal{M}^{m'}$. The resulting σ-algebra is denoted $\mathcal{M}^{m'}_{se}$ and is generated by the sets of the form

$$\tilde{H}_T = \{y \in \ell^{m'}_{se} : y_{0:T} \in H_T\}, \text{ with } H_T \in \mathcal{R}^{(T+1)m'} \tag{6}$$

for $T \geq 0$. As for the dynamic systems of interest, we constrain in this paper the mechanisms to be causal, i.e., the distribution of $P_T M u$ should be the same as that of $P_T M P_T u$ for any $u \in \ell^m_{re}$ and any time T. In other words, the values u_t for $t > T$ do not influence the values of the mechanism output up to time T. 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 2: Consider an adjacency relation Adj *on* ℓ^m_{re}. *Let* $\epsilon, \delta \geq 0$. *For a mechanism* $M : \ell^m_{re} \times \Omega \to \ell^{m'}_{se}$, *the following are equivalent*

(a) *M is (ϵ, δ)-differentially private.*

(b) *For all u, u' in ℓ^m_{re} such that* $\mathrm{Adj}(u, u')$, *we have*

$$\begin{aligned} \mathbb{P}((Mu)_{0:T} \in A) \leq e^{\epsilon}\, \mathbb{P}((Mu')_{0:T} \in A) + \delta, \\ \forall T \geq 0, \forall A \in \mathcal{R}^{(T+1)m'}. \end{aligned} \tag{7}$$

Proof:

a) ⇒ b): If M is (ϵ, δ)-differentially private, then for u, u' adjacent, and for all $H \in \mathcal{M}^{m'}_{se}$, we have $\mathbb{P}(Mu \in H) \leq e^{\epsilon}\, \mathbb{P}(Mu' \in H) + \delta$. In particular, for a given integer $T \geq 0$, we can restrict our attention to the sets $\tilde{H}_T$ of the form (6). In this case, we have immediately $\mathbb{P}(Mu \in \tilde{H}_T) = \mathbb{P}((Mu)_{0:T} \in H_T)$ since the events are the same.

b) ⇒ a): Conversely, consider two adjacent signal $u, u' \in \ell^m_{re}$, and let $S \in \mathcal{M}^{m'}_{se}$, for which we want to show (1). Fix $\eta > 0$. There exists $T \geq 0$ and $H_T \in \mathcal{R}^{(T+1)m'}$ such that $\mathbb{P}(Mu \in S \Delta \tilde{H}_T) \leq \eta$ and $\mathbb{P}(Mu' \in S \Delta \tilde{H}_T) \leq \eta$, where $A \Delta B := (A \setminus B) \cup (B \setminus A)$ denotes the symmetric difference. This is a consequence for example of the fact that the finite-dimensional cylinder sets form an algebra and of the argument in the proof of [24, Th. 3.1.10]. We then have

$$\begin{aligned} \mathbb{P}(Mu \in S) &\leq \mathbb{P}(Mu \in \tilde{H}_T) + \eta = \mathbb{P}((Mu)_{0:T} \in H_T) + \eta \\ &\leq e^{\epsilon}\, \mathbb{P}((Mu')_{0:T} \in H_T) + \delta + \eta \\ &= e^{\epsilon}\, \mathbb{P}(Mu' \in \tilde{H}_T) + \delta + \eta \\ &\leq e^{\epsilon}\, \mathbb{P}(Mu' \in S) + \delta + \eta(1 + e^{\epsilon}). \end{aligned}$$

Since η can be taken arbitrarily small, the differential privacy definition (1) holds. ■

B. Basic Dynamic Mechanisms

Recall (see, e.g., [28]) that for a system $\mathcal{G}$ with inputs in ℓ^m_{re} and output in $\ell^{m'}_{se}$, its ℓ_r-to-ℓ_s incremental gain $\gamma_{r,s}(\mathcal{G})$ is defined as the smallest number γ such that

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

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

$$\mathcal{G}(u_1, \ldots, u_n) = \sum_{i=1}^{n} \mathcal{G}_i u_i, \tag{8}$$

where $\mathcal{G}_i : \ell^{m_i}_{r_i e} \to \ell^{m'}_{se}$, 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 4: Let $\mathcal{G}$ be defined as in (8) and consider the adjacency relation (5). Then the mechanism $Mu = \mathcal{G}u + w$, where w is a white noise with $w_t \sim \mathrm{Lap}(B/\epsilon)^{m'}$, for $\epsilon > 0$ and $B \geq \max_{1 \leq i \leq n}\{\gamma_{r_i,1}(\mathcal{G}_i)\, b_i\}$, is ϵ-differentially private. For $\epsilon, \delta > 0$, the mechanism is (ϵ, δ)-differentially private if $w_t \sim \mathcal{N}(0, \sigma^2 I_{m'})$, with $\sigma \geq \kappa(\delta, \epsilon) \max_{1 \leq i \leq n}\{\gamma_{r_i,2}(\mathcal{G}_i)\, b_i\}$.

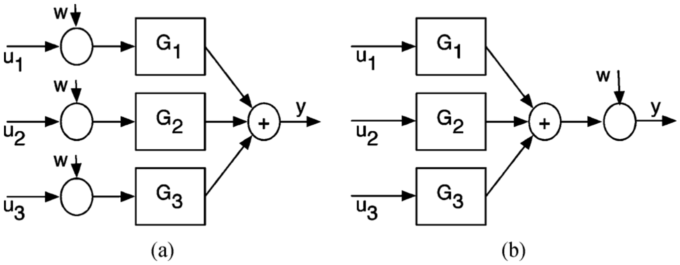

Fig. 2. Two architectures for differential privacy. (a) Input perturbation. (b) Output perturbation.

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

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

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 2, since (7) is satisfied for all T. ■

Corollary 1: Let $\mathcal{G}$ be defined as in (8), with $\mathcal{G}_i$ an LTI system, $r_i = 2$, and $\|\mathcal{G}_i\|_\infty < \infty$, for all $1 \le i \le n$. Let $\epsilon, \delta > 0$. 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 $\sigma \ge \kappa(\delta, \epsilon) \max_{1 \le i \le n}\{\|\mathcal{G}_i\|_\infty b_i\}$, is (ϵ, δ)-differentially private for (5).

C. Filter Approximation Set-Ups for Differential Privacy

Let $r_i = 2$ for all i, $\mathcal{G}$ be an LTI system as in 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 $\mathcal{G}$, depicted on Fig. 2. 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 $\kappa(\delta, \epsilon)^2 B \max_{1 \le i \le n}\{\|\mathcal{G}_i\|_\infty^2\}$. 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. For example, if n is small or if the bandwidths of the individual transfer functions $\mathcal{G}_i$ do not overlap, the error bound for the input perturbation scheme can be smaller. Another advantage of this scheme is that the users can release differentially private signals themselves without relying on a trusted server. However, there are cryptographic means for achieving the output perturbation scheme without centralized trusted server as well, see, e.g., [29].

1) Example 4: 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 $1 \le i \le n$. 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.

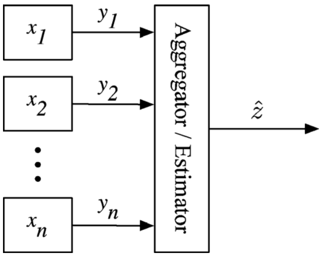

Fig. 3. Kalman filtering setup.

IV. Differentially Private Kalman Filtering

We now discuss the Kalman filtering problem subject to a differential privacy constraint. Compared to Section III, it is assumed here that more is publicly known about the dynamics of the processes producing the individual input signals, and this knowledge can be exploited in the design of privacy mechanisms with better performance. Section V describes an application of the mechanisms presented here to a traffic monitoring problem.

A. Differentially Private Kalman Filter

Consider a set of n linear systems, each with independent dynamics

$$x_{i,t+1} = A_i x_{i,t} + B_i w_{i,t}, \quad t \ge 0, \quad 1 \le i \le n \tag{9}$$

where w_i is a standard zero-mean Gaussian white noise process with covariance $\mathbb{E}[w_{i,t} w_{i,t'}] = \delta_{t-t'} I$, and the initial condition $x_{i,0}$ is a Gaussian random variable with mean $\bar{x}_{i,0}$, independent of the noise process w_i. System i, for $1 \le i \le n$, sends measurements

$$y_{i,t} = C_i x_{i,t} + D_i w_{i,t} \tag{10}$$

to a data aggregator. We assume for simplicity that the matrices D_i are full row rank. Fig. 3 shows this initial setup.

The data aggregator aims at releasing a signal that asymptotically minimizes the MSE with respect to a linear combination of the individual states. That is, the quantity of interest to be estimated at each period is $z_t = \sum_{i=1}^n L_i x_{i,t}$, where L_i are given matrices, and we are looking for a causal estimator $\hat{z}$ constructed from the signals $y_i, 1 \le i \le n$, solution of $\min_{\hat{z}} \lim_{T \to \infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left[\|z_t - \hat{z}_t\|_2^2\right]$. The data $\bar{x}_{i,0}, A_i, B_i, C_i, D_i, L_i, 1 \le i \le n$, are assumed to be public information. For all $1 \le i \le n$, we assume that the pairs (A_i, C_i) are detectable and the pairs (A_i, B_i) are stabilizable. In the absence of privacy constraint, the minimum mean squared error (MMSE) estimator is $\hat{z}_t = \sum_{i=1}^n L_i \hat{x}_{i,t}$, with $\hat{x}_{i,t}$ provided by the steady-state Kalman filter estimating the state of system i from y_i [30], and denoted $\mathcal{K}_i$ in the following.

Suppose now that the publicly released estimate should guarantee the differential privacy of the participants. This requires that we first specify an adjacency relation on the appropriate space of datasets. Let $x = [x_1^T, \dots, x_n^T]^T$ and $y = [y_1^T, \dots, y_n^T]^T$ denote the global state and measurement signals. Assume say that the mechanism is required to guarantee differential privacy for a subset $\mathcal{S}_i := \{i_1, \dots, i_k\}$ of the coordinates of the state trajectory x_i. Let the selection matrix S_i be the diagonal matrix with $[S_i]_{jj} = 1$ if $j \in \mathcal{S}_i$, and $[S_i]_{jj} = 0$ otherwise. Hence $S_i v$ sets the coordinates of a vector v that do not belong to the set $\mathcal{S}_i$ to zero. Fix a vector $\rho \in \mathbb{R}_+^n$. The adjacency relation considered here is

$$\begin{aligned}&\mathrm{Adj}_{\mathcal{S}}^{\rho}(x,x') \text{ iff for some } i,\ \|S_i x_i - S_i x_i'\|_2 \le \rho_i \qquad (11)\\ &(I - S_i)x_i = (I - S_i)x_i', \text{ and } x_j = x_j' \text{ for all } j \neq i.\end{aligned}$$

In words, two adjacent global state trajectories differ by the values of a single participant, say i. Moreover, for differential privacy guarantees we are constraining the range in energy variation in the signal $S_i x_i$ of participant i to be at most ρ_i^2. Hence, the distribution on the released results should be essentially the same if a participant's state signal value $S_i x_{i,t_0}$ at some single specific time t_0 were replaced by $S_i x'_{i,t_0}$ with $\|S_i(x_{i,t_0} - x'_{i,t_0})\| \le \rho_i$, but the privacy guarantee should also hold for smaller instantaneous deviations on longer segments of trajectory.

In the language of the previous sections, the dataset consists here of the state trajectories $x = (x_1, \dots, x_n)$, we have the adjacency relation (11) defined on this space, and the query of interest (i.e., in the absence of privacy constraint) is the signal $\hat{z} = \sum_{i=1}^n L_i \mathcal{K}_i y_i$ obtained via Kalman filtering. Note however the slight variation due to the additional constraint: here the data aggregator itself does not have direct access to the data x on which the adjacency relation is defined but only to the measurements y. Other adjacency relations could be considered, in particular, these could be defined directly on the output space of measured signals y. Finally, we aim at designing a mechanism M producing a signal My approaching as closely as possible the MMSE estimate $\hat{z}$, while maintaining differential privacy for (11).

First, in the rest of this subsection, we follow the approach of Section III-C, which simply consists in perturbing the original Kalman filter by adding privacy-preserving white Gaussian noise to certain signals. For the input noise injection mechanism, the noise can be added by each participant directly to their transmitted signal y_i. Namely, since for two state trajectories x_i, x_i' adjacent according to (11) we have $x_i - x_i' = S_i(x_i - x_i')$, the variation for the corresponding measured signals can be bounded as follows:

$$\begin{aligned}\|y_i - y_i'\|_2 = \|C_i S_i (x_i - x_i')\|_2 &= \|C_i S_i S_i (x_i - x_i')\|_2\\ &\le \sigma_{\max}(C_i S_i)\rho_i.\end{aligned}$$

Hence, differential privacy can be guaranteed if participant i adds to y_i a white Gaussian noise with covariance matrix $\kappa(\delta,\epsilon)^2 \rho_i^2 \sigma_{\max}^2(C_i S_i) I_{p_i}$, where p_i is the dimension of $y_{i,t}$. Note that in this sensitivity computation the measurement noise $D_i w_i$ has the same realization independently of the considered variation in x_i. With this perturbation, the signals transmitted by the participants are already differentially private. Moreover, in contrast to the basic input perturbation mechanism of Section III-C, the data aggregator here can modify the original Kalman filters $\mathcal{K}_i$ by viewing the privacy-preserving noise as additional measurement noise and improve the asymptotic estimation performance, see Section V.

Next, consider the output noise injection mechanism. Since we assume that $\bar{x}_0^i$ is public information, the initial condition $\hat{x}_{i,0}$ of each state estimator is fixed. Consider now two state trajectories x, x', adjacent according to (11), and let $\hat{z}, \hat{z}'$ be the corresponding estimates produced by the Kalman filters. We have

$$\begin{aligned}\hat{z} - \hat{z}' &= L_i \mathcal{K}_i (y_i - y_i') = L_i \mathcal{K}_i C_i S_i (x_i - x_i')\\ &= L_i \mathcal{K}_i C_i S_i S_i (x_i - x_i').\end{aligned}$$

Hence, $\|\hat{z} - \hat{z}'\|_2 \le \gamma_i \rho_i$, where γ_i is the $\mathcal{H}_\infty$ norm of the transfer function $L_i \mathcal{K}_i C_i S_i$. Thus we have the following theorem.

Theorem 5: Let $\epsilon, \delta > 0$. A mechanism releasing $(\sum_{i=1}^n L_i \mathcal{K}_i y_i) + \gamma\, \kappa(\delta,\epsilon)\, \nu$, where ν is a standard white Gaussian noise independent of $\{w_i\}_{1 \le i \le n}, \{x_{i,0}\}_{1 \le i \le n}$, and $\gamma = \max_{1 \le i \le n}\{\gamma_i \rho_i\}$, with γ_i the $\mathcal{H}_\infty$ norm of $L_i \mathcal{K}_i C_i S_i$, is (ϵ,δ)-differentially private for the adjacency relation (11).

B. Filter Redesign for Stable Systems

In the rest of Section IV, we aim at improving the output perturbation mechanism of Theorem 5, by redesigning the filter to optimize the overall MSE performance. This MSE is controlled both by the dynamics of the filter as well as the amount of privacy-preserving noise introduced at the output, which is a function of the $\mathcal{H}_\infty$ norm of the filter. Hence, we pursue the design of a filter that balances quality of estimation and size of its $\mathcal{H}_\infty$ norm.

We consider the design of n filters of the form

$$\hat{x}_{i,t+1} = F_i \hat{x}_{i,t} + G_i y_{i,t} \qquad (12)$$

$$\hat{z}_{i,t} = H_i \hat{x}_{i,t} + K_i y_{i,t} \qquad (13)$$

for $1 \le i \le n$, where F_i, G_i, H_i, K_i are matrices to determine. The overall system in front of the privacy-preserving noise source produces the estimate $\hat{z}_t = \sum_{i=1}^n \hat{z}_{i,t}$ of the signal z. Assume first in this section that the system matrices A_i are stable, in which case we also restrict the filter matrices F_i to be stable. Moreover, we only consider the design of full order filters, i.e., the dimensions of F_i are greater or equal to those of A_i, for all $1 \le i \le n$.

Denote the overall state for each system and associated filter by $\tilde{x}_i = [x_i^T, \hat{x}_i^T]^T$. The combined dynamics from w_i to the estimation error $e_i := z_i - \hat{z}_i$ can be written

$$\begin{aligned}\tilde{x}_{i,t+1} &= \tilde{A}_i \tilde{x}_{i,t} + \tilde{B}_i w_{i,t}\\ e_{i,t} &= \tilde{C}_i \tilde{x}_{i,t} + \tilde{D}_i w_{i,t}\end{aligned}$$

where

$$\tilde{A}_i = \begin{bmatrix} A_i & 0 \\ G_i C_i & F_i \end{bmatrix}, \quad \tilde{B}_i = \begin{bmatrix} B_i \\ G_i D_i \end{bmatrix},$$

$$\tilde{C}_i = [\, L_i - K_i C_i \quad -H_i \,], \quad \tilde{D}_i = -K_i D_i.$$

The steady-state MSE for the i^{th} estimator is then $\lim_{t \to \infty} \mathbb{E}[e_{i,t}^T e_{i,t}]$. Let us now consider the additional impact of the privacy-preserving noise on the overall MSE.

Considering as in the previous subsection the sensitivity of filter i's output to a change from a state trajectory x to an adjacent one x' according to (11), and letting $\delta x_i = x_i - x'_i = S_i(x_i - x'_i) = S_i \delta x_i$, we see that the change in the output of filter i follows the dynamics:

$$\begin{aligned}\delta \hat{x}_{i,t+1} &= F_i \delta \hat{x}_{i,t} + G_i C_i S_i \delta x_i \\ \delta \hat{z}_i &= H_i \delta \hat{x}_{i,t} + K_i C_i S_i \delta x_i.\end{aligned}$$

Hence, the ℓ_2-sensitivity can be measured by the $\mathcal{H}_\infty$ norm of the transfer function

$$\left[\begin{array}{c|c} F_i & G_i C_i S_i \\ \hline H_i & K_i C_i S_i \end{array}\right]. \tag{14}$$

Replacing the Kalman filter in Theorem 5, the MSE for the resulting output perturbation mechanism guaranteeing (ϵ, δ)-privacy is then

$$\left(\sum_{i=1}^{n} \|\tilde{C}_i(zI - \tilde{A}_i)^{-1}\tilde{B}_i + \tilde{D}_i\|_2^2\right) + \kappa(\delta,\epsilon)^2 \max_{1 \le i \le n} \{\gamma_i^2 \rho_i^2\}$$

with $\gamma_i := \|H_i(zI - F_i)^{-1} G_i C_i S_i + K_i C_i S_i\|_\infty$. Thus, minimizing this MSE leads us to the following optimization problem:

$$\min_{\mu_i, \lambda, F_i, G_i, H_i, K_i} \quad \sum_{i=1}^{n} \mu_i + \kappa(\delta,\epsilon)^2 \lambda \tag{15}$$

$$\text{s.t.} \ \forall \ 1 \le i \le n, \|\tilde{C}_i(zI - \tilde{A}_i)^{-1}\tilde{B}_i + \tilde{D}_i\|_2^2 \le \mu_i, \tag{16}$$

$$\rho_i^2 \|H_i(zI - F_i)^{-1} G_i C_i S_i + K_i C_i S_i\|_\infty^2 \le \lambda. \tag{17}$$

Assume without loss of generality that $\rho_i > 0$ for all i, since the privacy constraint for the signal x_i vanishes if $\rho_i = 0$. The following theorem gives a convex sufficient condition in the form of linear matrix inequalities (LMIs) guaranteeing that a choice of filter matrices F_i, G_i, H_i, K_i satisfies the constraints (16)-(17).

Theorem 6: The constraints (16)-(17), for some $1 \le i \le n$, *are satisfied if there exists matrices* $W_i, Y_i, Z_i, \hat{F}_i, \hat{G}_i, \hat{H}_i, \hat{K}_i$ *such that* $\mathrm{Tr}(W_i) < \mu_i$

$$\begin{bmatrix} W_i & (L_i - \hat{K}_i C_i - \hat{H}_i) & (L_i - \hat{K}_i C_i) & -\hat{K}_i D_i \\ * & Z_i & Z_i & 0 \\ * & * & Y_i & 0 \\ * & * & * & I \end{bmatrix} \succ 0$$

$$\begin{bmatrix} Z_i & Z_i & Z_i A_i & Z_i A_i & Z_i B_i \\ * & Y_i & (M_i + \hat{F}_i) & M_i & (Y_i B_i + \hat{G}_i D_i) \\ * & * & Z_i & Z_i & 0 \\ * & * & * & Y_i & 0 \\ * & * & * & * & I \end{bmatrix} \succ 0$$

$$\text{and} \quad \begin{bmatrix} Z_i & Z_i & 0 & 0 & 0 & 0 \\ * & Y_i & 0 & \hat{F}_i & 0 & \hat{G}_i C_i S_i \\ * & * & \frac{\lambda}{\rho_i^2} I & \hat{H}_i & 0 & \hat{K}_i C_i S_i \\ * & * & * & Z_i & Z_i & 0 \\ * & * & * & * & Y_i & 0 \\ * & * & * & * & * & I \end{bmatrix} \succ 0$$

where $M_i := Y_i A_i + \hat{G}_i C_i$. *If these conditions are satisfied, one can recover admissible filter matrices* F_i, G_i, H_i, K_i *as*

$$\begin{aligned} F_i = V_i^{-1} \hat{F}_i \hat{Z}_i^{-1} U_i^{-T}, \ \ G_i = V_i^{-1} \hat{G}_i \\ H_i = \hat{H}_i Z_i^{-1} U_i^{-T}, \ \ K_i = \hat{K}_i \end{aligned} \tag{18}$$

where U_i, V_i *are any two nonsingular matrices such that* $V_i U_i^T = I - Y_i Z_i^{-1}$.

Proof: For simplicity of notation, let us remove the subscript i in the constraints (16)-(17), since we are considering the design of the filters individually. Also, define $\bar{\lambda} = \lambda / \rho^2$. The condition (16) is satisfied if and only if there exist matrices $W, \tilde{P}_1$ such that [31]

$$\mathrm{Tr}(W) < \mu, \begin{bmatrix} W & \tilde{C} & \tilde{D} \\ * & \tilde{P}_1 & 0 \\ * & * & I \end{bmatrix} \succ 0, \begin{bmatrix} \tilde{P}_1 & \tilde{P}_1 \tilde{A} & \tilde{P}_1 \tilde{B} \\ * & \tilde{P}_1 & 0 \\ * & * & I \end{bmatrix} \succ 0. \tag{19}$$

For the constraint (17), first note that we have equality of the transfer functions

$$\left[\begin{array}{c|c} F & GCS \\ \hline H & KCS \end{array}\right] = \left[\begin{array}{cc|c} A_1 & 0 & 0 \\ 0 & F & GCS \\ \hline 0 & H & KCS \end{array}\right]$$

for any matrix A_1, in particular for A_1 the zero matrix of the same dimensions as A. With this choice, denote

$$\bar{A} = \begin{bmatrix} 0 & 0 \\ 0 & F \end{bmatrix}, \ \ \bar{B} = \begin{bmatrix} 0 \\ GCS \end{bmatrix}, \bar{C} = [0 \quad H], \bar{D} = KCS.$$

Then the constraint (17) can be rewritten $\|\bar{C}(sI - \bar{A})^{-1}\bar{B} + \bar{D}\|_\infty < \bar{\lambda}$, and is satisfied if and only if there exists a matrix $\tilde{P}_2$, of the same dimensions as $\tilde{P}_1$, such that [31]

$$\begin{bmatrix} \tilde{P}_2 & 0 & \tilde{P}_2 \bar{A} & \tilde{P}_2 \bar{B} \\ * & \bar{\lambda} I & \bar{C} & \bar{D} \\ * & * & \tilde{P}_2 & 0 \\ * & * & * & I \end{bmatrix} \succ 0. \tag{20}$$

The sufficient condition of the theorem is obtained by adding the constraint

$$\tilde{P} := \tilde{P}_1 = \tilde{P}_2 \tag{21}$$

and using the change of variable suggested in [32, p. 902]. Namely, assume that there are matrices $F, G, H, K, \tilde{P}$, and W satisfying (19), (20), and (21). We partition the positive definite matrix $\tilde{P}$ and its inverse as

$$\tilde{P} = \begin{bmatrix} Y & V \\ V^T & \hat{Y} \end{bmatrix}, \quad \tilde{P}^{-1} = \begin{bmatrix} X & U \\ U^T & \hat{X} \end{bmatrix}.$$

Note that $YX + VU^T = I$. Define

$$J_1 = \begin{bmatrix} X & I \\ U^T & 0 \end{bmatrix}, \quad J_2 = \begin{bmatrix} I & Y \\ 0 & V^T \end{bmatrix}. \tag{22}$$

Then we have $\tilde{P} J_1 = J_2$. Moreover

$$J_1^T \tilde{P} J_1 = \begin{bmatrix} X & I \\ I & Y \end{bmatrix}, \quad J_1^T \tilde{P} \tilde{B} = \begin{bmatrix} B \\ YB + VGD \end{bmatrix}$$

$$J_1^T \tilde{P} \tilde{A} J_1 = \begin{bmatrix} AX & A \\ YAX + VGCX + VFU^T & YA + VGC \end{bmatrix}$$
$$\tilde{C} J_1 = [(L - KC)X - HU^T \quad L - KC].$$

Similarly, $J_1^T \tilde{P} \bar{A} J_1 = \begin{bmatrix} 0 & 0 \\ VFU^T & 0 \end{bmatrix}$, $J_1^T \tilde{P} \bar{B} = \begin{bmatrix} 0 \\ VGCS \end{bmatrix}$, $\bar{C} J_1 = [HU^T \quad 0]$. Let $Z = X^{-1}$. Consider first the congruence transformations

- of the first LMI in (19) by $\mathrm{diag}(I, J_1, I)$ and then by $\mathrm{diag}(I, Z, I, I)$;
- of the second LMI in (19) by $\mathrm{diag}(J_1, J_1, I)$, and then by $\mathrm{diag}(Z, I, Z, I, I)$;
- and of the LMI (20) by $\mathrm{diag}(J_1, I, J_1, I)$, and then by $\mathrm{diag}(Z, I, I, Z, I, I)$.

Then, the transformation $\hat{F} = VFU^T Z$, $\hat{G} = VG$, $\hat{H} = HU^T Z$, between the filter matrix variables F, G, H and the new variables $\hat{F}, \hat{G}, \hat{H}$ leads to the LMIs of the theorem. Hence these LMIs are necessarily satisfied if the constraints (19), (20) are satisfied together with (21).

Now suppose that the LMIs of the theorem are satisfied. Since $Z \succ 0$, we can define $X = Z^{-1}$. Moreover, since $\begin{bmatrix} Z & Z \\ Z & Y \end{bmatrix} \succ 0$, we have $Y \succ X^{-1}$ by taking the Schur complement, and so $I - XY$ is nonsingular. Hence, we can find two $n \times n$ nonsingular matrices U, V such that $UV^T = I - XY$. Then define the nonsingular matrices J_1, J_2 as in (22), let $\tilde{P} = J_2 J_1^{-1}$, and define the matrices F, G, H, K as in (18). Since J_1 is nonsingular, we can then reverse the congruence transformations to recover (19), (20), which shows that the constraints (16), (17) are satisfied. ■

Remark 2: Note that the problem (15) is also linear in μ_i, λ. These variables can then be minimized subject to the LMI constraints of Theorem 6 in order to design a good filter trading off estimation error and ℓ_2-sensitivity to minimize the overall MSE. However, including these variables directly in the optimization problem can lead to ill-conditioning in the inversion of the matrices U_i, V_i in (18), a phenomenon discussed together with a recommended fix in [32, p. 903]. In addition, minimizing the objective (15) subject to the LMI constraints is different from solving (15)–(17), due to the conservativeness of the conditions in Theorem 6. As in mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problems, one could consider more complex algorithms to reduce this conservativeness [33]. Consider now, instead of (15), the objective $\sum_{i=1}^n \mu_i + \alpha\, \lambda$, where the parameter $\alpha \geq 0$ replaced $\kappa(\delta, \epsilon)^2$, subject to the LMIs of Theorem 6. By setting $\alpha = 0$, we recover exactly the Kalman filter. Hence by performing a one-dimensional search over α we can attempt to improve the overall MSE of the output mechanism over the basic Kalman filter design.

C. Unstable Systems

If the dynamics (9) are not stable, the linear filter design approach presented in the previous paragraph is not valid. To handle this case, we can further restrict the class of filters. As before we minimize the estimation error variance together with the sensitivity measured by the $\mathcal{H}_\infty$ norm of the filter. Starting from the general linear filter dynamics (12), (13), we can consider designs where $\hat{x}_i$ is an estimate of x_i, and set $H_i = L_i, K_i = 0$, so that $\hat{z}_i = L_i \hat{x}_i$ is an estimate of $z_i = L_i x_i$. The error dynamics $e_i := x_i - \hat{x}_i$ then satisfies

$$e_{i,t+1} = (A_i - G_i C_i) x_{i,t} - F_i \hat{x}_{i,t} + (B_i - G_i D_i) w_{i,t}.$$

Setting $F_i = (A_i - G_i C_i)$ gives an error dynamics independent of x_i

$$e_{i,t+1} = (A_i - G_i C_i) e_{i,t} + (B_i - G_i D_i) w_{i,t} \qquad (23)$$

and leaves the matrix G_i as the only remaining design variable. Note however that the resulting class of filters contains the (one-step delayed) Kalman filter. To obtain a bounded error, there is an implicit constraint on G_i that $A_i - G_i C_i$ should be stable.

Now, following the discussion in the previous subsection, minimizing the MSE while enforcing differential privacy leads to the following optimization problem:

$$\min_{\mu_i, \lambda, G_i} \quad \sum_{i=1}^n \mu_i + \kappa(\delta, \epsilon)^2 \lambda \qquad (24)$$
$$\text{s.t. } \|L_i (zI - (A_i - G_i C_i))^{-1} (B_i - G_i D_i)\|_2^2 \leq \mu_i, \quad (25)$$
$$\rho_i^2 \|L_i (zI - (A_i - G_i C_i))^{-1} G_i C_i S_i\|_\infty^2 \leq \lambda \qquad (26)$$
$$\forall\, 1 \leq i \leq n.$$

Again, one can efficiently check a sufficient condition, in the form of the LMIs of the following theorem, guaranteeing that the constraints (25), (26) are satisfied. Optimizing over the variables λ_i, μ_i, G_i can then be done using semidefinite programming.

Theorem 7: The constraints (25)-(26), for some $1 \leq i \leq n$, are satisfied if there exists matrices $Y_i, X_i, \hat{G}_i$ such that

$$\mathrm{Tr}(Y_i L_i^T L_i) < \mu_i, \quad \begin{bmatrix} Y_i & I \\ I & X_i \end{bmatrix} \succ 0$$
$$\begin{bmatrix} X_i & X_i A_i - \hat{G}_i C_i & X_i B_i - \hat{G}_i D_i \\ * & X_i & 0 \\ * & * & I \end{bmatrix} \succ 0 \qquad (27)$$
$$\text{and} \quad \begin{bmatrix} X_i & 0 & X_i A_i - \hat{G}_i C_i & \hat{G}_i C_i S_i \\ * & \frac{\lambda}{\rho_i^2} I & L_i & 0 \\ * & * & X_i & 0 \\ * & * & * & I \end{bmatrix} \succ 0. \qquad (28)$$

If these conditions are satisfied, one can recover an admissible filter matrix G_i by setting $G_i = X_i^{-1} \hat{G}_i$.

Proof: As in Theorem (6), we simplify the notation below by omitting the subscript i. First, from the error dynamics (23), the constraint (25) is satisfied if and only if there exists a positive definite matrix P such that [31]

$$\mathrm{Tr}(PL^T L) < \mu$$
$$F_i P F_i^T + (B_i - G_i D_i)(B_i - G_i D_i)^T \prec P$$

with $F_i = A_i - G_i C_i$. By letting $X = P^{-1}$, introducing the slack variable Y, the change of variable $\hat{G} = XG$, and using the Schur complement, these conditions are equivalent to the existence of two positive definite matrices X, Y such that (27) is satisfied. The LMI (28) derived from (26) is standard [31], see also (20). As in Theorem 6, we restrict the search in this LMI to the same matrix X as in (27), which results in a convex problem but introduces some conservatism. ∎

V. Traffic Monitoring Example

A. System Description

Consider a simplified description of a traffic monitoring system, inspired by real-world implementations and associated privacy concerns as discussed in [6], [34] for example. There are n participating vehicles traveling on a straight road segment. Vehicle i, for $1 \leq i \leq n$, is represented by its state $x_{i,t} = [\xi_{i,t}, \dot{\xi}_{i,t}]^T$, with ξ_i and $\dot{\xi}_i$ its position and velocity, respectively. This state evolves as a second-order system with unknown random acceleration inputs

$$x_{i,t+1} = \begin{bmatrix}1 & T_s\\ 0 & 1\end{bmatrix} x_{i,t} + \sigma_1 \begin{bmatrix}T_s^2/2 & 0\\ T_s & 0\end{bmatrix} w_{i,t}$$

where T_s is the sampling period, $w_{i,t}$ is a standard white Gaussian noise, and $\sigma_1 > 0$. Assume for simplicity that the noise signals w_j for different vehicles are independent. The traffic monitoring service collects GPS measurements from the vehicles [6], i.e., receives noisy readings of the positions at the sampling times

$$y_{i,t} = [1 \quad 0]\, x_{i,t} + \sigma_2\, [0 \quad 1]\, w_{i,t}, \text{ with } \sigma_2 > 0.$$

The purpose of the traffic monitoring service is to continuously provide an estimate of the traffic flow velocity on the road segment, which is approximated by releasing at each sampling period an estimate of the average velocity of the participating vehicles, i.e., of the quantity

$$z_t = \frac{1}{n} \sum_{i=1}^{n} \dot{\xi}_{i,t}. \tag{29}$$

With a larger number of participating vehicles, the sample average (29) represents the traffic flow velocity more accurately. However, while individuals are generally interested in the aggregate information provided by such a system, e.g., to estimate their commute time, they do not wish their own trajectories to be publicly revealed, since these might contain sensitive information about their driving behavior, frequently visited locations, etc. Privacy-preserving mechanisms for such location-based services are often based on ad-hoc temporal and spatial cloaking of the measurements [6], [35]. However, in the absence of a quantitative definition of privacy and a clear model of the adversary's capabilities, it is common that proposed techniques are later argued to be deficient [36], [37]. The temporal cloaking scheme proposed in [6] for example aggregates the speed measurements of k users successively crossing a given line, but does not necessarily protect individual trajectories against adversaries exploiting temporal relationships between these aggregated measurements [36].

B. Numerical Example

We now discuss some differentially private estimators introduced in Section IV, in the context of this example. All individual systems are identical, hence we drop the subscript i in the notation. Assume that the selection matrix is $S = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$, that $\rho = 100$ m in (11), and that we have $n = 200$ participants. Let $T_s = 1$ s, $\sigma_1 = 1\, m.s^{-2}$, $\sigma_2 = 10\, m$. A single Kalman filter denoted $\mathcal{K}$ is designed to provide an estimate $\hat{x}_i$ of each state vector x_i, so that in absence of privacy constraint the final estimate would be

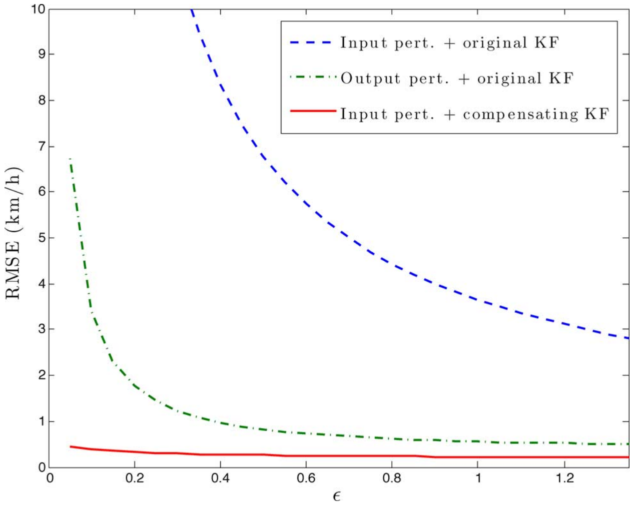

Fig. 4. Steady-state RMSE of the average velocity estimate for three mechanisms, as a function of the privacy parameter ϵ.

$$\hat{z} = [0 \quad \tfrac{1}{n}] \sum_{i=1}^{n} \mathcal{K} y_i = [0 \quad 1]\, \mathcal{K} \left(\frac{1}{n} \sum_{i=1}^{n} y_i \right).$$

We designed the four mechanisms of Section IV for various values of the privacy parameters ϵ, δ. For the output perturbation mechanisms, we used the approach described in Remark 2 to tradeoff estimation error and $\mathcal{H}_\infty$-norm of the filter. For this scenario however, the $\mathcal{H}_\infty$-norm of the Kalman filter is already quite small, and we could only improve marginally the MSE of the Kalman filter based output perturbation mechanism, typically by less than one percent (the improvement can be more significant for smaller values of σ_2 for example). Hence, in the following, we restrict our discussion of output perturbation mechanisms to the simplest scheme that does not redesign the original Kalman filter.

Fig. 4 shows the steady-state root-mean-square error (RMSE) of the mechanisms for different values of ϵ, with fixed $\delta = 0.05$. The input perturbation mechanism, while essentially unusable with the original Kalman filter, shows clearly the best performance when the filter is redesigned by taking the privacy-preserving noise into account as additional measurement noise. The higher performance of this mechanism is especially noticeable in the high-privacy (hence, high-noise) regime, i.e., as ϵ becomes small.

However, other measures of performance are also of interest, and in particular Fig. 5 illustrates the convergence time of the input and output perturbation mechanisms. Here, the filters are simply initialized with an incorrect value of the initial average velocity, but this also serves to illustrate situations where we could have a sudden change in traffic velocity, e.g., due to the formation of a traffic jam. In such cases, it is desirable to have fast convergence of the filter, e.g., in order to warn downstream

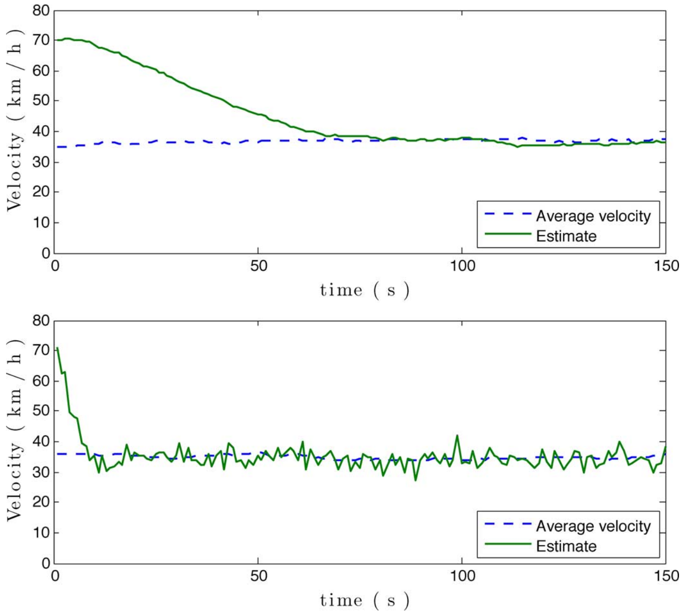

Fig. 5. Two differentially private average velocity estimates: input perturbation and compensating Kalman filter (top), output perturbation and original Kalman filter (bottom), for $\epsilon = 0.3$, $\delta = 0.05$. The filters are initialized with the same incorrect initial mean velocity (70 km/h, instead of 35 km/h). The input perturbation mechanism shows better asymptotic performance but worse convergence time.

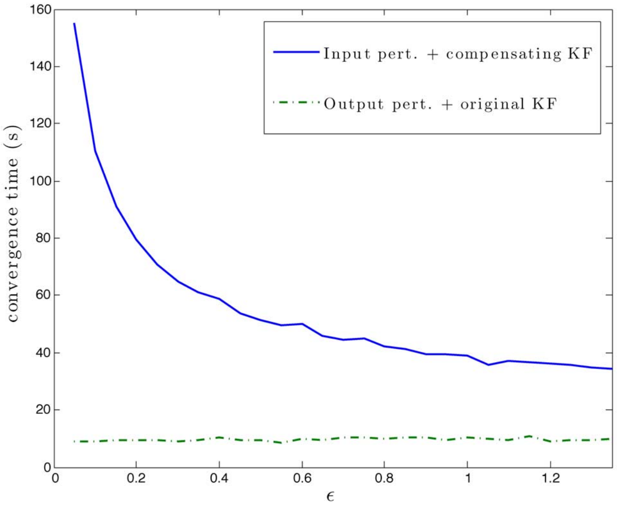

Fig. 6. Time taken by the mechanisms, starting from an incorrect initial average velocity estimate of 70 km/h, to approach the true average vehicle velocity (35 km/h at $t = 0$) within 10% for the first time. The convergence time is estimated by averaging over 20 simulations for each value of ϵ, as illustrated on Fig. 5. Here $\delta = 0.05$.

drivers sufficiently soon. For $\epsilon = 0.3, \delta = 0.05$, the output perturbation mechanism converges in few seconds, whereas the input perturbation mechanism takes more than a minute to converge. In this case, the higher asymptotic RMSE of the first mechanism, which nonetheless remains below 2 km/h, might very well be acceptable in view of the much improved convergence speed. Fig. 6 shows the dependence of the convergence time for these two mechanisms as ϵ varies.

VI. Filtering Event Streams

This section considers an application scenario motivated by the work of [13] and [14]. 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 [13], [14], two signals u and u' are adjacent if and only if they differ by one at a single time, or equivalently

$$\text{Adj}(u, u') \text{ iff } \|u - u'\|_1 = 1. \tag{30}$$

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* [13], that is, hide to some extent the presence or absence of an event at a particular time. This could for example prevent the inference of individual transactions from publicly available collaborative filtering outputs, as in [5]. Now, 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 [13], [14] consider specifically the design of a private counter or accumulator, i.e., a system producing an output signal 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 [15], using a technique similar to [14] based on binary trees. Here we develop certain approximations of general stable LTI systems that preserve event-level privacy. We first make the following remark.

Lemma 3: Let $\mathcal{G}$ be a SISO LTI system with impulse response g. Then for the adjacency relation (30) 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

$$\begin{aligned}\|\mathcal{G}u - \mathcal{G}u'\|_p &= \|\mathcal{G}(u - u')\|_p = \|\mathcal{G}\delta\|_p = \|g * \delta\|_p \\ &= \|g\|_p.\end{aligned}$$

■

We continue to measure the utility of specific schemes throughout this section by the MSE between the published and desired outputs. Similarly to our discussion at the end of Section III, there are two straightforward mechanisms that provide differentially private approximations of $\mathcal{G}$. One can add white noise w directly on 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 setup shown on Fig. 7. The previous mechanisms are recovered when $\mathcal{G}_1$ or $\mathcal{G}_2$ is the identity operator. To show that one can

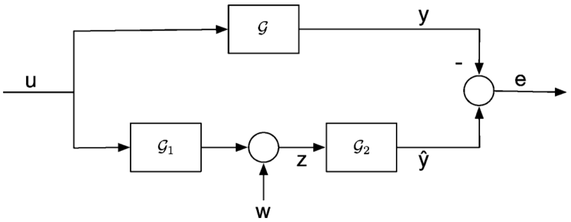

Fig. 7. Differentially private filter approximation setup.

improve the utility of the mechanism with this setup, consider the following choice of filters $\mathcal{G}_1$ and $\mathcal{G}_2$. Let $\mathcal{G}_1$ be a causal LTI system with a causal inverse denoted $\mathcal{G}_1^{-1}$, such that both $\mathcal{G}_1$ and $\mathcal{G}_1^{-1}$ have a square summable impulse response. In the following, we use the term *minimum phase* to refer to such a system $\mathcal{G}_1$. Let $\mathcal{G}_2 := \mathcal{G}\mathcal{G}_1^{-1}$. We call this particular choice of filters a *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 a ZFE mechanism with input filter $\mathcal{G}_1$ is

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

Hence, we are led to consider the following problem:

$$\begin{aligned} &\min_{\mathcal{G}_1} \|\mathcal{G}_1\|_2^2 \|\mathcal{G}\mathcal{G}_1^{-1}\|_2^2 \\ &= \min_{\mathcal{G}_1} \frac{1}{4\pi^2} \int_{-\pi}^{\pi} |\mathcal{G}_1(e^{j\omega})|^2 d\omega \int_{-\pi}^{\pi} \left|\frac{\mathcal{G}(e^{j\omega})}{\mathcal{G}_1(e^{j\omega})}\right|^2 d\omega \end{aligned}$$

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

Theorem 8: Let $\mathcal{G}$ be SISO LTI system with $\|\mathcal{G}\|_2 < \infty$. We have, for any minimum phase system $\mathcal{G}_1$

$$e_{mse}^{ZFE}(\mathcal{G}_1) \geq \kappa(\epsilon, \delta)^2 \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |\mathcal{G}(e^{j\omega})| d\omega\right)^2. \qquad (31)$$

If moreover $\mathcal{G}$ satisfies the Paley–Wiener condition $\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|\mathcal{G}(e^{j\omega})| d\omega > -\infty$, this lower bound on the mean-squared error of the ZFE mechanism can be attained by some minimum phase system $\mathcal{G}_1$ such that $|\mathcal{G}_1(e^{j\omega})|^2 = |\mathcal{G}(e^{j\omega})|$, for almost every $\omega \in [-\pi, \pi)$.

Proof: By the Cauchy–Schwarz inequality

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

hence the bound. Moreover, equality is attained if and only if there exists $\lambda \in \mathbb{R}$ such that, for almost every ω in $[-\pi, \pi)$,

$$|\mathcal{G}_1(e^{j\omega})| = \lambda \left|\frac{\mathcal{G}(e^{j\omega})}{\mathcal{G}_1(e^{j\omega})}\right|, \text{ i.e., } |\mathcal{G}_1(e^{j\omega})|^2 = \lambda|\mathcal{G}(e^{j\omega})|.$$

Now $|\mathcal{G}(e^{j\omega})|$ is a nonnegative function on the unit circle, and if it satisfies the Paley–Wiener condition, it has a minimum phase spectral factor $\mathcal{G}_1$ satisfying $|\mathcal{G}_1(e^{j\omega})|^2 = |\mathcal{G}(e^{j\omega})|$ almost everywhere [38, p. 242], and thus the performance bound can be attained. ■

The MSE obtained for the best ZFE mechanism in Theorem 8 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, better performance 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 [39]. 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 [13] and [14], 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 $\hat{u}_t = u_t + w_t$

$$H(\hat{u}_t) = \begin{cases} 1, & \hat{u}_t \geq 1/2, \\ 0, & \hat{u}_t < 1/2. \end{cases}$$

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, consider using an 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. Let us assume that $\mathcal{G}_1$ is already chosen, e.g., according to Theorem 8 (this choice is not optimal any more if $\mathcal{G}_2$ is not $\mathcal{G}\mathcal{G}_1^{-1}$). Moreover, assume that it is publicly known that u is wide-sense stationary with mean and autocorrelation denoted

$$\mathbb{E}[u_t] = \mu, \;\; \mathbb{E}[u_s u_t] =: R_u[s-t].$$

From this data, the second-order statistics of y and z on Fig. 1 are also known, in particular

$$R_z = f * \tilde{f} * R_u + \sigma^2\delta, R_{yz} = g * \tilde{f} * R_u$$

where $\sigma^2 = \kappa(\delta, \epsilon)^2\|\mathcal{G}_1\|_2^2$, δ is the impulse signal, g and f are the impulse responses of $\mathcal{G}$ and $\mathcal{G}_1$, and $\tilde{f}_t = f_{-t}$. We then design $\mathcal{G}_2$ to minimize the MSE

$$\mathbb{E}[|y_t - \hat{y}_t|^2].$$

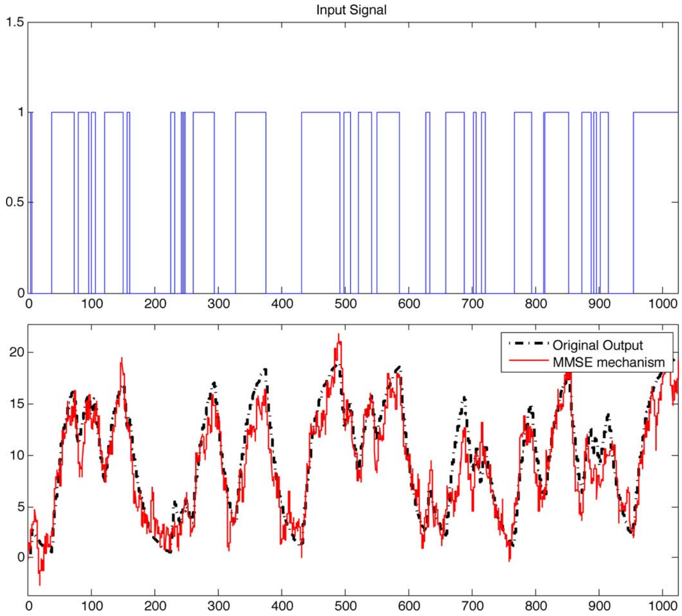

Fig. 8. Sample path for the MMSE mechanism.

For simplicity, consider the case where $\mathcal{G}_2$ is restricted to be a finite-impulse response filter, i.e.,

$$\hat{y}_t = (\mathcal{G}_2 z)_t = \sum_{k=0}^{N} h_k z_{t-k}$$

with N the order of the filter. The vector $h = [h_0, \ldots, h_N]^T$ is the solution of the Wiener–Hopf equations [38]

$$\begin{bmatrix} R_z[0] & R_z[1] & \ldots & R_z[N] \\ R_z[1] & R_z[0] & \ldots & R_z[N-1] \\ \vdots & \vdots & \vdots & \vdots \\ R_z[N] & \ldots & \ldots & R_z[0] \end{bmatrix} h = \begin{bmatrix} R_{yz}[0] \\ \vdots \\ R_{yz}[N] \end{bmatrix}.$$

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.

1) Example 5: Fig. 8 illustrates the differentially private output obtained by the MMSE mechanism approximating the filter $\mathcal{G} = 1/(s(z)+0.05)$, with $s(z)$ the bilinear transformation

$$s(z) = 2\frac{1 - z^{-1}}{1 + z^{-1}}.$$

The input signal is binary valued and the privacy parameters are set to $\epsilon = \ln 3$, $\delta = 0.05$. For this specific input, the empirical MSE of the ZFE is 5.8, compared to 4.6 for the MMSE mechanism. The simpler scheme with noise added at the input is essentially unusable, since its MSE is $\kappa(\delta, \epsilon)^2 \|\mathcal{G}\|_2^2 \approx 30.1$. Adding a detector reduces this MSE to about 17.

B. Related Work

Some papers closely related to the event filtering problem considered in this section are [13]–[15], [40]. As previously mentioned, [13], [40] 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.* [15] 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.

Finally, as with the MMSE estimation mechanism, one can try to use additional information about the input signals to calibrate the amount of noise introduced by the privacy mechanism. For example, if there exists a sparse representation of the signal in some basis (such as a Fourier or a wavelet basis), then one can try to perturb the representation coefficients in this alternate basis. For example, [40] perturbs the largest coefficients of the discrete Fourier transform of the signal. A difficulty with such approaches is that they are typically not causal and not recursive, requiring an amount of processing that increases with time.

VII. 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 [40]. We believe that research on privacy issues is critical to encourage the development of future cyber-physical systems, which typically rely on the users data to improve their efficiency. Numerous directions of study are open to develop privacy-preserving signal processing systems, including designing better filtering mechanisms, and understanding design tradeoffs between privacy or security and performance in large-scale control systems.

Acknowledgment

The authors would like to thank A. Roth for providing valuable insight into the notion of differential privacy.

References

[1] J. Le Ny and G. J. Pappas, “Differentially private Kalman filtering,” in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2012, pp. 1618–1625.

[2] J. Le Ny and G. J. Pappas, “Differentially private filtering,” in *Proc. Conf. Decision Control*, Maui, HI, USA, Dec. 2012.

[3] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[4] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets (how to break anonymity of the Netflix Prize dataset),” in *Proc. IEEE Symp. Security and Privacy*, 2008, pp. 111–125.

[5] J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov, ““you might also like”: Privacy risks of collaborative filtering,” in *Proc. IEEE Symp. Security Privacy*, Berkeley, CA, USA, May 2011, pp. 231–246.

[6] B. Hoh, T. Iwuchukwu, Q. Jacobson, M. Gruteser, A. Bayen, J.-C. Herrera, R. Herring, D. Work, M. Annavaram, and J. Ban, “Enhancing privacy and accuracy in probe vehicle based traffic monitoring via virtual trip lines,” *IEEE Trans. Mobile Comput.*, vol. 11, no. 5, pp. 849–864, May 2012.

[7] H. Zhang and J. Bolot, "Anonymization of location data does not work: A large-scale measurement study," in *Proc. 17th Annu. Int. Conf. Mobile Comput. and Network.*, 2011.

[8] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. 3rd Theory of Cryptogr. Conf.*, 2006, pp. 265–284.

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

[10] C. Dwork, K. Kenthapadi, F. McSherry, I. M. M. Naor, and Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Adv. Cryptol.-EUROCRYPT '06*, 2006, pp. 486–503.

[11] A. Roth, "New algorithms for preserving differential privacy," Ph.D., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2010.

[12] C. Li, M. Hay, V. Rastogi, G. Miklau, and A. McGregor, "Optimizing linear counting queries under differential privacy," in *Principles of Database Syst. (PODS)*, 2010.

[13] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observations," in *Proc. ACM Symp. Theory Comput. (STOC)*, Cambridge, MA, USA, Jun. 2010.

[14] T.-H. H. Chan, E. Shi, and D. Song, "Private and continual release of statistics," *ACM Trans. Info. Syst. Security*, vol. 14, no. 3, pp. 26:1–26:24, Nov. 2011.

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

[16] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proc. CCS Workshop Privacy Electron. Soc. (WPES)*, Raleigh, NC, USA, Oct. 2012.

[17] D. Varodayan and A. Khisti, "Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prag, Czech Republic, 2011, pp. 1932–1935.

[18] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "A theory of privacy and utility in databases," Tech. Rep. Princeton Univ., Princeton, NJ, USA, 2011.

[19] A. Blum, C. Dwork, F. McSherry, and K. Nissim, "Practical privacy: The SuLQ framework," in *Proc. 24th ACM SIGMOD-SIGACT-SIGART Symp. Principles of Database Syst. (PODS)*, New York, NY, USA, 2005, pp. 128–138.

[20] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Colloq. Automata, Lang., Programm. (ICALP)*, 2006, vol. 4052, Lecture Notes in Computer Science.

[21] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness, Knowl.-Based Syst.*, vol. 10, no. 05, pp. 557–570, 2002.

[22] G. Duncan and D. Lambert, "Disclosure-limited data dissemination," *J. Amer. Statist. Associat.*, vol. 81, no. 393, pp. 10–28, Mar. 1986.

[23] S. Gomatam, A. F. Karr, J. P. Reiter, and A. P. Sanil, "Data dissemination and disclosure limitation in a world without microdata: A risk-utility framework for remote access analysis servers," *Statist. Sci.*, vol. 20, no. 2, pp. 163–177, 2005.

[24] R. M. Dudley, *Real Analysis and Probability*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[25] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. IEEE Symp. Foundat. Comput. Sci.*, 2007, pp. 94–103.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[27] L. Breiman, *Probability*. Philadelphia, PA, USA: SIAM, 1992, Classics in Appl. Math..

[28] A. van der Schaft, *L2-Gain and Passivity Techniques in Nonlinear Control*. Berlin, Germany: Springer-Verlag, 2000.

[29] E. Shi, T.-H H. Chan, E. Rieffel, R. Chow, and D. Song, "Privacy-preserving aggregation of time-series data," in *Proc. 18th Annu. Netw. Distrib. Syst. Sec. Symp. (NDSS '11)*, Feb. 2011.

[30] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. New York, NY, USA: Dover, 2005.

[31] R. E. Skelton, T. Iwasaki, and K. Grigoriadis, *A Unified Algebraic Approach to Linear Control Design*. New York, NY, USA: Taylor & Francis, 1998.

[32] C. Scherer, P. Gahinet, and M. Chilali, "Multiobjective output-feedback control via LMI optimization," *IEEE Trans. Autom. Control*, vol. 42, no. 7, pp. 896–911, Jul. 1997.

[33] C. W. Scherer, "An efficient solution to multi-objective control problems with LMI objectives," *Syst. Control Lett.*, vol. 40, pp. 43–57, 2000.

[34] X. Sun, L. Munoz, and R. Horowitz, "Mixture Kalman filter based highway congestion mode and vehicle density estimator and its application," in *Proc. Amer. Control Conf.*, Jul. 2004, pp. 2098–2103.

[35] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. 1st Int. Conf. Mobile Syst., Applicat. Services (MobiSys'03)*, 2003, pp. 31–42.

[36] R. Shokri, J. Freudiger, M. Jadliwala, and J.-P. Hubaux, "A distortion-based metric for location privacy," in *Proc. CCS Workshop Privacy Electron. Soc. (WPES)*, 2009.

[37] R. Shokri, C. Troncoso, C. Diaz, J. Freudiger, and J.-P. Hubaux, "Unraveling an old cloak k-anonymity for location privacy," in *Proc. CCS Workshop Privacy Electron. Soc. (WPES)*, 2010.

[38] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY, USA: Springer, 1994.

[39] J. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill, 2000.

[40] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proc. ACM Conf. Manage. Data (SIGMOD)*, Indianapolis, IN, USA, Jun. 2010.

Jerome Le Ny (S'05–M'09) received the B.S. degree from the Ecole Polytechnique, Palaiseau, France, in 2001, the M.Sc. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2003, and the Ph.D. degree in aeronautics and astronautics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2008.

He has been an Assistant Professor in the Department of Electrical Engineering, École Polytechnique de Montréal, Montreal, QC, Canada, since May 2012. From 2008 to 2012, he was a Postdoctoral Researcher with the GRASP Laboratory at the University of Pennsylvania. His research interests include robust and stochastic control, scheduling and dynamic resource allocation problems, with applications to autonomous and embedded systems, multi-robot systems, and transportation systems.

George J. Pappas (S'90–M'91–SM'04–F'09) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 1998, for which he received the Eliahu Jury Award for Excellence in Systems Research.

He is currently the Joseph Moore Professor of Electrical and Systems Engineering at the University of Pennsylvania, Philadelphia, PA, USA. He is a member of the General Robotics, Automation, Sensing and Perception (GRASP) Laboratory and the PRECISE Center for Embedded Systems. His current research interests include hybrid and embedded systems, hierarchical control systems, distributed control systems, nonlinear control systems, with applications to robotics, unmanned aerial vehicles, biomolecular networks, and green buildings.

Dr. Pappas has received numerous awards, including the National Science Foundation (NSF) CAREER Award in 2002, the NSF Presidential Early Career Award for Scientists and Engineers in 2002, the 2009 George S. Axelby Outstanding Paper Award, and the 2010 Antonio Ruberti Outstanding Young Researcher Prize.