Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Differentially Private Kalman Filtering

Jerome Le Ny and George J. Pappas

Abstract— **This paper studies the $\mathcal{H}_2$ (Kalman) filtering problem in the situation where a signal estimate must be constructed based on inputs from individual participants, whose data must remain private. This problem arises in emerging applications such as smart grids or intelligent transportation systems, where users continuously send data to third-party aggregators performing global monitoring or control tasks, and require guarantees that this data cannot be used to infer additional personal information. To provide formal privacy guarantees against adversaries with arbitrary side information, we rely on the notion of *differential privacy* introduced relatively recently in the database literature. This notion is extended to dynamic systems with many participants contributing independent input signals, and mechanisms are then proposed to solve the $\mathcal{H}_2$ filtering problem with a differential privacy constraint. A method for mitigating the impact of the privacy-inducing mechanism on the estimation performance is described, which relies on controlling the $\mathcal{H}_\infty$ norm of the filter. Finally, we discuss an application to a privacy-preserving traffic monitoring system.**

I. INTRODUCTION

Many large-scale systems, such as smart grids, population health monitoring systems, or traffic monitoring systems, rely on the participation of the users to provide reliable data in real-time, e.g., power consumption, sickness symptoms, or GPS coordinates. However, for privacy or security reasons, the participants benefiting from these services generally do not want to release more information than strictly necessary. Unfortunately, examples of unintended loss of privacy already abound. Indeed, it is possible to infer from the trace of a smart meter the daily activities of a household [1], to re-identify an anonymous GPS trace by correlating it with publicly available information such as a person's home and work locations [2], or to infer individual transactions on commercial websites from temporal changes in public recommendation systems [3]. Providing rigorous privacy guarantees is thus crucial to encourage participation and ultimately realize the benefits promised by these systems.

J. Le Ny is with the department of Electrical Engineering, Ecole Polytechnique de Montreal, QC H3C 3A7, 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.

In our recent work [4], we introduced privacy concerns in the the context of systems theory, by relying on the notion of *differential privacy* [5], a particularly successful definition of privacy used in the database literature. 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 [6].

Most work related to privacy is concerned with the analysis of static databases, whereas cyber-physical systems require mechanisms working with dynamic, time-varying data streams. Recently, information-theoretic approaches have been proposed to guarantee some level of privacy when releasing time series [7], [8]. However, the resulting privacy guarantees only hold if the statistics of the participants' data streams obey the assumptions made (typically stationarity and distributional assumptions), and require the explicit statistical modeling of all available side information. In contrast, differential privacy is a worst-case notion that holds independently of any probabilistic assumption on the dataset, and controls the information leakage against adversaries with arbitrary side information [6].

In this paper, we pursue our work on differential privacy for dynamical systems [4], by considering the $\mathcal{H}_2$ filtering problem (or steady-state Kalman filtering) with a differential privacy constraint. The goal is to minimize an estimation error variance for a desired linear combination of the participants' state trajectories, based on their contributed measurements, while guaranteeing the privacy of the individual signals. In contrast to the generic filtering mechanisms presented in [4], we show here how a model of the participants' dynamics can be used to publish more accurate results, without compromising the differential privacy guarantee if this model is not accurate. Section II provides some technical background on differential privacy and Section III describes a basic mechanism enforcing privacy for dynamical systems by injecting additional white noise. As shown in [4], accurate private results can be published for filters with small incremental gains with respect to the individual input channels. This leads us in Section IV to present a modification of the standard Kalman filter, essentially controlling its $\mathcal{H}_\infty$ norm simultaneously with the steady-state estimation error, in order to minimize the impact of the privacy-preserving mechanism. Finally, Section V describes an application to a traffic monitoring system relying on location traces from the participants to provide an average velocity estimate on a road segment. Most proofs are omitted from this extended abstract and will appear in the full version of the paper [9].

II. Differential Privacy

In this section we review the notion of differential privacy [5] as well as a basic mechanism that can be used to achieve it when the released data belongs to a finite-dimensional vector space. We refer the reader to the surveys by Dwork, e.g., [10], 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 data tables, or a signal space). A *mechanism* is just a map $M : \mathsf{D} \times \Omega \to \mathsf{R}$, for some measurable output space R, such that for any element $d \in \mathsf{D}$, $M(d, \cdot)$ is a random variable, typically writen 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. Next, we introduce the definition of differential privacy. 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 $\text{Adj}(d, d')$ 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 $\text{Adj}(d, d')$, we have

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

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

The 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. This result is similar in spirit to the data processing inequality from information theory. 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 $d \in \mathsf{D}$. 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 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.

B. A Basic Differentially Private Mechanism

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 a basic mechanism 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 $k > 0$, 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 [5].

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

$$\Delta_{\mathsf{R}} q := \max_{d, d' : \text{Adj}(d, d')} \|q(d) - q(d')\|_{\mathsf{R}}.$$

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 $\Delta_p q$.

A differentially private mechanism proposed in [11], modifies an answer to a numerical query by adding iid zero-mean noise distributed according to a Gaussian distribution. 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 [11].

Theorem 2: 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 2 can be written $\sigma(\delta, \epsilon) = \kappa(\epsilon, \delta)\Delta_2 q$. It can be shown that $\kappa(\delta, \epsilon)$ behaves roughly as $O(\ln(1/\delta))^{1/2}/\epsilon$. For example, to guarantee (ϵ, δ)-differential privacy with $\epsilon = \ln(2)$ and $\delta = 0.05$, we obtain that the standard deviation of the Gaussian noise introduced should be about 2.65 times the ℓ_2-sensitivity of q.

III. Differentially Private Dynamic Systems

In this section we review the notion of differential privacy for dynamic systems, following [4]. 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. 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 $\ell_{p,e}^m$ the space of sequences with values in $\mathbb{R}^m$ and such that $x \in \ell_{p,e}^m$ 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. 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 $u_i \in \ell_{r_i,e}^{m_i}$ 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 $l_{r,e}^m = \ell_{r_1,e}^{m_1} \times \ldots \times \ell_{r_n,e}^{m_n}$ 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

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

Note that in (3) two signals u_i, u_i' are considered different if there exists some time t at which $u_{i,t} \neq u_{i,t}'$.

A. The Dynamic Gaussian Mechanism

Recall (see, e.g., [12]) that for a system $\mathcal{G}$ with inputs in $\ell_{r,e}^m$ and output in $\ell_{s,e}^{m'}$, its ℓ_r-to-ℓ_s incremental gain $\gamma_{r,s}^{inc}(\mathcal{G})$ is defined as the smallest number γ such that

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

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

$$\mathcal{G}(u_1, \ldots, u_n) = \sum_{i=1}^n \mathcal{G}_i u_i, \tag{4}$$

where $\mathcal{G}_i : \ell_{r_i,e}^{m_i} \to \ell_{s,e}^{m'}$, for all $1 \leq i \leq n$. The next theorem generalizes the Gaussian mechanism of Theorem 2 to causal dynamic systems.

Theorem 3: Let $\mathcal{G}$ be defined as in (4) and consider the adjacency relation (3). Then the mechanism $Mu = \mathcal{G}u + w$, where w is a white noise with $w_t \sim \mathcal{N}(0, \sigma^2 I_{m'})$ and $\sigma \geq \kappa(\delta, \epsilon) \max_{1 \leq i \leq n}\{\gamma_{r_i,2}^{inc}(\mathcal{G}_i)\, b_i\}$, is (ϵ, δ)-differentially private.

Corollary 1: Let $\mathcal{G}$ be defined as in (4) 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).

B. 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 $b_1^2 = \ldots = b_n^2 = \mathcal{E}$ 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. 1. 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 $\sigma^2 = \kappa(\delta, \epsilon)^2\mathcal{E}$. These perturbations on each input channel are then passed through $\mathcal{G}$, leading

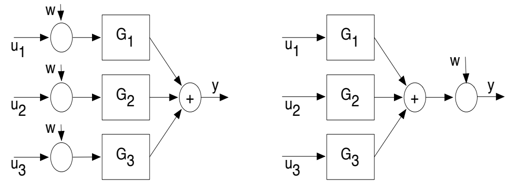

Fig. 1. Two architectures for differential privacy. (a) Input perturbation. (b) Output perturbation.

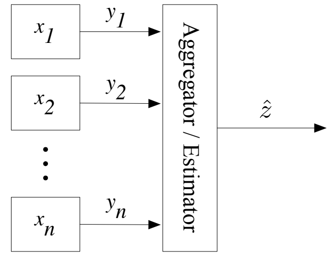

Fig. 2. Kalman filtering set-up.

to a mean squared error (MSE) for the output equal to $\kappa(\delta,\epsilon)^2\mathcal{E}\|\mathcal{G}\|_2^2 = \kappa(\delta,\epsilon)^2\mathcal{E}\sum_{i=1}^n \|\mathcal{G}_i\|_2^2$. 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 $\kappa(\delta,\epsilon)^2\mathcal{E}\max_{1\leq i\leq n}\{\|\mathcal{G}_i\|_\infty^2\}$. 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. 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., [13].

IV. Kalman Filtering

We now discuss the Kalman filtering problem subject to a differential privacy constraint. Compared to the previous section, for Kalman filtering it is assumed that more is publicly known about the dynamics of the processes producing the individual signals. The goal here is to guarantee differential privacy for the individual state trajectories. Section V describes an application of the privacy mechanisms presented here to a traffic monitoring problem.

A. A Differentially Private Kalman Filter

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

$$x_{i,t+1} = A_i x_{i,t} + B_i w_{i,t}, \;\; t \geq 0, \;\; 1 \leq i \leq n, \quad (5)$$

where w_i is a standard zero-mean Gaussian white noise process with covariance $\mathbb{E}[w_{i,t}w_{i,t'}] = \delta_{t-t'}$, 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 $1 \leq i \leq n$, sends measurements

$$y_{i,t} = C_i x_{i,t} + D_i w_{i,t} \quad (6)$$

to a data aggregator. We assume for simplicity that the matrices D_i are full row rank. Figure 2 shows this initial set-up.

The data aggregator aims at releasing a signal that asymptotically minimizes the minimum mean squared error 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 $y_i, 1 \leq i \leq n$, 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 $\bar{x}_{i,0}, A_i, B_i, C_i, D_i, L_i, 1 \leq i \leq n$, are assumed to be public information. For all $1 \leq i \leq n$, 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 optimal 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 [14], and denoted $\mathcal{K}_i$ in the following.

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

$$\text{Adj}_{\mathcal{S}}^{\rho}(x, x') \text{ iff for some } i, \; \|T_i x_i - T_i x_i'\|_2 \leq \rho_i, \quad (7)$$
$$(I - T_i)x_i = (I - T_i)x_i', \text{and } x_j = x_j' \text{ for all } j \neq i.$$

In words, two adjacent global output signals differ by the trajectory of a single participant, say i. Moreover, for differential privacy guarantees we are constraining the range in energy variation in the signal $T_i x_i$ 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 output signal value $T_i x_{i,t_0}$ at some single specific time t_0 were replaced by $T_i x'_{i,t_0}$ with $\|T_i(x_{i,t_0} - x'_{i,t_0})\| \leq \rho_i$, but the privacy guarantee should also hold for smaller instantaneous deviations on longer segments of trajectory.

Depending on which signals on Fig. 2 are actually published, and similarly to the discussion of Section III-B, there are different points at which a privacy inducing noise can be introduced. First, for the input noise injection mechanism, the noise can be added by each participant directly to their transmitted measurement signal y_i. Namely, since for two state trajectories x_i, x'_i adjacent according to (7) we have for the corresponding measured signals

$$\|y_i - y'_i\|_2 = \|C_i T_i(x_i - x'_i)\|_2,$$

differential privacy can be guaranteed if participant i adds to y_i a white Gaussian noise with covariance matrix $\kappa(\delta,\epsilon)^2 \rho_i^2 \sigma_{\max}^2(C_i T_i) I_{p_i}$, 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. At the data aggregator, this additional noise can be taken into account in the design of the Kalman filter, since it can simply be viewed as an additional measurement noise. Again, this mechanism is simple to implement when the participants do not trust the data aggregator, since the transmitted signals are already differentially private.

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 (7), and let $\hat{z}, \hat{z}'$ be the corresponding estimates produced. We have

$$\hat{z} - \hat{z}' = L_i \mathcal{K}_i (y_i - y'_i) = L_i \mathcal{K}_i C_i T_i (x_i - x'_i),$$

where $\mathcal{K}_i$ is the i^{th} Kalman filter. Hence $\|\hat{z} - \hat{z}'\|_2 \leq \gamma_i \rho_i$, where γ_i is the $\mathcal{H}_\infty$ norm of the transfer function $L_i \mathcal{K}_i C_i T_i$. We thus have the following theorem.

Theorem 4: 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 $\{w_i\}_{1 \leq i \leq n}, \{x_{i,0}\}_{1 \leq i \leq n}$, and $\gamma = \max_{1 \leq i \leq n}\{\gamma_i \rho_i\}$, with γ_i the $\mathcal{H}_\infty$ norm of $L_i \mathcal{K}_i C_i T_i$, is differentially private for the adjacency relation (7).

B. Filter Redesign for Stable Systems

In the case of the output perturbation mechanism, one can potentially improve the MSE of the filter with respect to the Kalman filter considered in the previous subsection. Namely, consider the design of n filters of the form

$$\hat{x}_{i,t+1} = F_i \hat{x}_{i,t} + G_i y_{i,t} \quad (8)$$

$$\hat{z}_{i,t} = H_i \hat{x}_{i,t} + K_i y_{i,t}, \quad (9)$$

for $1 \leq i \leq n$, where F_i, G_i, H_i, K_i are matrices to determine. The estimator considered is

$$\hat{z}_t = \sum_{i=1}^n \hat{z}_{i,t},$$

so that each filter output $\hat{z}_i$ should minimize the steady-state error variance with $z_i = L_i x_i$, and the released signal $\hat{z}$ should guarantee the differential privacy with respect to (7). 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 $1 \leq i \leq n$.

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 then be written

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

where

$$\tilde{A}_i = \begin{bmatrix} A_i & 0 \\ G_i C_i & F_i \end{bmatrix}, \quad \tilde{B}_i = \begin{bmatrix} B_i \\ G_i D_i \end{bmatrix},$$
$$\tilde{C}_i = \begin{bmatrix} L_i - K_i C_i & -H_i \end{bmatrix}, \quad \tilde{D}_i = -K_i D_i.$$

The steady-state MSE for the i^{th} estimator is then $\lim_{t\to\infty} \mathbb{E}[e_{i,t}^T e_{i,t}]$. Moreover, we are interested in designing filters with small $\mathcal{H}_\infty$ norm, in order to minimize the amount of noise introduced by the privacy-preserving mechanism, which ultimately impacts 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 (7), and letting $\delta x_i = x_i - x'_i = T_i(x_i - x'_i) = T_i \delta x_i$, we see that the change in the output of filter i follows the dynamics

$$\delta \hat{x}_{i,t+1} = F_i \delta \hat{x}_{i,t} + G_i C_i T_i \delta x_i$$
$$\delta \hat{z}_i = H_i \delta \hat{x}_{i,t} + K_i C_i T_i \delta x_i.$$

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

$$\left[\begin{array}{c|c} F_i & G_i C_i T_i \\ \hline H_i & K_i C_i T_i \end{array}\right]. \tag{10}$$

Simply replacing the Kalman filter in Theorem 4, the MSE for the output perturbation mechanism guaranteeing (ϵ, δ)-privacy is then

$$\sum_{i=1}^n \|\tilde{C}_i(zI - \tilde{A}_i)^{-1}\tilde{B}_i + \tilde{D}_i\|_2^2 + \kappa(\delta,\epsilon)^2 \max_{1\leq i\leq n}\{\gamma_i^2 \rho_i^2\},$$
$$\text{with } \gamma_i := \|H_i(zI - F_i)^{-1}G_iC_iT_i + K_iC_iT_i\|_\infty.$$

Hence minimizing this MSE leads us to the following optimization problem

$$\min_{\mu_i,\lambda,F_i,G_i,H_i,K_i} \quad \sum_{i=1}^n \mu_i + \kappa(\delta,\epsilon)^2\lambda \tag{11}$$
$$\text{s.t. } \forall\, 1 \leq i \leq n, \|\tilde{C}_i(zI - \tilde{A}_i)^{-1}\tilde{B}_i + \tilde{D}_i\|_2^2 \leq \mu_i, \tag{12}$$
$$\rho_i^2 \|H_i(zI - F_i)^{-1}G_iC_iT_i + K_iC_iT_i\|_\infty^2 \leq \lambda. \tag{13}$$

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 (12)-(13). These LMIs can be obtained using the change of variable technique described in [15].

Theorem 5: The constraints (12)-(13), for some $1 \leq i \leq n$, 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 $\text{Tr}(W_i) < \mu_i$, and the LMIs (14), (15) shown next page are satisfied.

If these conditions are satisfied, one can recover admissible filter matrices F_i, G_i, H_i, K_i by setting

$$F_i = V_i^{-1}\hat{F}_i\hat{Z}_i^{-1}U_i^{-T}, \;\; G_i = V_i^{-1}\hat{G}_i,$$
$$H_i = \hat{H}_i Z_i^{-1}U_i^{-T}, \;\; K_i = \hat{K}_i, \tag{16}$$

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

Note that the problem (11) is also linear in μ_i, λ. These variables can then be minimized subject to the LMI constraints of Theorem 5 in order to design a good filter trading off estimation error and ℓ^2-sensitivity to minimize the overall MSE.

C. Unstable Systems

If the dynamics (5) 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 (8), (9), 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 $z_i = L_ix_i$. The error dynamics $e_i := x_i - \hat{x}_i$ then satisfies

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

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

$$e_{i,t+1} = (A_i - G_iC_i)e_{i,t} + (B_i - G_iD_i)w_{i,t}, \tag{17}$$

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 $A_i - G_iC_i$ should be stable.

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

$$\min_{\mu_i,\lambda,G_i} \quad \sum_{i=1}^n \mu_i + \kappa(\delta,\epsilon)^2\lambda \tag{18}$$
$$\text{s.t. } \forall\, 1 \leq i \leq n,$$
$$\|L_i(zI - (A_i - G_iC_i))^{-1}(B_i - G_iD_i)\|_2^2 \leq \mu_i, \tag{19}$$
$$\rho_i^2\|L_i(zI - (A_i - G_iC_i))^{-1}G_iC_iT_i\|_\infty^2 \leq \lambda. \tag{20}$$

Again, one can efficiently check a sufficient condition, in the form of the LMIs of the following theorem, guaranteeing that the constraints (19), (20) are satisfied. Optimizing over the variables λ_i, μ_i, G_i can then be done using semidefinite programming.

Theorem 6: The constraints (19)-(20), for some $1 \leq i \leq n$, are satisfied if there exists matrices $Y_i, X_i, \hat{G}_i$ such that

$$\text{Tr}(Y_iL_i^TL_i) < \mu_i, \quad \begin{bmatrix} Y_i & I \\ I & X_i \end{bmatrix} \succ 0, \tag{21}$$

$$\begin{bmatrix} X_i & X_iA_i - \hat{G}_iC_i & X_iB_i - \hat{G}_iD_i \\ * & X_i & 0 \\ * & * & I \end{bmatrix} \succ 0, \tag{22}$$

$$\text{and} \quad \begin{bmatrix} X_i & 0 & X_iA_i - \hat{G}_iC_i & \hat{G}_iC_iT_i \\ * & \frac{\lambda}{\rho_i^2}I & L_i & 0 \\ * & * & X_i & 0 \\ * & * & * & I \end{bmatrix} \succ 0. \tag{23}$$

If these conditions are satisfied, one can recover an admissible filter matrice G_i by setting

$$G_i = X_i^{-1}\hat{G}_i.$$

$$\begin{bmatrix} W_i & (L_i - \hat{K}_i C_i - \hat{H}_i) & (L_i - \hat{K}_i C_i) & -\hat{K}_i D_i \\ * & Z_i & Z_i & 0 \\ * & * & Y_i & 0 \\ * & * & * & I \end{bmatrix} \succ 0, \quad \begin{bmatrix} Z_i & Z_i & 0 & 0 & 0 & 0 \\ * & Y_i & 0 & \hat{F}_i & 0 & \hat{G}_i C_i T_i \\ * & * & \frac{\lambda}{\rho_i^2} I & \hat{H}_i & 0 & \hat{K}_i C_i T_i \\ * & * & * & Z_i & Z_i & 0 \\ * & * & * & * & Y_i & 0 \\ * & * & * & * & * & I \end{bmatrix} \succ 0, \quad (14)$$

$$\begin{bmatrix} Z_i & Z_i & Z_i A_i & Z_i A_i & Z_i B_i \\ * & Y_i & (Y_i A_i + \hat{G}_i C_i + \hat{F}_i) & (Y_i A_i + \hat{G}_i C_i) & (Y_i B_i + \hat{G}_i D_i) \\ * & * & Z_i & Z_i & 0 \\ * & * & * & Y_i & 0 \\ * & * & * & * & I \end{bmatrix} \succ 0. \quad (15)$$

V. A Traffic Monitoring Example

Consider a simplified description of a traffic monitoring system, inspired by real-world implementations and associated privacy concerns as discussed in [2], [16] 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

$$x_{i,t+1} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} x_{i,t} + \sigma_{i1} \begin{bmatrix} T_s^2/2 & 0 \\ T_s & 0 \end{bmatrix} w_{i,t},$$

where T_s is the sampling period, $w_{i,t}$ is a standard white Gaussian noise, and $\sigma_{i1} > 0$. Assume for simplicity that the noise signals w_j for different vehicles are independent. The traffic monitoring service collects GPS measurements from the vehicles [2], thus getting noisy readings of the positions at the sampling times

$$y_{i,t} = \begin{bmatrix} 1 & 0 \end{bmatrix} x_{i,t} + \sigma_{i2} \begin{bmatrix} 0 & 1 \end{bmatrix} w_{i,t},$$

with $\sigma_{i2} > 0$.

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

$$z_t = \frac{1}{n} \sum_{i=1}^{n} \dot{\xi}_{i,t}. \quad (24)$$

With a larger number of participating vehicles, the sample average (24) 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 individual 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 [2], [17]. However, in the absence of a clear model of the adversary capabilities, it is common that proposed techniques are later argued to be deficient [18].

A. Numerical Example

We now discuss some differentially private estimators introduced above, 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 $T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, that $\rho = 100$ m, $T_s = 1s$, $\sigma_{i1} = \sigma_{i2} = 1$, and $\epsilon = \ln 3$, $\delta = 0.05$. 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

$$\hat{z} = \begin{bmatrix} 0 & \frac{1}{n} \end{bmatrix} \sum_{i=1}^{n} \mathcal{K} y_i = \begin{bmatrix} 0 & 1 \end{bmatrix} \mathcal{K} \left(\frac{1}{n} \sum_{i=1}^{n} y_i \right).$$

Finally, assume that we have $n = 200$ participants, and that their mean initial velocity is 45 km/h.

In this case, the input noise injection scheme without modification of the Kalman filter is essentially unusable since its steady-state Root-Mean-Square-Error (RMSE) is almost 26 km/h. However, modifying the Kalman filter to take the privacy inducing noise into account as additional measurement noise leads to the best RMSE of all the schemes discussed here, of about 0.31 km/h. Using the Kalman filter $\mathcal{K}$ with the output noise injection scheme leads to an RMSE of 2.41 km/h. Moreover in

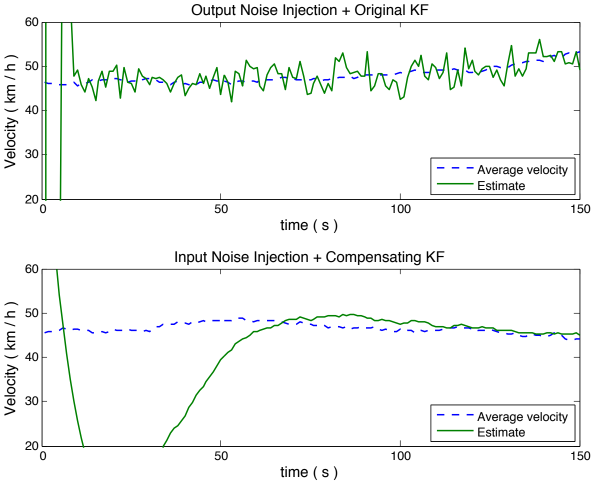

Fig. 3. Two differentially private average velocity estimates, with $n = 200$ users. The Kalman filters are initialized with the same incorrect initial mean velocity (75 km/h), in order to evaluate their convergence time.

this case $\|\mathcal{K}\|_\infty = 0.57$ is quite small, and trying to balance estimation with sensitivity using the LMI of Theorem 6 (by minimizing the MSE while constraining the $\mathcal{H}_\infty$ norm rather than using the objective function (18)) only allowed us to reduce this RMSE to 2.31 km/h. However, an issue that is not captured in these steady-state estimation error measures is that of convergence time of the filters. This is illustrated on Fig. 3, which shows a trajectory of the average velocity of the participants, together with the estimates produced by the input noise injection scheme with compensating Kalman filter and the output noise injection scheme following $\mathcal{K}$. Although the RMSE of the first scheme is much better, its convergence time of more than 1 min, due to the large measurement noise assumed, is much larger. This can make this scheme impractical, e.g., if the system is supposed to respond quickly to an abrupt change in average velocity.

VI. Conclusion

We have discussed mechanisms for preserving the differential privacy of individual users transmitting measurements of their state trajectories to a trusted central server releasing sanitized filtered outputs based on these measurements. Decentralized versions of these mechanisms can in fact be implemented in the absence of trusted server by means of cryptographic techniques [19]. Further research on privacy issues associated with emerging large-scale information processing and control systems is critical to encourage their development. Moreover, obtaining a better understanding of the design trade-offs between privacy or security and performance in these systems raises interesting system theoretic questions.

References

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

[2] 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 Transactions on Mobile Computing*, vol. 11, no. 5, pp. 849–864, May 2012.

[3] J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov, ""you might also like": Privacy risks of collaborative filtering," in *IEEE Symposium on Security and Privacy*, Berkeley, CA, May 2011.

[4] J. Le Ny and G. J. Pappas, "Differentially private filtering," in *Proceedings of the Conference on Decision and Control*, Maui, HI, December 2012.

[5] 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.

[6] 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

[7] 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.

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

[9] J. Le Ny and G. J. Pappas, "Differentially private filtering," available at http://arxiv.org/abs/1207.4305.

[10] 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.

[11] 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.

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

[13] E. Shi, T.-H. H. Chan, E. Rieffel, R. Chow, and D. Song, "Privacy-preserving aggregation of time-series data," in *Proceedings of 18th Annual Network and Distributed System Security Symposium (NDSS 2011)*, February 2011.

[14] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Dover, 2005.

[15] C. Scherer, P. Gahinet, and M. Chilali, "Multiobjective output-feedback control via LMI optimization," *IEEE Transactions on Automatic Control*, vol. 42, no. 7, pp. 896–911, July 1997.

[16] X. Sun, L. Munoz, and R. Horowitz, "Mixture Kalman filter based highway congestion mode and vehicle density estimator and its application," in *Proceedings of the American Control Conference*, July 2004, pp. 2098–2103.

[17] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *ACM MobiSys*, 2003.

[18] R. Shokri, J. Freudiger, M. Jadliwala, and J.-P. Hubaux, "A distortion-based metric for location privacy," in *Proceedings of the CCS Workshop on Privacy in the Electronic Society (WPES)*, 2009.

[19] 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.