Chapter 4
Statistical Parameter Privacy

Germán Bassi, Ehsan Nekouei, Mikael Skoglund and Karl H. Johansson

Abstract We investigate the problem of sharing the outcomes of a parametric source with an untrusted party while ensuring the privacy of the parameters. We propose privacy mechanisms which guarantee parameter privacy under both Bayesian statistical as well as information-theoretic privacy measures. The properties of the proposed mechanisms are investigated and the utility-privacy trade-off is analyzed.

4.1 Introduction

As the costs of digitizing, storing, and analyzing real-world data constantly decrease, more and more parts of our lives are increasingly being done through digital means. Once this information leaves our control, it can be duplicated and inspected at will. The data might be collected without our explicit knowledge and consent, e.g., our online behavior which is used to personalize the ads with see, or we might actively seek to share the information, e.g., by publishing our movie preferences on social media. In the latter case, there is usually a benefit or utility to be gained by sharing data. However, revealing sensitive information might be undesired for many people even if there is some gain involved. Moreover, if the benefit obtained is directly related to the fidelity of the shared information, a natural trade-off between utility and privacy arises.

Over the past two decades, there has been a surge of research on the problem of utility versus privacy and the design of privacy mechanisms to safely share data.

G. Bassi · M. Skoglund · K. H. Johansson
KTH Royal Institute of Technology, Stockholm, Sweden
e-mail: germanb@kth.se

M. Skoglund
e-mail: skoglund@kth.se

K. H. Johansson
e-mail: kallej@kth.se

E. Nekouei (✉)
City University of Hong Kong, Kowloon Tong, Hong Kong
e-mail: enekouei@cityu.edu.hk

F. Farokhi (ed.), *Privacy in Dynamical Systems*,
https://doi.org/10.1007/978-981-15-0493-8_4

Two of the most well-known approaches for providing privacy in databases are k-anonymity [20] and differential privacy [11]. These strategies are normally implemented when data from different users are employed to estimate statistical population parameters; it is desired that the estimated values are close to the true parameters while the amount of information revealed about any particular user is low. This goal is achieved by suppressing values in the database, by clustering similar values, or by distorting the values with noise. As an example, a group of users participating in a medical survey might have their names removed from the database, their ages assigned to specific age-groups, and their weights modified by the addition of a zero-mean random variable. The effect in the inference performance of such privacy mechanisms has been addressed more recently; for instance, minimax risk bounds and minimax optimal estimation procedures for several canonical families of problems are studied in [3, 10] (see also the references therein).

A complementary problem to the one just described appears when the shared data should closely mirror the real one but must prevent an observer from learning some specific patterns or sensitive statistics of the raw data. This is commonly the case for users who share a stream of data with a third party in order to obtain a service; the more faithful the data is, the better the service provided but the easier the analysis of hidden information. As an example, a user may submit a scanned document to an online service with the goal of performing optical character recognition, however, the user might not want the online service to infer any personality traits in the handwriting or the author's identity with respect to previously submitted documents. In this work, we focus on this second type of problems.

To the best of our knowledge, this is a less explored direction of research in privacy. The reason might be strictly technological since it was not until recently that large amounts of data from a single user could be collected; on the other hand, tiny bits of information from vast numbers of users have been compiled in databases for many years now. The design of this type of privacy filter is also inherently more complex since each entry in the sequence cannot be processed independently and the statistics of the whole sequence must be taken into account. Recently, the authors in [12] define a privacy-preserving strategy that minimizes the Fisher information about the private parameters in the released data. The Cramér–Rao lower bound [9, Theorem 11.10.1] establishes that this strategy maximizes a lower bound on the mean square error of any unbiased estimator of the parameters. The interested reader is referred to [5] for a broader study on the performance degradation of *any* parameter estimator due to a privacy filter. In particular, the authors investigate the relationship between the length n of the data sequence and the inference performance of the parameters. The main result from this work, a privacy filter that hinders the estimation of sensitive parameters, is reproduced in the present manuscript.

Several different works have studied the general relationship between utility and privacy. The authors of [6] introduce a general framework of utility versus privacy where the former is defined as a bounded distortion and the latter as a log-loss cost, which yields a trade-off similar to the rate-distortion function. If both utility and privacy are measured using entropy or mutual information, a different utility-privacy region is provided in [7]; however, only the two extreme cases of perfect privacy and perfect utility are properly characterized therein. The concept of maximal correlation as a measure of privacy is introduced in [1] and it is shown in [2] that this measure is equivalent to maximizing the MMSE of the private data given the shared data. The authors of [13] also define a problem similar to rate-distortion, where privacy is determined by mutual information, and they characterize optimal asymptotic leakages for i.i.d. and general privacy mechanisms. Finally, the use of a secret key to hinder the success of an eavesdropper is addressed in [23]; the authors argue for the use of distortion-based privacy metrics instead of stronger information-theoretic ones to reduce the size of the secret key.

Information privacy and the design of privacy filters have also been studied in dynamic settings. The authors of [22] study the design of privacy-preserving filters in a cloud-based control problem using the notion of directed information as the privacy metric. Le Ny and Pappas in [15] propose privacy-preserving filtering algorithms for ensuring the privacy of states or measurements of dynamical systems, based on differential privacy. The authors of [21] study the state estimation problem in a distribution power network subject to differential privacy constraints for the consumers. Wang et al. in [24] propose privacy-preserving mechanisms for ensuring the privacy of the initial states and the preferred target way-points in a distributed multi-agent control system. Privacy-preserving average consensus algorithms, for preserving the privacy of initial states, are addressed in [17, 18].

4.1.1 Organization

After the comprehensive introduction into the problem of parameter privacy, the rest of this chapter is devoted to the analysis of two different privacy-preserving filters. To facilitate the reading, we have organized the work as follows.

In Sect. 4.2, we present the system model for the problem of parameter privacy and some important definitions. In particular, we link the performance of the parameter estimation to the mutual information between the parameter and the released data. The section ends with the overview of the two different privacy filters described in this work. The first of these filters is introduced in Sect. 4.3. An achievable scheme that distorts the shared data is outlined; the proposed privacy mechanism seeks to confuse the adversary by introducing an auxiliary parameter that behaves as the true one. It is shown that the filter limits the amount of information released to the eavesdropper. A Gaussian example is used to illustrate the trade-off between distortion and privacy. In Sect. 4.4, the structure of the second privacy filter is studied. In this scheme, the privacy filter design problem is posed as a convex optimization problem which achieves the Pareto boundary of the distortion-privacy region. An upper bound on the leakage of private information under this scheme is obtained. The distortion-privacy trade-off for this filter is studied with a numerical example.

4.1.2 Notation

In the rest of this chapter, lowercase letters such as x and y are mainly used to represent constants or realizations of random variables, capital letters such as X and Y stand for the random variables in itself, and calligraphic letters such as $\mathcal{X}$ and $\mathcal{Y}$ are reserved for sets. In the case of Greek letters, we use Θ, θ, and $\mathbf{\Theta}$ to denote a random variable, its realization, and its support set, respectively.

We use X^n to denote the sequence of independent and identically distributed (i.i.d.) random variables $\{X_k\}_{k=1}^n$. Given three random variables X, Y, and Z, if its joint probability distribution can be decomposed as $p(xyz) = p(x)p(y|x)p(z|y)$, then they form a Markov chain, denoted by $X \to Y \to Z$.

Entropy is denoted by $H(\cdot)$ and mutual information, $I(\cdot;\cdot)$. Throughout the work and unless stated otherwise, log refers to logarithm in base 2.

4.2 System Model and Overview of Results

In this section, we first introduce the general model studied in this work and some useful definitions. We then present an overview of results for two particular privacy-preserving filters.

4.2.1 System Model

Consider the three-user problem depicted in Fig. 4.1, where Alice wants to share with Bob the outcomes of a random parametric source she observes. The value of the parameter, which in this work constitutes private information that Alice does not want to disclose, might even be unknown to her. For example, the observation might be a handwritten note (a sequence of characters) by Alice while the parameter represents her personality traits.

In the absence of any constraint on the rate of information between the users, Alice may choose to directly send the observed sequence of values. However, the communication is overheard by Eve, who is interested in characterizing the statistical properties of the random parametric source, i.e., estimate the parameter. In order to protect her privacy, Alice needs to share a distorted version of the source, but one

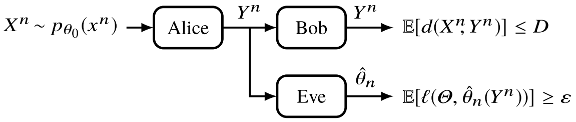
$X^n \sim p_{\theta_0}(x^n)$
Alice
Y^n
Bob
Y^n
$\mathbb{E}[d(X^n, Y^n)] \leq D$
Eve
$\hat{\theta}_n$
$\mathbb{E}[\ell(\Theta, \hat{\theta}_n(Y^n))] \geq \varepsilon$

Fig. 4.1 General system model

that it is still useful for Bob. In our previous example with the handwritten note, the font style may change as long as Bob is able to correctly read the text.

More precisely, we assume that Alice observes n samples of the random variable $X \in \mathcal{X}$ where the samples are i.i.d. with respect to the distribution P_{θ_0}. The probability measure P_{θ_0} is a member of a parameterized family of distributions $\mathcal{P}_\Theta = \{P_\theta : \theta \in \boldsymbol{\Theta}\}$ on a measurable space, where θ_0 is a point in the interior of $\boldsymbol{\Theta}$. Moreover, $p_{\theta_0}(x)$ is the probability density function (PDF) of P_{θ_0} with respect to a fixed σ-finite measure $\mu(dx)$; it is assumed that $p_{\theta_0}(x)$ is non-zero almost everywhere on $\mathcal{X}$ and the corresponding probability measure, P_{θ_0}, is absolutely continuous with respect to Lebesgue measure.

The value of the parameter is chosen randomly by nature according to the known prior distribution $p(\theta)$ with respect to Lebesgue measure; thus, the parameter is regarded as a random variable, which we denote Θ.[1] As previously mentioned, Alice produces a distorted sequence Y^n which is based on the observed sequence X^n and shares it with Bob. The channel between Alice and Bob has no rate limitation, and the purpose of distorting the sequence is to prevent Eve from increasing her knowledge about the unknown parameter Θ beyond what is specified by the prior distribution.

We further assume that Bob has no advantage over Eve. The communication is received by both users with the same level of quality and they are all aware of the strategy employed by Alice to distort the observed sequence. If a certain stochastic transformation is used to increase the privacy, the particular realization of the mapping is unknown to both Bob and Eve.

4.2.2 Useful Definitions and Preliminary Result

We present here some definitions needed to characterize the loss in the fidelity of the sequence Y^n with respect to X^n and the increase in privacy. We start with some general notions.

Definition 4.1 Let X denote an absolutely continuous random variable with probability density function $p_X(x)$. Then, the *differential entropy* of X is defined as

$$h(X) = -\int_{\mathcal{X}} p_X(x) \log p_X(x)\, dx. \tag{4.1}$$

Definition 4.2 Consider the random variable X distributed according to the probability density function $p_\theta(x)$, where θ is a parameter taking values in $\mathbb{R}^d$. Then, the *Fisher information matrix* about θ contained in $X|_{\Theta=\theta}$ is defined as a $d \times d$ matrix with the (i, j)th entry given by

[1] Note that Θ stands for the parameter taken as a random variable, whereas $\boldsymbol{\Theta}$ corresponds to the parameter space.

$$\left[\mathbf{I}_X(\theta)\right]_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log p_\theta(x)\right)\left(\frac{\partial}{\partial\theta_j}\log p_\theta(x)\right)\,\middle|\, \Theta = \theta\right] \quad 1 \le i, j \le d. \tag{4.2}$$

Definition 4.3 The *distortion* between the sequences x^n and y^n is defined as

$$d(x^n, y^n) \triangleq \frac{1}{n}\sum\nolimits_{i=1}^{n} d(x_i, y_i)\,, \tag{4.3}$$

where the distortion function d is a mapping $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$.

Definition 4.4 The *privacy* (distortion) between the parameter θ and an estimate $\tilde{\theta}$, i.e., $\ell(\theta, \tilde{\theta})$, is given by the mapping $\ell : \boldsymbol{\Theta} \times \boldsymbol{\Theta} \to \mathbb{R}^+$, where we assume that $\inf_{\tilde{\theta}} \ell(\theta, \tilde{\theta}) = 0$ for all θ.

Definition 4.5 A distortion-privacy pair (D, ϵ) is *achievable* in this problem if there exists $N > 0$ and a privacy-preserving mapping (stochastic kernel) $f_n : \mathcal{X}^n \to \mathcal{Y}^n$ such that

$$\mathbb{E}[d(X^n, Y^n)] \le D, \tag{4.4}$$

$$\inf_{\hat{\theta}_n} \mathbb{E}[\ell(\Theta, \hat{\theta}_n(Y^n))] \ge \epsilon\,, \tag{4.5}$$

for $n > N$, where $Y^n = f_n(X^n)$ and the infimum is taken over all measurable functions $\hat{\theta}_n : \mathcal{Y}^n \to \boldsymbol{\Theta}$ that are possible estimators of the parameter Θ.

An important information-theoretic function is the rate-distortion (RD) function. We introduce it here for completeness.

Definition 4.6 ([9, Sect. 10.2]) The *(information) rate-distortion function* for a random variable Θ with distortion measure $\ell(\cdot, \cdot)$ is defined as

$$R_{\Theta,\ell}(D) \triangleq \min_{p(\tilde{\theta}|\theta):\, \mathbb{E}[\ell(\Theta,\tilde{\Theta})] \le D} I(\Theta; \tilde{\Theta}). \tag{4.6}$$

We assume that there exists $D \ge 0$ such that $R_{\Theta,\ell}(D)$ is finite.

We note that the RD function has the following properties:

- The infimum over $D \ge 0$ such that $R_{\Theta,\ell}(D)$ is finite is denoted $D_{\min}$; the corresponding rate is $R_{\max} \triangleq \lim_{D \to D_{\min}^+} R_{\Theta,\ell}(D)$.
- The RD function $R_{\Theta,\ell}(D)$ is a non-increasing convex function of D on the interval $(D_{\min}, \infty)$. It is monotonically decreasing on the interval $(D_{\min}, D_{\max})$ and constant with $R_{\Theta,\ell}(D) = R_{\min}$ on $[D_{\max}, \infty)$.
- The inverse function $R_{\Theta,\ell}^{-1}(r)$ is well defined on $(R_{\min}, R_{\max})$ and monotonically decreasing. This function is known as the *distortion-rate* (DR) *function*.

Before proceeding with the overview of results in the next subsection, we present an important lemma that relates Eve's performance in estimating the random parameter Θ with the sequence transmitted by Alice. From Eve's point of view, the setting of Fig. 4.1 is a statistical inference problem of a random quantity Θ that cannot be directly observed; only an indirect measurement Y^n is obtained. However, the pair (Θ, Y^n) has a given joint probability distribution, and thus Eve may calculate an estimate $\hat{\theta}_n(Y^n)$ of the parameter Θ [14].

Lemma 4.1 ([5, Lemma 2]) *For* any *estimator* $\hat{\theta}_n$ *and any distortion function* $\ell(\cdot, \cdot)$*, the* expected privacy *(the Bayes risk of the estimator) is bounded from below by:*

$$\mathbb{E}[\ell(\Theta, \hat{\theta}_n)] \geq R^{-1}_{\Theta,\ell}(I(\Theta; Y^n)) , \tag{4.7}$$

where $R^{-1}_{\Theta,\ell}(\cdot)$ *is the DR function of the random variable* Θ*.*

Proof The proof follows from the data-processing inequality [9, Sect. 2.8], the Markov chain

$$\Theta \rightarrow Y^n \rightarrow \hat{\theta}_n(Y^n), \tag{4.8}$$

and the definition of the RD function in (4.6). Please refer to [5] for more details. □

The preceding lemma is quite powerful in that it allows us to bound the performance of *any* estimator $\hat{\theta}_n$ of Θ based on Y^n without knowing how that estimator is calculated. Analytical solutions for the DR function in (4.7), on the other hand, are only known for a handful of random variables and distortion measures. In many situations, we may need to employ the looser Shannon lower bound or compute the RD function numerically using the Blahut–Arimoto algorithm [9, Chap. 10].

Gaussian Example

Assume that Alice observes n i.i.d. samples of the process $X|_{\Theta=\theta} \sim \mathcal{N}(\theta, \sigma_X^2)$, where the mean is fixed throughout the process but it has an unknown value. Additionally, assume that $\Theta \sim \mathcal{N}(0, \sigma_\Theta^2)$, and consider the square error distortion function for the estimation at Eve, i.e., $\ell(\theta, \hat{\theta}) \triangleq (\theta - \hat{\theta})^2$.

In this setting, the expected distortion is the mean square error, and the DR function is [9, Theorem 10.3.2]

$$R^{-1}_{\Theta,\ell}(r) = \sigma_\Theta^2 2^{-2r}. \tag{4.9}$$

If Alice sends the sequence X^n without any distortion, i.e., $Y = X$, then

$$I(\Theta; X^n) = \frac{1}{2}\log\left(\frac{\sigma_X^2 + n\sigma_\Theta^2}{\sigma_X^2}\right) = \frac{1}{2}\log\frac{n\sigma_\Theta^2}{\sigma_X^2} + o(1) , \tag{4.10}$$

which according to Lemma 4.1 yields

$$\mathrm{MSE}_{\hat{\theta}_n} \geq \frac{\sigma_X^2\sigma_\Theta^2}{\sigma_X^2 + n\sigma_\Theta^2} \simeq \frac{\sigma_X^2}{n} , \tag{4.11}$$

where the approximation is valid for large n such that $\sigma_X^2 \ll n\sigma_\Theta^2$. We see that the lower bound approaches 0 as $n \to \infty$; hence, there might exist an estimator that attains a vanishing mean square error.

We note that in this particular example the general lower bound in Lemma 4.1 is tight. The right-hand side of (4.11) is known to be the minimum mean square error for the considered estimation problem; thus, it can be attained by a specific estimator: the MMSE estimator [14, Sect. 11.4].

We conclude this part with an important remark. Lemma 4.1 states that the expected distortion of the unknown estimator $\hat{\theta}_n$ is bounded from below by a monotonically decreasing function (the DR function) of the mutual information between Θ and Y^n. Consequently, the mutual information $I(\Theta; Y^n)$ should be *minimized* in order to hinder Eve's estimation performance.

4.2.3 Overview of Results

Two different privacy-preserving filters are presented in this work. The following is an overview of their characteristics and performance; both filters are analyzed in more detail in Sects. 4.3 and 4.4.

4.2.3.1 First Scheme

For smooth parametric families $\mathcal{P}_\Theta$ with a *continuous* parameter $\Theta \in \boldsymbol{\Theta} \subset \mathbb{R}^d$, it is a well-known fact that $I(\Theta; X^n) \propto \frac{d}{2}\log n$ (see e.g., [8, 19]). Therefore, without properly distorting X^n, the amount of information about Θ gathered by the eavesdropper increases with n. According to Lemma 4.1, Eve might thus be able to estimate the parameter with arbitrarily high precision.

In Sect. 4.3, we analyze a privacy-preserving filter like the one depicted in Fig. 4.2, where the auxiliary random parameter Φ is added to prevent an accurate estimation of Θ. Eve is only able to estimate a function of the true and auxiliary parameters, i.e., only $\Psi = \psi(\Theta, \Phi)$ may be estimated, where the function $\psi(\cdot)$ depends on the statistics of the source and the filter. Namely, the randomness in the auxiliary parameter acts as noise of a virtual noisy channel for the inference problem: $I(\Theta; Y^n) = I(\Theta; \Psi) + O(1)$. Consequently, we prevent Eve from collecting unlimited amount of information about Θ as n increases.

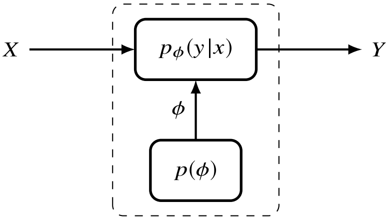

Fig. 4.2 Stochastic filter. Both the mapping $\mathcal{X} \to \mathcal{Y}$ and the choice of the mapping are random

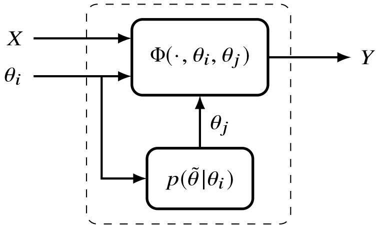
X
θ_i
$\Phi(\cdot, \theta_i, \theta_j)$
Y
θ_j
$p(\tilde{\theta}|\theta_i)$

Fig. 4.3 Deterministic filter for $\Theta = \theta_i$ and $\tilde{\Theta} = \theta_j$. The mapping $\Phi\left(\cdot, \theta_i, \theta_j\right)$ is deterministic but the choice of $\tilde{\Theta}$ is stochastic

4.2.3.2 Second Scheme

Under the second scheme, we assume that $\mathcal{X} = \mathbb{R}^d$ and that the parameter set $\boldsymbol{\Theta}$ has a finite cardinality. In the proposed filter and at each time k, Alice generates the random variable Y_k according to the mapping

$$Y_k = \Phi\left(X_k, \theta, \tilde{\theta}\right), \tag{4.12}$$

where $\tilde{\theta}$ belongs to $\boldsymbol{\Theta}$ and $\Phi\left(\cdot, \theta, \tilde{\theta}\right)$ is a map from $\mathbb{R}^d$ to $\mathbb{R}^d$. The map $\Phi\left(\cdot, \theta, \tilde{\theta}\right)$ is designed such that the common PDF of $\{Y_1, \ldots, Y_n\}$ is equal to $p_{\tilde{\theta}}(x)$. We assume that θ is known by Alice and $\tilde{\theta}$ is selected, by the privacy filter, to simultaneously ensure the privacy of θ and accuracy of the revealed information to Bob. The construction of the map $\Phi\left(\cdot, \theta, \tilde{\theta}\right)$ and the generation of $\tilde{\theta}$ are discussed in Sect. 4.4. Figure 4.3 shows a pictorial representation of the privacy filter under the second scheme.

4.3 First Scheme

In this section, we focus on parameter sets $\boldsymbol{\Theta} \subset \mathbb{R}^d$. The unfavorable result obtained in the Gaussian example in the preceding section, in particular (4.10) and (4.11), is not an isolated case but rather the norm for most (well-behaved) parametric sources. Given that the parameter $\Theta \in \mathbb{R}^d$, i.e., it has an infinite precision, it is expected that the mutual information (4.10) grows unboundedly; hence, an observer is able to estimate the parameter Θ with arbitrarily low error as n increases.

The asymptotic behavior in (4.10) is a special case of a much larger set of results. Specifically, if the density $p_\theta(x)$ satisfies suitable smoothness conditions, it is shown by Clarke and Barron [8] that

$$I(\Theta; X^n) = \frac{d}{2} \log \frac{n}{2\pi e} + h(\Theta) + \frac{1}{2} \mathbb{E}\big[\log |\mathbf{I}_X(\Theta)|\big] + o(1)\,, \tag{4.13}$$

where d is the dimension of the parameter space, $h(\Theta)$ is the differential entropy of the parameter, and $\mathbf{I}_X(\theta)$ is the Fisher information matrix about θ contained in $X|_{\Theta=\theta}$. In the aforementioned Gaussian example, we obtain (4.10) from (4.13) by noting that $d = 1$, $h(\Theta) = \frac{1}{2} \log 2\pi e \sigma_\Theta^2$, and $\mathbf{I}_X(\theta) = 1/\sigma_X^2\ \forall \theta \in \mathbb{R}$. For all the parametric sources where (4.13) holds, e.g., all the exponential families, Eve may attain good inference performance if Alice reveals her observation X^n without distortion.

4.3.1 A Simple Privacy-Preserving Strategy

Let us define the privacy filter as a conditional distribution of Y given X belonging to a parametric family of distributions $\mathcal{P}_\Phi = \{P_\phi : \phi \in \boldsymbol{\Phi}\}$, where $\boldsymbol{\Phi} \subset \mathbb{R}^{d'}$, and where the auxiliary random parameter Φ is distributed according to the prior distribution $p(\phi)$. Therefore, $\{\mathcal{P}_\Phi, p(\phi)\}$ determines the privacy filter.

In a well-designed privacy filter, the auxiliary parameter Φ combines with Θ in a way that the sequence Y^n is consistent with the observation of a parametric family of distributions $\mathcal{P}_\Psi = \{P_\psi : \psi \in \boldsymbol{\Psi}\}$, where $\psi = \psi(\theta, \phi)$ and $\boldsymbol{\Psi} \subset \mathbb{R}^{d'}$. In other words, Ψ is a *sufficient statistic* for Y [9, 16]. Eve may thus be able to estimate ψ with arbitrarily low error as n increases but she has a non-vanishing uncertainty about θ given by the randomness in ϕ.

Theorem 4.1 *If the privacy filter $\{\mathcal{P}_\Phi, p(\phi)\}$ satisfies some suitable smoothness conditions (defined in the proof), it achieves according to Def. 4.5 all distortion-privacy pairs (D, ϵ) such that $\mathbb{E}[d(X^n, Y^n)] \le D$ and $\epsilon \le R^{-1}_{\Theta,\ell}(I(\Theta; Y^n))$, where*

$$I(\Theta; Y^n) = h(\Psi) - h(\Phi) + \frac{1}{2} \mathbb{E}\left[\log \frac{|\mathbf{I}_Y(\Psi)|}{|\mathbf{I}_Y(\Phi)|}\right] + o(1)\,. \tag{4.14}$$

Therefore, the privacy level ϵ remains asymptotically bounded away from zero.

Remark 4.1 Since Φ is independent of Θ, we have that $h(\Phi) = h(\Psi|\Theta)$, and thus (4.14) is equivalently

$$I(\Theta; Y^n) = I(\Theta; \Psi) + O(1). \tag{4.15}$$

In other words, the privacy filter creates a noisy channel for the parameter.

Before presenting the proof of the theorem, we revisit the Gaussian example.

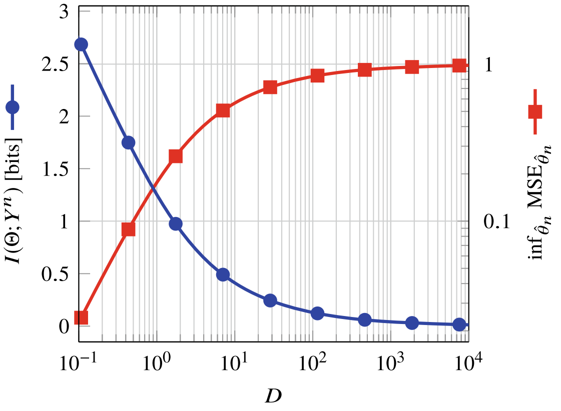

Fig. 4.4 Trade-off between the distortion level D and the corresponding maximum privacy level $\epsilon = \inf_{\hat{\theta}_n} \text{MSE}_{\hat{\theta}_n}$. The relation between the curves for $I(\Theta; Y^n)$ and $\text{MSE}_{\hat{\theta}_n}$ is given by the DR function (4.9). These curves are calculated assuming $\sigma_\Theta^2 = 1$ and $\sigma_Z^2 = 10^{-3}$

Gaussian Example (cont.)

Let us continue with the Gaussian example where we now consider the square error distortion function for the reconstruction at Bob, i.e., $d(x, y) \triangleq (x - y)^2$.

The privacy filter $\{\mathcal{P}_\Phi, p(\phi)\}$ is chosen to mimic the parametric source that Alice tries to protect. In particular, for a fixed $\sigma_Z^2 < D$, the auxiliary parameter Φ is drawn uniformly at random from the interval $[D - \sigma_Z^2, D - \sigma_Z^2]$ and the filter's output is constructed in an i.i.d. manner: for each time $i \in [1 : n]$, $Y_i = X_i + Z_i$, where $Z_i|_{\Phi=\phi} \sim \mathcal{N}(\phi, \sigma_Z^2)$ and is independent of X_i. This choice of filter satisfies $\mathbb{E}[d(X^n, Y^n)] \le D$ and the smoothness conditions needed for Theorem 4.1.

With this privacy-preserving strategy, the eavesdropper observes n i.i.d. samples of the process $Y|_{\Psi=\psi} \sim \mathcal{N}(\psi, \sigma_X^2 + \sigma_Z^2)$, where the mean $\Psi = \Phi + \Theta$ is distributed according to

$$p(\psi) = \frac{1}{2\sqrt{D}} \left[F\left(\frac{\psi + \sqrt{D}}{\sigma_\Theta} \right) - F\left(\frac{\psi - \sqrt{D}}{\sigma_\Theta} \right) \right] \tag{4.16}$$

and $F(\cdot)$ is the cumulative distribution function of the standard normal distribution. We may easily obtain the Fisher information

$$\mathbf{I}_Y(\psi) = \mathbf{I}_Y(\phi) = (\sigma_X^2 + \sigma_Z^2)^{-1}, \tag{4.17}$$

which holds for all values of ϕ and ψ. Therefore, using (4.14) we have that

$$I(\Theta; Y^n) = h(\Psi) - \log 2\sqrt{D} + o(1)\,, \tag{4.18}$$

where $h(\Psi)$ does not have a closed-form expression and has to be calculated numerically using (4.16).

The mutual information (4.18) and the corresponding lower bound from Lemma 4.1 are plotted in Fig. 4.4 for different values of D; we note that the DR function is found in (4.9). The curves show the effect of the privacy-preserving strategy and the trade-off between the distortion level D and the loss in Eve's inference performance.

4.3.2 *Proof of Theorem 4.1*

In order to protect her privacy, Alice needs to distort the observed sequence x^n with some randomness that behaves like the parameter θ from the point of view of Eve.

Given the privacy filter $\{\mathcal{P}_\Phi, p(\phi)\}$, Alice selects a distribution P_ϕ, whose probability density function with respect to a fixed σ-finite measure $\mu(dy)$ is $p_\phi(y|x)$, according to the prior $p(\phi)$. Then, given the original sequence x^n and the specific distribution P_ϕ, she transmits the distorted symbols $Y_i \sim p_\phi(y|x_i)$ for $i \in [1:n]$. Therefore, the joint density of the sequences is:

$$p_{\theta,\phi}(x^n, y^n) = \prod\nolimits_{i=1}^{n} p_\phi(y_i|x_i) p_\theta(x_i) \, . \tag{4.19}$$

Due to the i.i.d. nature of the source and the privacy filter, and conditioned on the true value of the parameters, the sequence observed by Eve is distributed according to

$$p_{\theta,\phi}(y) = \int_{\mathcal{X}} p_\phi(y|x) p_\theta(x) \mu(dx) \, , \tag{4.20}$$

where both θ and ϕ are unknown to her. As previously mentioned, privacy is possible if, in the marginal density (4.20), the auxiliary parameter ϕ combines with θ such that $\psi = \psi(\theta, \phi)$ is a sufficient statistic for the parametric family of distributions $\mathcal{P}_\Psi$.

In this case, we may expand the quantity of interest as follows

$$\begin{aligned} I(\Theta; Y^n) &= I(\Theta, \Phi; Y^n) - I(\Phi; Y^n|\Theta) \\ &= I(\Psi; Y^n) - I(\Phi; Y^n|\Theta) \, , \end{aligned} \tag{4.21}$$

where the last equality is due to Ψ being a sufficient statistic for (θ, ϕ), i.e., the Markov chain $(\Theta, \Phi) \to \Psi \to Y^n$ holds. Assuming all the probability distributions involved in (4.21) satisfy suitable smoothness conditions, we may characterize the asymptotic behavior of both terms similarly to (4.13). In broad terms, these conditions are:

- the densities $p_\theta(x)$ and $p_\phi(y|x)$ are twice continuously differentiable almost everywhere, and the first and second derivatives of $\log p_\theta(x)$ and $\log p_\phi(y|x)$ are square-integrable;
- the priors are continuous and positive almost everywhere;
- the appropriate Fisher information matrices are positive definite; and,
- the posterior distribution of the parameters concentrates around the true value as n increases.

We refer the reader to [5] for more details on these conditions. If these conditions are satisfied, the first term on the right-hand side of (4.21) may be written as

$$I(\Psi; Y^n) = \frac{d'}{2} \log \frac{n}{2\pi e} + h(\Psi) + \frac{1}{2}\,\mathbb{E}\big[\log |\mathbf{I}_Y(\Psi)|\big] + o(1)\,. \tag{4.22}$$

On the other hand, the second term is an expectation on Θ:

$$I(\Phi; Y^n|\Theta) = \int_{\boldsymbol{\Theta}} I(\Phi; Y^n|\theta) p(\theta) d\theta\,, \tag{4.23}$$

where $I(\Phi; Y^n|\theta)$ implies that the parameter is now fixed and known. Then, $I(\Phi; Y^n|\theta)$ is equal to

$$I(\Phi; Y^n|\theta) = \frac{d'}{2} \log \frac{n}{2\pi e} + h(\Phi) + \frac{1}{2}\,\mathbb{E}\big[\log |\mathbf{I}_Y(\Phi)|\big] + o(1)\,. \tag{4.24}$$

Joining these results, we obtain the expression (4.14), which concludes the proof of Theorem 4.1. □

4.4 Second Scheme

For the rest of the chapter, we assume that each random variable $X_k = [X_k^1, \ldots, X_k^d]^\top$ takes values in $\mathbb{R}^d$. Furthermore, X_k^l denotes the lth entry of the random variable X_k while $X_k^{1:l-1}$ denotes the collection of the first $l-1$ entries of X_k. In this section, we also assume that the parameter set $\boldsymbol{\Theta}$ consists of m elements, i.e., $\boldsymbol{\Theta} = \{\theta_1, \ldots, \theta_m\}$.

The conditional cumulative distribution function (CDF) of X^l given $X^{1:l-1} = x^{1:l-1}$ and $\Theta = \theta_i$ is defined as

$$F_{l,\theta_i}\big(z \,\big|x^{1:l-1}\big) = \int_{-\infty}^{z} p_{\theta_i}\big(x \,\big|x^{1:l-1}\big)\, dx\,, \tag{4.25}$$

where $p_{\theta_i}\big(x \,\big|x^{1:l-1}\big)$ is the conditional PDF of X^l given $(X^{1:l-1} = x^{1:l-1}, \Theta = \theta_i)$ which is computed by Bayes' rule and the marginalization of $p_\theta(x)$. We use the convention that $p_{\theta_i}\big(x \,\big|x^{1:0}\big) = p_{\theta_i}(x)$. Note that $F_{l,\theta_i}(\cdot\,|\cdot)$ is a map from $\mathbb{R}^l$ to [0, 1] and is non-decreasing in the first argument when the second argument is fixed. We use $F^{-1}_{l,\theta_i}\big(\cdot \,\big|x^{1:l-1}\big)$ to denote the inverse of the function $F_{l,\theta_i}\big(\cdot \,\big|x^{1:l-1}\big)$ for $1 \le l \le d$.

The second privacy filter comprises the map $\Phi\Big(\cdot, \theta, \tilde{\theta}\Big)$ and a stochastic kernel for generating $\tilde{\theta}$. We first describe the structure of $\Phi\Big(\cdot, \theta, \tilde{\theta}\Big)$. Given $\Theta = \theta_i, \tilde{\Theta} = \theta_j$ and the observation at time k, i.e., $X_k = [X_k^1, \ldots, X_k^d]^\top$, Alice sequentially generates the entries of $Y_k = [Y_k^1, \ldots, Y_k^d]^\top$ as follows. For $1 \le l \le d$, the lth entry of Y_k, i.e., Y_k^l, is generated according to

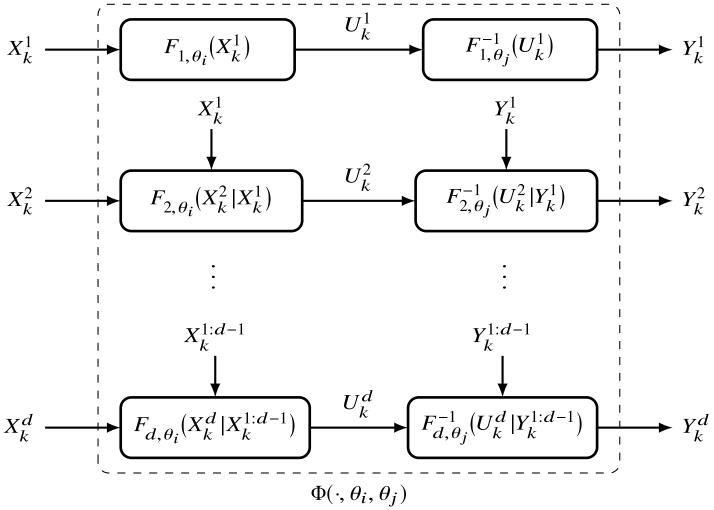
X_k^1
$F_{1,\theta_i}(X_k^1)$
U_k^1
$F^{-1}_{1,\theta_j}(U_k^1)$
Y_k^1
X_k^1
Y_k^1
X_k^2
$F_{2,\theta_i}(X_k^2|X_k^1)$
U_k^2
$F^{-1}_{2,\theta_j}(U_k^2|Y_k^1)$
Y_k^2
$X_k^{1:d-1}$
$Y_k^{1:d-1}$
X_k^d
$F_{d,\theta_i}(X_k^d|X_k^{1:d-1})$
U_k^d
$F^{-1}_{d,\theta_j}(U_k^d|Y_k^{1:d-1})$
Y_k^d
$\Phi(\cdot,\theta_i,\theta_j)$

Fig. 4.5 The structure of $\Phi\left(\cdot,\theta,\tilde{\theta}\right)$ for $\Theta=\theta_i$ and $\tilde{\Theta}=\theta_j$

$$Y_k^l=\phi^l\left(X_k^{1:l},\theta_i,\theta_j\right), \tag{4.26}$$

where $\phi^l\left(X_k^{1:l},\theta_i,\theta_j\right)=F^{-1}_{l,\theta_j}\left(U_k^l\,\middle|Y_k^{1:l-1}\right)$ and $U_k^l=F_{l,\theta_i}\left(X_k^l\,\middle|X_k^{1:l-1}\right)$. The non-linear map $\Phi\left(\cdot,\theta_i,\theta_j\right)$ can be written as $\Phi\left(\cdot,\theta_i,\theta_j\right)=[\phi^1\left(\cdot,\theta_i,\theta_j\right),\ldots,\phi^d\left(\cdot,\theta_i,\theta_j\right)]^\top$. Figure 4.5 shows the structure of $\Phi\left(\cdot,\theta_i,\theta_j\right)$ for $\Theta=\theta_i$ and $\tilde{\Theta}=\theta_j$.

The following lemma studies the statistical properties of the output of this privacy-preserving filter.

Lemma 4.2 *Consider the construction above and assume that $\Theta=\theta_i$ and $\tilde{\Theta}=\theta_j$ for $1\le i,j\le m$. Then, the sequence of random variables $\{Y_k\}_k$ are jointly independent and distributed according to $p_{\theta_j}(x)$.*

Proof See [4]. □

We next discuss the optimal generation of $\tilde{\Theta}$. The parameter $\tilde{\Theta}$ is selected from the set $\boldsymbol{\Theta}$ using a stochastic kernel. More precisely, given $\Theta=\theta_i$, the value of $\tilde{\Theta}$ is generated according to the following stochastic kernel

$$\tilde{\Theta}=\theta_j \;\text{ w.p. }\; P_{ji}=\mathsf{Pr}\left(\tilde{\Theta}=\theta_j\middle|\Theta=\theta_i\right), \tag{4.27}$$

where $\sum_j P_{ji}=1$ for all i and w.p. stands for *with probability*. The randomization probabilities are designed such that the accuracy of the output of the privacy filter is maximized while a certain privacy level for the parameter Θ is achieved.

To discuss the design of the randomization probabilities, we first define the privacy metric as follows. The privacy level of the parameter Θ is captured by the mutual information between the Θ and $\tilde{\Theta}$ which is defined as

$$I(\Theta;\tilde{\Theta}) = \sum_{i,j} \Pr\left(\Theta = \theta_i, \tilde{\Theta} = \theta_j\right) \log \frac{\Pr\left(\Theta = \theta_i, \tilde{\Theta} = \theta_j\right)}{\Pr(\Theta = \theta_i) \Pr\left(\tilde{\Theta} = \theta_j\right)} \tag{4.28}$$

Note that when the privacy metric equal to zero, $\tilde{\Theta}$ contains no information about Θ and the maximum privacy level is achieved.

The optimal randomization probabilities are obtained, by minimizing the average distortion between X^n and Y^n subject to a privacy level of Θ, using the following optimization problem

$$\begin{aligned} \underset{\{P_{ji}\}_{j,i}}{\text{minimize}} \quad & \mathbb{E}[d\left(X^n, Y^n\right)] \\ & P_{ji} \geq 0, \forall i, j \\ & \sum_j P_{ji} = 1, \quad \forall i \\ & I(\Theta;\tilde{\Theta}) \leq I_0 \end{aligned} \tag{4.29}$$

The next theorem states that the optimization problem above is a convex optimization problem. Hence, the optimal randomization probabilities can be computed efficiently.

Theorem 4.2 *The optimal privacy filter design problem in* (4.29) *is a convex optimization problem.*

Proof See [4]. □

To study the privacy level of Θ under the proposed scheme, consider an estimator of Θ based on Y^n, denoted by $\hat{\Theta}(Y^n)$. Using Fano's inequality [9], the error probability of any estimator of Θ based on Y^n, can be bounded from below as

$$\Pr\left(\Theta \neq \hat{\Theta}\left(Y^n\right)\right) \geq \frac{H(\Theta\,|Y^n\,) - 1}{\log|\boldsymbol{\Theta}|}, \tag{4.30}$$

where $H(\Theta\,|Y^n\,)$ denotes the conditional entropy of Θ given Y^n. Using the definition of mutual information, we have that

$$H\left(\Theta\,\middle|Y^n\right) = H(\Theta) - I(\Theta; Y^n). \tag{4.31}$$

Notice that the following Markov chain holds

$$\Theta \to \tilde{\Theta} \to Y^n \to \hat{\Theta}\left(Y^n\right). \tag{4.32}$$

Thus, according to the data-processing inequality, we have that

$$I(\Theta; \hat{\Theta}\left(Y^n\right)) \leq I(\Theta; Y^n) \leq I(\Theta; \tilde{\Theta}). \tag{4.33}$$

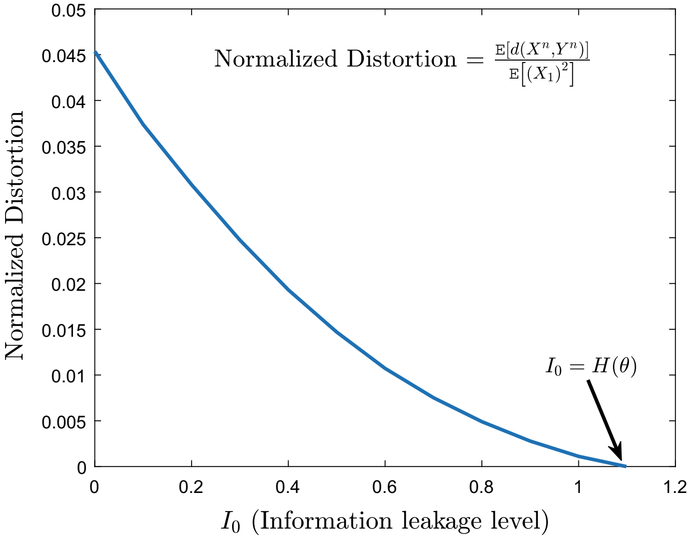

Fig. 4.6 Normalized distortion, under the second scheme, as a function of the privacy level

Hence, the upper bound on the mutual information between Θ and $\tilde{\Theta}$ in (4.29) ensures the privacy of Θ. That is, according to Fano's inequality, the privacy constraint imposes a lower bound on the performance of Bob in recovering Θ using the output of the privacy filter.

Gaussian Example

In this section, the distortion-privacy trade-off is numerically studied for a Gaussian information source under the second scheme. In our numerical result, X^n is modeled as a sequence of i.i.d. Gaussian random variables with zero mean and variance $\Theta \in \boldsymbol{\Theta} = \{1, 2, 3\}$. It is assumed that Θ is uniformly distributed over $\boldsymbol{\Theta}$.

Figure 4.6 shows the optimal level of the normalized distortion between the input and the output of the privacy filter as a function of the privacy level I_0. According to this figure, the minimum distortion level is achieved when I_0 is equal to $H(\Theta)$. In this example, the optimal randomization probabilities are computed using the fmincon solver in MATLAB®. Note that, when $\Theta = \tilde{\Theta}$, the input and output of the privacy filter are the same. Thus, the distortion is zero in this case and the leakage of the private information is at its maximum level.

Moreover, the distortion level increases as the leakage level of private information becomes small, since the mutual information between Θ and $\tilde{\Theta}$ decreases. The maximum distortion is attained when Θ and $\tilde{\Theta}$ are statistically independent. In this case, perfect privacy is achieved since the leakage level of private information is equal to zero.

4.5 Final Remarks

In this chapter, we studied the problem of statistical parameter privacy wherein the outputs of a parametric source are shared with an untrusted party. The objective is to design privacy filters which ensure the accuracy of the shared information while guaranteeing the privacy of the parameters.

Two different schemes were proposed for the statistical parameter privacy problem, where the mutual information was used as the privacy measure. In the first scheme, it was assumed that the parameter belonged to a continuous alphabet and the mutual information was exploited as a proxy for a Bayesian statistical metric of privacy. On the other hand, the parameter was assumed to belong to a finite set of possibilities under the second scheme, and the mutual information was used directly as the privacy measure via Fano's inequality. The optimal distortion-privacy trade-off was analyzed for this scheme.

Acknowledgements This work was supported by the Knut and Alice Wallenberg foundation and the Swedish Foundation for Strategic Research.

References

1. Asoodeh S, Alajaji F, Linder T (2015) On maximal correlation, mutual information and data privacy. In: 2015 IEEE 14th Canadian workshop on information theory (CWIT), pp 27–31
2. Asoodeh S, Alajaji F, Linder T (2016) Privacy-aware MMSE estimation. In: 2016 IEEE international symposium on information theory (ISIT), pp 1989–1993
3. Barber RF, Duchi JC (2014) Privacy and statistical risk: formalisms and minimax bounds. arXiv:1412.4451 [cs, math, stat]
4. Bassi G, Nekouei E, Skoglund M, Johansson KH (2019) Statistical parameter privacy. Technical report, KTH Royal Institute of Technology, Stockholm, Sweden. https://www.dropbox.com/s/zx49o4oxn95k3c9/main.pdf?dl=0
5. Bassi G, Piantanida P, Skoglund M (2018) Lossy communication subject to statistical parameter privacy. In: 2018 IEEE international symposium on information theory (ISIT), pp 1031–1035
6. du Calmon FP, Fawaz N (2012) Privacy against statistical inference. In: 2012 50th annual Allerton conference on communication, control, and computing (Allerton), pp 1401–1408
7. Chakraborty S, Bitouzé N, Srivastava M, Dolecek L (2013) Protecting data against unwanted inferences. In: 2013 IEEE information theory workshop (ITW), pp 1–5
8. Clarke BS, Barron AR (1990) Information-theoretic asymptotics of Bayes methods. IEEE Trans Inf Theory 36(3):453–471
9. Cover TM, Thomas JA (2006) Elements of information theory, 2nd edn. Wiley, New York, NY
10. Duchi JC, Jordan MI, Wainwright MJ (2018) Minimax optimal procedures for locally private estimation. J Am Stat Assoc 113(521):182–201
11. Dwork C, McSherry F, Nissim K, Smith A (2006) Calibrating noise to sensitivity in private data analysis. In: Halevi S, Rabin T (eds) Theory of cryptography. Springer, Berlin, pp 265–284
12. Farokhi F, Sandberg H (2018) Fisher information as a measure of privacy: preserving privacy of households with smart meters using batteries. IEEE Trans Smart Grid 9(5):4726–4734
13. Kalantari K, Sankar L, Kosut O (2017) On information-theoretic privacy with general distortion cost functions. In: 2017 IEEE international symposium on information theory, pp 2865–2869
14. Kay SM (1993) Fundamentals of statistical signal processing: estimation theory. Prentice-Hall, Upper Saddle River, NJ, USA

15. Le Ny J, Pappas GJ (2014) Differentially private filtering. IEEE Trans Autom Control 59(2):341–354
16. Lehmann EL, Casella G (1998) Theory of point estimation, 2nd edn. Springer, New York, NY
17. Mo Y, Murray RM (2017) Privacy preserving average consensus. IEEE Trans Autom Control 62(2):753–765
18. Nozari E, Tallapragada P, Cortés J (2017) Differentially private average consensus: obstructions, trade-offs, and optimal algorithm design. Automatica 81:221–231
19. Rissanen J (1986) Stochastic complexity and modeling. Ann Stat 14(3):1080–1100
20. Samarati P, Sweeney L (1998) Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, Computer Science Laboratory, SRI International
21. Sandberg H, Dán G, Thobaben R (2015) Differentially private state estimation in distribution networks with smart meters. In: 2015 54th IEEE conference on decision and control (CDC), pp 4492–4498
22. Tanaka T, Skoglund M, Sandberg H, Johansson K (2017) Directed information as privacy measure in cloud-based control. Technical report, KTH Royal Institute of Technology, Sweden. https://arxiv.org/abs/1705.02802
23. Tsai CY, Agarwal GK, Fragouli C, Diggavi S (2017) A distortion based approach for protecting inferences. In: 2017 IEEE international symposium on information theory (ISIT), pp 1913–1917
24. Wang Y, Huang Z, Mitra S, Dullerud GE (2017) Differential privacy in linear distributed control systems: entropy minimizing mechanisms and performance tradeoffs. IEEE Trans Control Netw Syst 4(1):118–130