Privacy-Preserving Release of Aggregate Dynamic Models

Jerome Le Ny
Department of Electrical Engineering
Ecole Polytechnique de Montréal
C.P. 6079, succ. Centre-ville
Montréal, QC H3C 3A7, Canada
jerome.le-ny@polymtl.ca

George J. Pappas
Department of Electrical & Systems Engineering
University of Pennsylvania
200 South 33rd Street
Philadelphia, PA 19104, USA
pappasg@seas.upenn.edu

ABSTRACT

New solutions proposed for the monitoring and control of large-scale systems increasingly rely on sensitive data provided by end-users. As a result, there is a need to provide guarantees that these systems do not unintentionally leak private and confidential information during their operation. Motivated by this context, this paper discusses the problem of releasing a dynamic model describing the aggregate input-output dynamics of an ensemble of subsystems coupled via a common input and output, while controlling the amount of information that an adversary can infer about the dynamics of the individual subsystems. Such a model can then be used as an approximation of the true system, e.g., for controller design purposes. The proposed schemes rely on the notion of differential privacy, which provides strong and quantitative privacy guarantees that can be used by individuals to evaluate the risk/reward trade-offs involved in releasing detailed information about their behavior.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Control theory*; C.3 [**Special-Purpose and Application-Based Systems**]: Signal processing systems

General Terms

Security, Algorithms

Keywords

Differential Privacy; Dynamical Systems; System Identification

1. INTRODUCTION

Many emerging large-scale monitoring and control systems are expected to leverage vast amount of data provided by individual users to provide real-time services more efficiently. Examples include smart grids with advanced metering capabilities [15], intelligent transportation systems exploiting users location traces [11], or real-time population health monitoring systems [16]. The privacy concerns associated with the deployment of such pervasive monitoring networks have often been overlooked, in view of the many promised benefits. However, with certain recent setbacks in the rollout of smart meter projects for example, it is becoming increasingly clear that not addressing these concerns rigorously at design time puts at real risk the adoption of these technologies.

In contrast to these real-time systems, there is extensive work in the statistics and database literature on disclosure limitation and privacy-preserving publication of data [5, 6]. In particular, the recently proposed notion of differential privacy [2, 6, 8] has been adopted in many works as a definition of privacy offering quantitative guarantees. This notion characterizes certain randomized algorithms that produce answers to statistical queries according to a distribution that is not very sensitive to the presence or absence of the data of any single individual. As a result, people contributing their data to a database are guaranteed that they are not dramatically increasing the ability of an adversary to infer private information about them, even by linking the released answers to the queries to other available sources of information.

Previous work on privacy for dynamic systems addressed the problem of releasing time-series, with or without real-time constraints, under various definitions of privacy [3, 9, 13,14,18]. In this paper however, we consider the problem of releasing a *model*, rather than actual signals, capturing the aggregate dynamic behavior of a large number of participants, while restricting the accuracy with which the dynamics of any single participant can be inferred. Such models are useful for system simulation, forecasts, or control design. For example, consider an Independent System Operator (ISO) in an electricity market requesting from industrial producers and consumers a model of how fast and by how much they would ramp down and up their production and consumption when exposed to fluctuating spot prices, in order to implement a stable demand response scheme [17]. By observing the actual operation of such a scheme, .e.g., through the temporal behavior of electricity prices, it might be possible to infer some information about these confidential models, such as the equipment owned by specific companies or the value they extract from consuming a certain amount of electricity.

HiCoNS'13, April 9–11, 2013, Philadelphia, Pennsylvania, USA.

After some preliminaries on differential privacy in Section 2, we formulate in Section 3 a simple aggregate model release scenario of the type described in the previous paragraph, involving linear scalar dynamics to describe the individual subsystems. Three schemes preserving differential privacy for the subsystems are then proposed, based respectively on perturbation of the parameters of the transfer function, on perturbation of the Markov parameters or impulse response of a sampled version of the model, and on sampling and perturbing the frequency response. The ability of these schemes to capture accurately the global input-output behavior of the aggregate system is then discussed using simulated examples in Section 4.

2. TOOLS FOR DIFFERENTIAL PRIVACY

2.1 Definitions

We first introduce the notion of differential privacy [7, 8], which is a property of certain mechanisms accessing datasets containing private information to answer queries. Let us fix some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let D be a space of datasets of interest. A *mechanism* is just a map $M : \mathsf{D} \times \Omega \to \mathsf{R}$, for some measurable output space $(\mathsf{R}, \mathcal{M})$, where $\mathcal{M}$ denotes a σ-algebra, such that for any element $d \in \mathsf{D}$, $M(d)$ is a random variable. A mechanism can be viewed as a probabilistic algorithm to answer a specific query q, which is a map $q : \mathsf{D} \to \mathsf{R}$.

Differential privacy is defined with respect to a choice of symmetric binary relation Adj on D, called adjacency, defined so that $\mathrm{Adj}(d, d')$ if and only if d and d' differ by the data of a single participant. A differentially private mechanism produces randomized outputs, with a distribution that does not change much for two adjacent datasets. As a result, an individual choosing to contribute its data is guaranteed that this choice won't dramatically increase the ability of an adversary to infer additional private information about him.

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

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

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

In this paper, the input data of the participants belongs to a vector space. We now introduce two adjacency relations that are useful for this situation. They bound the variations allowed in the individual input data for which the condition (1) can be guaranteed. Let the space of datasets of interest for n participants be a product vector space $\mathsf{D} = \mathsf{D}_1 \times \ldots \times \mathsf{D}_n$, where D_i is equipped with a norm $\|\cdot\|_i$. Let $x, x' \in \mathsf{D}$ and $\rho \in \mathbb{R}^n_{>0}$. The first binary relation controls the absolute variation

$$\mathrm{Adj}_a^{\rho}(x, x') \text{ iff for some } i, \|x_i - x_i'\|_i \leq \rho_i, \quad \text{and } x_j = x_j', \forall j \neq i. \qquad (2)$$

Many algorithms, notably in numerical analysis, have a sensitivity that is typically measured by relative variations rather than absolute variations, see, e.g., [19]. To capture these situations, we also introduce the following binary relation, for $\eta \in \mathbb{R}^n_{>0}$,

$$\mathrm{Adj}_r^{\eta}(x, x') \text{ iff for some } i, \frac{\|x_i - x_i'\|_i}{\min\{\|x_i\|_i, \|x_i'\|_i\}} \leq \eta_i, \quad \text{and } x_j = x_j', \forall j \neq i. \qquad (3)$$

The relation is undefined is $\min\{\|x_i\|_i, \|x_i'\|_i\} = 0$ in (3). Note that a differentially private mechanism for this adjacency relation produces similar outputs if a single participants changes its data from x_i to any $x_i' = x_i(1 + \tilde{\eta}_i)$, with $|\tilde{\eta}_i| \leq \eta_i$.

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 without looking back at the original data. This result is similar in spirit to the data processing inequality from information theory [4]. 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*

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

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

A proof of this theorem can be found in [13]. 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 shows that a mechanism M_2 accessing a dataset only indirectly via the output of a differentially private mechanism M_1 cannot weaken the privacy guarantee. Hence post-processing can be used freely to improve the *accuracy* of an output, without having to worry about a possible loss of privacy.

2.2 Basic Mechanisms for Numerical Queries

Two basic mechanisms [7, 8], introduced in the next theorem, achieve differential privacy by additively perturbing the answers to numerical queries. They involve the following notion of sensitivity of a query.

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

Next, recall that the Laplace distribution with mean zero and scale parameter b, denoted $\mathrm{Lap}(b)$, has density $p(x; b) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$ and variance $2b^2$. Moreover, for $w \in \mathbb{R}^k$ with w_i iid and $w_i \sim \mathrm{Lap}(b)$, denoted $w \sim \mathrm{Lap}(b)^k$, we have $p(w; b) = (\frac{1}{2b})^k \exp\left(-\frac{\|w\|_1}{b}\right)$, $\mathbb{E}[\|w\|_1] = b$, and $\mathbb{P}(\|w\|_1 \geq tb) = e^{-t}$. The multidimensional normal distribution with

mean μ and covariance matrix Σ is denoted $\mathcal{N}(\mu, \Sigma)$. Finally, the $\mathcal{Q}$-function is defined as

$$\mathcal{Q}(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du.$$

THEOREM 2. *Let $q : \mathsf{D} \to \mathbb{R}^k$ be a query. Then the Laplace mechanism $M_q : \mathsf{D} \times \Omega \to \mathbb{R}^k$ defined by $M_q(d) = q(d) + w$, with $w \sim Lap(b)^k$ and $b \geq \frac{\Delta_1 q}{\epsilon}$ is ϵ-differentially private. 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 $\kappa(\delta, \epsilon) = \frac{1}{2\epsilon}(K + \sqrt{K^2 + 2\epsilon})$. Note that it can be shown that $\kappa(\delta, \epsilon)$ can be bounded by $2\sqrt{2\ln(2/\delta)}/\epsilon$.

2.2.1 A Sign-Preserving Mechanism Adapted to Relative Variations

To conclude this section, we introduce a multiplicative perturbation mechanism, which can be useful to handle relative variations as in (3), and also to maintain sign consistency between an original real-valued query and the result provided by the mechanism. This last feature in particular is used in Section 3.1 to maintain the stability of a dynamic model while perturbing its poles.

Consider a query $q : \mathsf{D} \to \mathbb{R}_{>0}$. We would like to design a differentially private mechanism that preserves the positivity of the output. We have, for two datasets d, d'

$$\begin{aligned} |\ln q(d') - \ln q(d)| &= \left| \ln \left(\frac{q(d')}{q(d)} \right) \right| \\ &\leq \frac{|q(d') - q(d)|}{|q(d)|} \\ &\leq \frac{|q(d') - q(d)|}{\min\{|q(d)|, |q(d')|\}}, \end{aligned}$$

using the fact that $\ln x \leq x - 1$ for all $x > 0$. Suppose now that there exists γ such that for all $d, d' \in \mathsf{D}$,

$$\frac{|q(d') - q(d)|}{\min\{|q(d)|, |q(d')|\}} \leq \gamma \frac{\|d' - d\|}{\min\{\|d\|, \|d'\|\}}. \tag{5}$$

In particular, for d, d' such that $\mathrm{Adj}_r^\eta(d, d')$, we have $\|d' - d\| = \|d_i - d_i'\|_i$ for some i, and moreover $\|d\| \geq \|d_i\|_i$. Hence finally

$$|\ln q(d') - \ln q(d)| \leq \gamma \max_{1 \leq i \leq n} \eta_i =: \Gamma.$$

From Theorem 2, the mechanism $M_1(q) = \ln q(d) + Y$, with $Y \sim \mathrm{Lap}\left(\frac{\Gamma}{\epsilon}\right)$, is ϵ-differentially private, and is (ϵ, δ)-differentially private if $Y \sim \mathcal{N}(0, \kappa(\delta, \epsilon)^2 \Gamma^2)$. Taking exponentials and using Theorem 1, we obtain the following result, where $\lambda = \exp(Y)$.

THEOREM 3. *Let $q : \mathsf{D} \to \mathbb{R}_{>0}$ be a query satisfying (5). Then the mechanism $M(d) = \lambda q(d)$ is ϵ-differentially private for (3) if $\lambda \sim$ ln-Lap(Γ/ϵ) is a log-Laplace random variable, and is (ϵ, δ)-differentially private if $\lambda \sim$ ln-$\mathcal{N}(0, \kappa(\delta, \epsilon)^2 \Gamma^2)$ is a log-normal random variable.*

Bounds of the form (5) are frequent in numerical analysis [19]. In the following however, we use this mechanism in its most simple form, with $D = \mathbb{R}_{>0}$ and $q = \mathrm{id}$. Theorem 3 shows in particular that if the data of user i consists of a single positive number x_i, letting each user release $\lambda_i x_i$ with $\lambda_i \sim$ ln-Lap(η_i/ϵ) or $\lambda_i \sim$ ln-$\mathcal{N}(0, \kappa(\delta, \epsilon)^2 \eta_i^2)$ guarantees ϵ- or (ϵ, δ)-differential privacy for (3) respectively.

3. DYNAMIC MODEL PUBLICATION

In the rest of this paper, we consider the following scenario. A group of n users responds to a common scalar input signal $u : \mathbb{R}_+ \to \mathbb{R}$, according to their own dynamics described by a stable scalar first order differential equation

$$\dot{x}_i = -a_i x_i + b_i u, \;\; x_i(0) = x_{0,i} \in \mathbb{R}, \;\; a_i > 0, \forall i \in [n],$$

where $[n] := \{1, \ldots, n\}$. The group produces a measured aggregate scalar output signal $y : \mathbb{R}_+ \to \mathbb{R}$, which is a linear combination of the n individual states

$$y = c^T x,$$

where $c \in \mathbb{R}^n$ is a known vector. For example, u could correspond to a price signal, the individual states to deviations with respect to a nominal consumption level (with $b_i \leq 0$), which react to price changes with some inertia, and we could measure the total consumption, i.e., $c = \mathbf{1}_n$.

Each individual user is willing to provide his parameters a_i, b_i to a data aggregator, which then publishes a version $\hat{G}$ of the single input single output (SISO) system $G(s) = c^T(sI - A)^{-1}b$, describing the relationship between the common input and the aggregate output. Once released, this model can then be used by anyone to predict how the global system responds to a given input u. In particular, an adversary with access to the published model $\hat{G}$ is allowed to use any test input u and read the corresponding output y to try to estimate the parameters a_i, b_i of any specific participant. This adversary could have access to arbitrary side information for this purpose, e.g., he could know the parameters a_j, b_j of all the participants except i.

In order to encourage participants to provide their parameters to the data aggregator, we wish to make the mechanism releasing the global system model (ϵ, δ)-differentially private, for the following adjacency relation on $\mathsf{D} = \mathbb{R}_{>0}^n \times \mathbb{R}^n$

$$\mathrm{Adj}^{\eta,\rho}((a, b), (\hat{a}, \hat{b})) \text{ iff } \mathrm{Adj}_r^\eta(a, \hat{a}) \text{ and } \mathrm{Adj}_a^\rho(b, \hat{b}), \tag{6}$$

where $\eta, \rho \in \mathbb{R}_{>0}^n$ and Adj_r^η and Adj_a^ρ are defined in (3) and (2) respectively. Note that we protect relative variations for the location of the individual poles, which is more meaningful than absolute variations, e.g., due to the strong influence of the distance of the poles to the imaginary axis on the dynamics of the system.

The transfer function of the overall system is

$$G(s) = \sum_{i=1}^n \frac{c_i b_i}{s + a_i}. \tag{7}$$

Its order is bounded by n, the number of participants, which is assumed to be large. The data for the participants such that $c_i = 0$ can immediately be discarded, so without loss of generality, we can assume $c_i \neq 0, \forall i$. Up to a linear state transformation by $\frac{1}{n}\mathrm{diag}(c_1^{-1}, \ldots, c_n^{-1})$, we can moreover assume from now on that $c_i = 1/n, \forall i$, which will simplify the notation. Here the normalization factor $1/n$ is chosen to study the performance of the mechanisms more conveniently as n grows.

3.1 Parameter Perturbation

The first mechanism (generally called an input perturbation mechanism in the literature on differential privacy), consists in letting each participant perturb its own data to satisfy the differential privacy property directly, without relying on a central server. A mechanism releasing the scalars

$$\hat{a}_i = \lambda_i a_i, \;\; \hat{b}_i = b_i + \mu_i, \;\; 1 \leq i \leq n,$$

with μ_i, λ_i chosen as in Theorems 2 and 3, is 2ϵ- or $(2\epsilon, 2\delta)$-differentially private respectively, depending if Laplace or Gaussian random variables are used. According to Theorem 1, we can then publish the model

$$\hat{G}(s) = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{b}_i}{s + \hat{a}_i} \tag{8}$$

to achieve 2ϵ- or $(2\epsilon, 2\delta)$-differentially privacy.

This approach however can add an excessive amount of noise in some common situations. As an illustration, suppose that all users have the same transfer function $1/(s+1)$. Then $G(s) = 1/(s+1)$. Now, even if only the parameters b_i were to be protected, thus allowing us to leave $\hat{a}_i = 1$ for all i in (8), we would get

$$G(s) - \hat{G}(s) = \frac{1}{s+1} \frac{\sum_{i=1}^{n} \mu_i}{n},$$

i.e., the mean squared error (MSE) of this mechanism scales as $1/n$ at low frequencies, whereas one can reasonably hope to have schemes with an MSE approaching an $1/n^2$ scaling in this case. Another issue in this case is that despite the fact that the original system G is really of order 1, the model with perturbed parameters $\hat{a}_i$ remains of large order $n = 100$ with probability 1.

3.2 Impulse Response Perturbation

The next two schemes try to reduce of amount of privacy-preserving noise by taking advantage of the fact that a single term in (7) changes between two vectors of parameters adjacent according to (6). In this subsection, we produce a differentially private version of the impulse response of the system, and use it to rebuild a differentially private version of G, using again Theorem 1. First, we sample the systems with period h (using a zero-order hold), to obtain the difference equation

$$x_{i,k+1} = \alpha_i x_{i,k} + \beta_i u_k,$$
$$\text{with } \alpha_i = e^{-a_i h}, \;\; \beta_i = \int_0^h e^{-a_i \tau} d\tau \; b_i = \frac{1 - e^{-a_i h}}{a_i} b_i.$$

We then aim at releasing the impulse response or Markov parameters for the discrete-time system

$$v_0 = 0, \;\; v_k = \frac{1}{n} \sum_{i=1}^{n} \beta_i \alpha_i^{k-1}, k \geq 1.$$

Suppose that we wish to publish the first N non-trivial parameters $v = [v_1, \ldots, v_N]^T$. The Laplace mechanism asks that we add to each component v_k a random variable Y_k distributed according to $\text{Lap}(\Delta_1/\epsilon)$, where Δ_1 is the ℓ_1 sensitivity

$$\Delta_1 = \max_{\text{Adj}^{\eta,\rho}((a,b),(\hat{a},\hat{b}))} \|v - \hat{v}\|_1.$$

To compute this sensitivity, consider two adjacent parameter vectors (a, b) and $(\hat{a}, \hat{b})$, differing say by the data of the i^{th} participant. Then

$$\|v - \hat{v}\|_1 = \frac{1}{n} \sum_{k=0}^{N-1} |\beta_i \alpha_i^k - \hat{\beta}_i \hat{\alpha}_i^k|. \tag{9}$$

Since this quantity scales as $1/n$ (for N fixed), the MSE between the original impulse response and the perturbed one scales as $1/n^2$, which is an improvement over the scaling in Section 3.1. However, we still need to reconstruct a system approximation from the perturbed Markov parameters, which can be very sensitive to the presence of noise and therefore can cancel the benefits of the scaling. To pursue the computation of the sensitivity, we now make the following additional assumption on the location of the parameters of the system.

Assumption 1. There are publicly known scalars $\kappa_a, \kappa_b > 0$ such that $a_i \geq \kappa_a$ and $|b_i| \leq \kappa_b$, for all $i \in [n]$.

The following proposition, whose proof can be found in the Appendix, bounds the ℓ_1 sensitivity. For the parameters η, ρ appearing in the (3), we define $\eta_m = \max_{i \in [n]} \eta_i, \rho_m = \max_{i \in [n]} \rho_i$.

PROPOSITION 1. *Under Assumption 1, we have*

$$\Delta_1 \leq \frac{h}{n} \Bigg(\frac{(0.3\,\eta_m \kappa_b + \rho_m)(1 - e^{-N\kappa_a h})}{1 - e^{-\kappa_a h}} + \frac{(0.37\,\eta_m\,\kappa_b)(1 + (N-1)e^{-N\kappa_a h} - Ne^{-(N-1)\kappa_a h})}{(1 - e^{-\kappa_a h})^2} \Bigg). \tag{10}$$

Let $\bar{\Delta}_1$ denote the right-hand side of (10). Algorithm 1 summarizes our mechanism for publishing a differentially private version of the global dynamical system. From a perturbed sequence of N Markov parameters, we reconstruct an approximate version of G, using the MATLAB function `imp2ss`, which implements a model realization algorithm using an impulse response, proposed initially by Kung [12]. A discrete-time model is constructed first, and the sampling period h is then used to reconstruct a continuous-time system using the inverse Tustin transform [1]. Note finally that the mechanism is ϵ-differentially private, but the computations in the Appendix can be used to obtain the ℓ_2 sensitivity and design an (ϵ, δ)-differentially private mechanism.

Algorithm 1 Dynamic Model Publication via Approximate Realization

Require: h, sampling period; N, number of Markov parameters
Generate $\nu_i \sim \text{Lap}(\bar{\Delta}_1/\epsilon), \; i \in [n]$
$y_i \leftarrow v_i + \nu_i, \; i \in [n]$
$\hat{G} \leftarrow \texttt{imp2ss}(y, h)$

3.3 Frequency Response Perturbation

Instead of releasing the impulse response, it is perhaps more intuitive to release a set of N samples of the transfer function G, measured at a set of a priori fixed frequencies $\omega_1, \ldots, \omega_N$, which can moreover be chosen in the frequency range over which we wish to better approximate G. Let

$$f = [G(j\omega_1), \ldots G(j\omega_N)]$$

be the vector in $\mathbb{C}^N$ to be released. Equivalently, we want to release the real and imaginary parts of $f = [f_R, f_I]$, where

$$f_R = [\mathfrak{Re}(G(j\omega_1)), \dots \mathfrak{Re}(G(j\omega_N))],$$
$$f_I = [\mathfrak{Im}(G(j\omega_1)), \dots \mathfrak{Im}(G(j\omega_N))].$$

To compute the ℓ_2 sensitivity, consider the variation of the parameters of say participant i. We then compute the change in 2-norm

$$\begin{aligned}\Delta_{2,i}^2 &= \sum_{k=1}^{N} [\mathfrak{Re}(G(j\omega_k) - \hat{G}(j\omega_k))]^2 + [\mathfrak{Im}(G(j\omega_k) - \hat{G}(j\omega_k))]^2 \\ &= \sum_{k=1}^{N} |G(j\omega_k) - \hat{G}(j\omega_k)|^2 \\ &= \frac{1}{n^2} \sum_{k=1}^{N} \left| \frac{b_i}{j\omega_k + a_i} - \frac{\hat{b}_i}{j\omega_k + \hat{a}_i} \right|^2 . \end{aligned} \tag{11}$$

We now analyze the terms of the sum (11).

$$\begin{aligned}&\left| \frac{b_i}{j\omega_k + a_i} - \frac{\hat{b}_i}{j\omega_k + \hat{a}_i} \right|^2 = \frac{(b_i\hat{a}_i - \hat{b}_i a_i)^2 + \omega_k^2 (b_i - \hat{b}_i)^2}{(a_i^2 + \omega_k^2)(\hat{a}_i^2 + \omega_k^2)} \\ &= \frac{(b_i(\hat{a}_i - a_i) + (b_i - \hat{b}_i)a_i)^2 + \omega_k^2 (b_i - \hat{b}_i)^2}{(a_i^2 + \omega_k^2)(\hat{a}_i^2 + \omega_k^2)} \\ &\le \frac{2\kappa_b^2 (\hat{a}_i - a_i)^2 + 2\rho_i^2 a_i^2 + \omega_k^2 \rho_i^2}{(a_i^2 + \omega_k^2)(\hat{a}_i^2 + \omega_k^2)} \\ &\le \frac{2a_i^2(\kappa_b^2 \eta_i^2 + \rho_i^2) + \omega_k^2 \rho_i^2}{(a_i^2 + \omega_k^2)(\hat{a}_i^2 + \omega_k^2)}.\end{aligned}$$

Now, using $\frac{a_i^2}{a_i^2 + \omega_k^2} \le 1$, we get the bound

$$\Delta_2^2 \le \frac{1}{n^2} \sum_{k=1}^{N} \frac{2(\kappa_b^2 \eta_m^2 + \rho_m^2)}{\kappa_a^2 + \omega_k^2} + \frac{\omega_k^2 \rho_m^2}{(\kappa_a^2 + \omega_k^2)(\kappa_a^2 + \omega_k^2)}, \tag{12}$$

under Assumption 1. Note in particular that sampling at high frequencies contributes less to the sensitivity bound, or equivalently, it is harder to publish information about low frequencies than about high frequencies while achieving differential privacy.

Algorithm 2 summarizes our mechanism based on frequency response perturbation. Let $\bar{\Delta}_2^2$ denote the right-hand side of (12). We first perturb the coordinates of the vector (f_R, f_I) using additive Gaussian noise with variance proportional to $\bar{\Delta}_2^2$ to achieve (ϵ, δ)-differential privacy. We then estimate a transfer function $\hat{G}$ based on this frequency response data, using the MATLAB function `tfest` [10]. This function requires the user to specify the order `np` of the model to produce, which should not be chosen too large to avoid overfitting the perturbed values of the frequency response. Note that this order should be chosen a priori without taking the form of G into account, otherwise Theorem 1 would not apply.

4. SIMULATIONS

In this section, we discuss in more details the performance of two of the mechanisms previously presented, namely, the parameter perturbation scheme of Subsection 3.1 and the frequency response perturbation scheme of Subsection 3.3. We fix a priori `np` $= 5$ in Algorithm 2, and a vector ω of 20 sampled frequencies logarithmically spaced between 0.1 rad/s and 100 rad/s. We let $n = 100, \epsilon = \ln 3, \delta = 0.05, \eta_i = 0.2, \rho_i = 0.5, \forall i \in [n]$.

Algorithm 2 Dynamic Model Publication via Frequency Response Estimation

Require: ω, vector of frequencies to sample; `np`, number of poles desired in the released model
Generate $\nu_{R,i} \sim \mathcal{N}(0, \kappa(\delta,\epsilon)^2 \bar{\Delta}_2^2),\ i \in [n]$
Generate $\nu_{I,i} \sim \mathcal{N}(0, \kappa(\delta,\epsilon)^2 \bar{\Delta}_2^2),\ i \in [n]$
$\hat{f}_{R,i} \leftarrow f_{R,i} + \nu_{R,i},\ i \in [n]$
$\hat{f}_{I,i} \leftarrow f_{I,i} + \nu_{I,i},\ i \in [n]$
$\hat{G} \leftarrow$ `tfest`$(\hat{f}_R + i\hat{f}_I, \omega,$ `np`$)$

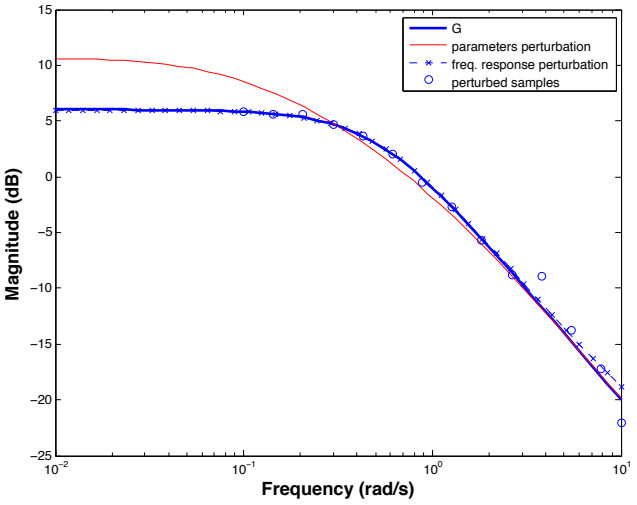

Figure 1: Examples of results produced by the mechanisms based on parameter perturbation and on reconstruction from noisy samples of the frequency response. The samples produced and used by `tfest` in Algorithm 2 are denoted with circles.

The first example is similar to the case described in Subsection 3.1, with each user associated to the same transfer function $1/(s+0.5)$, which is also equal to G. Sample models produced by the parameter perturbation and frequency response perturbation are shown on Fig. 2. For these particular outputs, the errors measured by the H_∞-norm of the difference between the produced and original models are 1.397 for the parameter perturbation scheme and 0.043 for Algorithm 2. In this case the parameter perturbation method produced a model with a large error at low frequencies.

The second example consists of a randomly generated model, where the parameters a_i and b_i are generated independently across users according to uniform distributions on the intervals $[0.5, 5]$, and $[0, 5]$ respectively. Sample outputs of the two mechanisms are reproduced on Fig. 2. In this specific output, we see that the approximation quality of the frequency response perturbation mechanism remains good at low frequencies, but much worse than the parameters perturbation scheme at high frequencies. This is mainly due to the fact that we use additive perturbations in Algorithm 2, which means that the noise tends to dominate the magnitude of the transfer function when the latter is small. Since in this case however the errors occur when the attenuation is already significant, the H_∞-norm of the model error

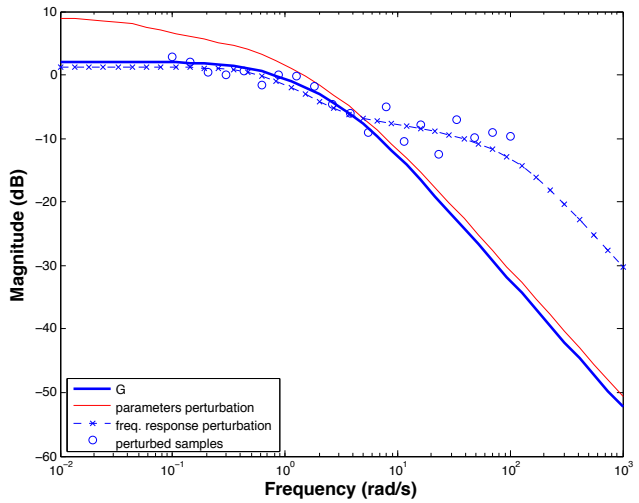

Figure 2: Examples of results produced by the mechanisms for a randomly generated model. We have $\|G - \hat{G}_1\|_\infty = 1.53$ and $\|G - \hat{G}_2\|_\infty = 0.28$, where G_1 is the model produced using parameter perturbation, and G_2 is the model produced via perturbed frequency response data.

for the frequency response perturbation mechanism still remains much smaller than for the parameter perturbation mechanism. The average H_∞ error obtained over 1000 randomly generated systems was found to be 0.97 for the parameter perturbation mechanism and 0.29 for the frequency response perturbation mechanism.

5. CONCLUSION

We have presented several privacy-preserving mechanisms that can be used to release a model describing the dynamics of a large group of users responding to a common input signal and producing an aggregate output signal. We expect the techniques described here to extend to more complex systems, with multiple inputs and outputs and potentially more coupled process dynamics. Future work will focus on analytically quantifying the approximation error achieved by the proposed mechanisms and will explore lower bounds on the model error achievable by differentially private release mechanisms.

Appendix: Proof of Proposition 1

As in (9), let us assume for concreteness that the data of the first participant is perturbed. From the Taylor-Lagrange formula, we have that for all $k \geq 0$,

$$|\hat{\beta}_1\hat{\alpha}_1^k - \beta_1\alpha_1^k| \leq k\tilde{\alpha}_1^{k-1}|\tilde{\beta}_1||\hat{\alpha}_1 - \alpha_1| + \tilde{\alpha}_1^k|\hat{\beta}_1 - \beta_1|,$$

where $\tilde{\alpha}_1 = \alpha_1 + \theta(\hat{\alpha}_1 - \alpha_1), \tilde{\beta}_1 = \beta_1 + \theta(\hat{\beta}_1 - \beta_1)$, for some $\theta \in (0,1)$. In particular,

$$\tilde{\alpha}_1 \leq e^{-\kappa_a h}, \quad |\tilde{\beta}_1| \leq \kappa_b h, \tag{13}$$

where the second inequality comes from the fact that

$$0 < \frac{1 - e^{-x}}{x} \leq 1, \forall x > 0.$$

We now bound the variations in the parameters α_1, β_1. Again, from the Taylor-Lagrange formula,

$$\begin{aligned}|\hat{\alpha}_1 - \alpha_1| = |e^{-\hat{a}_1 h} - e^{-a_1 h}| &\leq \tilde{a}_1 h e^{-\tilde{a}_1 h}\frac{|\hat{a}_1 - a_1|}{\tilde{a}_1} \\ &\leq 0.37\,\eta_m,\end{aligned}$$

where $\tilde{a}_1 = a_1 + \theta(\hat{a}_1 - a_1)$ for some $\theta \in (0,1)$, and we used the fact that $0 < xe^{-x} < 0.37$, for all $x > 0$. Finally,

$$\begin{aligned}|\hat{\beta}_1 - \beta_1| &= \left|\frac{1 - e^{-\hat{a}_1 h}}{\hat{a}_1 h}\hat{b}_1 h - \frac{1 - e^{-a_1 h}}{a_1 h}b_1 h\right| \\ &\leq \left|\frac{-1 + e^{-\tilde{a}_1 h} + \tilde{a}_1 h e^{-\tilde{a}_1 h}}{\tilde{a}_1^2 h^2}\right| |\tilde{b}_1|\, h\, |\hat{a}_1 h - a_1 h| \\ &\quad + \frac{1 - e^{-\tilde{a}_1 h}}{\tilde{a}_1 h} h|\hat{b}_1 - b_1| \\ &\leq 0.3\,\eta_m\,\kappa_b\,h + \rho_m h,\end{aligned}$$

where $\tilde{\alpha}_1 = \alpha_1 + \theta(\hat{\alpha}_1 - \alpha_1), \tilde{\beta}_1 = \beta_1 + \theta(\hat{\beta}_1 - \beta_1)$, for some $\theta \in (0,1)$, and we used the fact that $|(-1 + e^{-x} + xe^{-x})/x| < 0.3$, for all $x > 0$.

Hence overall we get the bounds, for all $k \geq 0$,

$$|\hat{\beta}_1\hat{\alpha}_1^k - \beta_1\alpha_1^k| \leq h(0.37\,k\,\tilde{\alpha}_1^{k-1}\,\kappa_b\,\eta_m + \tilde{\alpha}_1^k(0.3\,\eta_m\,\kappa_b + \rho_m)),$$

and the result of the proposition follows from the formulas

$$\sum_{k=0}^{N-1} k\alpha^{k-1} = \frac{1 + (N-1)\alpha^N - N\alpha^{N-1}}{(1-\alpha)^2}, \quad \sum_{k=0}^{N-1}\alpha^k = \frac{1-\alpha^N}{1-\alpha}.$$

as well as (13).

6. REFERENCES

[1] K. J. Åström and B. Wittenmark. *Computer-Controlled Systems: Theory and Design.* Prentice Hall, 3rd edition, 1997.

[2] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: the SuLQ framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS)*, pages 128–138, New York, NY, USA, 2005.

[3] T.-H. H. Chan, E. Shi, and D. Song. Private and continual release of statistics. *ACM Transactions on Information and System Security*, 14(3):26:1–26:24, November 2011.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, New York, NY, 1991.

[5] G. Duncan and D. Lambert. Disclosure-limited data dissemination. *Journal of the American Statistical Association*, 81(393):10–28, March 1986.

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

[7] 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*, pages 486–503, 2006.

[8] 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*, pages 265–284, 2006.

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

[10] H. Garnier, M. Mensler, and A. Richard. Continuous-time model identification from sampled data: Implementation issues and performance evaluation. *International Journal of Control*, 76(13):1337–1357, 2003.

[11] 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*, 11(5), May 2012.

[12] S. Kung. A new identification and model reduction algorithm via singular value decompositions. In *Proceedings of the Twelfth Asilomar Conference on Circuits, Systems and Computers*, pages 705–714., November 1978.

[13] J. Le Ny and G. J. Pappas. Differentially private filtering. September 2012. Conditionally accepted for publication in the IEEE Transactions on Automatic Control, available at http://arxiv.org/abs/1207.4305.

[14] J. Le Ny and G. J. Pappas. Differentially private Kalman filtering. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, October 2012.

[15] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin. Private memoirs of a smart meter. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pages 61–66, New York, NY, USA, 2010.

[16] B. Reis, C. Kirby, L. Hadden, K. Olson, A. McMurry, J. Daniel, and K. Mandl. Aegis: a robust and scalable real-time public health surveillance system. *Journal of the American Medical Informatics Association*, 14(5):581–588, 2007.

[17] M. Roozbehani, M. Dahleh, and S. Mitter. Dynamic pricing and stabilization of supply and demand in modern electric power grids. In *IEEE International Conference onf Smart Grid Communications*, October 2010.

[18] L. Sankar, S. R. Rajagopalan, and H. V. Poor. A theory of privacy and utility in databases. Technical report, Princeton University, February 2011.

[19] L. N. Trefethen and D. Bau, III. *Numerical Linear Algebra*. SIAM, 1997.