TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Privacy against state estimation

Citation for published version (APA):
Murguia, C., Shames, I., Farokhi, F., & Nešic, D. (2020). Privacy against state estimation: An optimization framework based on the data processing inequality. *IFAC-PapersOnLine*, *53*(2), 7368-7373. https://doi.org/10.1016/j.ifacol.2020.12.1260

DOI:
10.1016/j.ifacol.2020.12.1260

Document status and date:
Published: 01/01/2020

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 06. Dec. 2021

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 53-2 (2020) 7368–7373

Privacy Against State Estimation: An Optimization Framework based on the Data Processing Inequality

Carlos Murguia * **Iman Shames** ** **Farhad Farokhi** **
Dragan Nešić **

* *Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: c.g.murguia@tue.nl.*
** *Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, Australia, e-mails: iman.shames@unimelb.edu.au, farhad.farokhi@unimelb.edu.au, dnesic@unimelb.edu.au.*

Abstract: Information about the system state is obtained through noisy sensor measurements. This data is coded and transmitted to a trusted user through an unsecured communication network. We aim at keeping the system state private; however, because the network is not secure, opponents might access sensor data, which can be used to estimate the state. To prevent this, before transmission, we randomize coded sensor data by passing it through a probabilistic mapping, and send the corrupted data to the trusted user. Making use of the *data processing inequality*, we cast the synthesis of the probabilistic mapping as a convex program where we minimize the *mutual information* (our privacy metric) between two estimators, one constructed using the randomized sensor data and the other using the actual undistorted sensor measurements, for a *desired level of distortion*–how different coded sensor measurements and distorted data are allowed to be.

Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)

Keywords: Privacy, Stochastic Systems, Mutual Information, Data Processing Inequality.

1. INTRODUCTION

Technological advances have led to an alarming widespread loss of privacy in society and vulnerabilities within critical infrastructure. Adversaries might infer critical (private) information about the operation of systems from public data available on the internet and unsecured/public servers and communication networks. A motivating example for this privacy loss is the data collection, classification, and sharing by the Internet-of-Things (IoT), (Weber, 2010), which is often performed without the user's informed consent. Another example of privacy loss is the potential use of data from smart electrical meters by criminals, advertising agencies, and governments, for monitoring the presence and activities of occupants, (Rajagopalan et al., 2011; Tan et al., 2013). These privacy concerns show that there is an acute need for privacy preserving mechanisms capable of handling the new privacy challenges induced by a hyperconnected world, which, in turn, has attracted the attention of researchers from different fields (e.g., computer science, information theory, and control theory) in the broad area of privacy and security of Cyber-Physical Systems (CPSs), see, e.g., (Farokhi and Sandberg, 2017)-(Sultangazin and Tabuada, 2019).

In many engineering applications, information about the state of systems, X, is obtained through sensor measurements. Once this information is collected, it is sent to a trusted server for signal processing and decision-making purposes through communication networks. If the communication network is public/unsecured, opponents might access and estimate the system state. To avoid this, before transmission, we randomize sensor data by passing it through a probabilistic mapping, and send the corrupted data to the trusted server. This mapping is designed to hide (as much as possible) information about the state X. Note, however, that it is not desired to overly distort the original sensor data. We might change the data excessively for its legitimate use. Hence, when designing the probabilistic mappings, we need to take into account the trade-off between *privacy* and *distortion*. As *distortion metric*, we use the *mean squared error* between the original sensor data, Y, and its randomized version, $Z = G(Y)$, for some probabilistic mapping $G(\cdot)$. In this manuscript, we follow an information-theoretic approach to privacy. As privacy metric, we propose the *mutual information*, (Cover and Thomas, 1991), $I[\hat{X}(Y); \hat{X}(Z)]$, for given pair of estimators of the state, $\hat{X}(Y)$ and $\hat{X}(Z)$, obtained through the original sensor data Y and the distorted $Z = G(Y)$, respectively. We design the probabilistic mapping $G(Y)$ (characterized by the conditional probability distribution $p_{Y|Z}(y|z)$) to minimize $I[\hat{X}(Y); \hat{X}(Z)]$, for a *desired level of distortion* – how different quantized sensor measurements and distorted data are allowed to be. We pose the

⋆ This work was supported by the Australian Research Council (ARC) under the Project DP170104099; and the NATO Science for Peace and Security (SPS) PROGRAMME under the project SPS.SFP G5479.

2405-8963 Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2020.12.1260

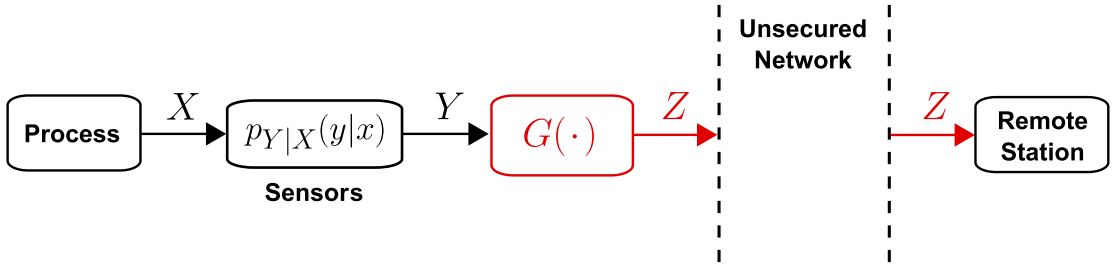

Fig. 1. System Configuration.

problem of synthesising $p_{Y|Z}(y|z)$ (the map $G(\cdot)$) as a convex program subject to linear constraints.

Randomizing data to increase privacy is common practice. In privacy of databases, a popular approach is differential privacy (Ny and Pappas, 2014; Dwork, 2008), where random vectors are added to the response of queries so that private information in the database cannot be inferred. Because it provides differential privacy guarantees, Laplace noise is usually used (Dwork and Roth, 2014). However, when maximal privacy with minimal distortion is desired, Laplace noise is generally not the optimal solution. This raises the fundamental question: for a given allowable distortion level, what is the randomizing mechanism achieving maximal privacy? This question has many possible answers depending on the particular privacy and distortion metrics being considered and the system configuration (Han et al., 2014)-(Wang et al., 2014). There are also results addressing this question from an information theoretic perspective, where information metrics – e.g., mutual information, entropy, Kullback-Leibler divergence, and Fisher information – are used to quantify privacy (Rajagopalan et al., 2011; Tan et al., 2013; Farokhi and Sandberg, 2017; Farokhi et al., 2015; Farokhi and Nair, 2016; du Pin Calmon and Fawaz, 2012; Salamatian et al., 2015; Belmega et al., 2015).

If the data to be kept private follows continuous probability distributions, finding the optimal additive noise to maximize privacy (for any privacy metric and even without considering distortion) is a difficult problem. If a close-form solution for the distribution is desired, the problem amounts to solving a set of nonlinear partial differential equations which, in general, might not have a solution, and even if they have it, it is hard to find (Farokhi and Sandberg, 2017). This problem has been addressed by imposing some particular structure on the considered distributions or assuming the data to be kept private is deterministic (Farokhi and Sandberg, 2017; Soria-Comas and Domingo-Ferrer, 2013; Geng and Viswanath, 2014). The authors in (Soria-Comas and Domingo-Ferrer, 2013; Geng and Viswanath, 2014) consider deterministic data sets and treat optimal distributions as distributions that concentrate probability around zero as much as possible while ensuring differential privacy. Under this framework, they obtain a family of piecewise constant density functions that achieve minimal distortion for a given level of privacy. Farokhi and Sandberg (2017) consider the problem of preserving the privacy of deterministic databases using constrained additive noise. They use the Fisher information and the Cramer-Rao bound to construct a privacy metric between noise-free data and the one with the additive noise and find the probability density function that minimizes it.

Most of the aforementioned papers propose optimal continuous distributions assuming deterministic data. However, in a cyber-physical-systems context, the inherent system dynamics and unavoidable system and sensor noise lead to stochasticity and thus existing tools do not fully fit this setting. As we prove in this manuscript, under some mild assumptions, we can cast the problem of finding the optimal probabilistic mapping as a constrained convex optimization.

2. NOTATION AND PRELIMINARIES

The symbol $\mathbb{R}$ stands for the real numbers, $\mathbb{R}_{>0}(\mathbb{R}_{\geq 0})$ denotes the set of positive (non-negative) real numbers. The symbol $\mathbb{N}$ stands for the set of natural numbers. The Euclidian norm in $\mathbb{R}^n$ is denoted by $||X||$, $||X||^2 = X^\top X$. For a discrete random vector X with alphabet $\mathcal{X} = \{x_1, \ldots, x_N\}$, $x_i \in \mathbb{R}^m$, $m, N \in \mathbb{N}$, $i \in \{1, \ldots, N\}$, we denote its probability mass function (pmf) as $p_X(x) = \Pr[X = x]$, $x \in \mathcal{X}$, where $\Pr[B]$ denotes probability of event B. We denote by "Simplex" the probability simplex defined by $\sum_{x\in\mathcal{X}} p_X(x) = 1$, $p_X(x) \geq 0$ for all $x \in \mathcal{X}$. We denote independence between two random vectors, X and Y, as $X \perp\!\!\!\perp Y$, and the expected value of X with $E[X]$. Given two numbers a and b, $b > 0$, the notation $a \bmod b$ stands for a modulo b, and for a vector $a = (a_1, \ldots, a_n)^\top$, $a_i \in \mathbb{R}_{>0}$, $i = 1, \ldots, n$, $a \bmod b = (a_1 \bmod b, \ldots, a_n \bmod b)^\top$.

2.1 Mutual Information

Definition 1. (Cover and Thomas, 1991) *Consider discrete random vectors, X and Y, with joint probability mass function $p(x, y)$ and marginal probability mass functions, $p(x)$ and $p(y)$, respectively. Their mutual information $I[X;Y]$ is defined as the relative entropy between the joint distribution and the product distribution $p(x)p(y)$:*

$$I[X;Y] := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$$

The mutual information $I[X;Y]$ between two jointly distributed vectors, X and Y, is a measure of the statistical dependence between X and Y.

3. METRICS AND PROBLEM FORMULATION

Let X be the state of some stochastic process that must be kept private. The alphabet and probability mass function of X are denoted as $\mathcal{X} = \{x_1, \ldots, x_N\}$, $x_i \in \mathbb{R}^{n_x}$, $n_x \in \mathbb{N}$,

$i \in \{1,\dots,N_X\}$ and $p_X(x) = \Pr[X = x]$, $x \in \mathcal{X}$, respectively. Information about the state is obtained through n_y sensors of the form $Y = h(X)$, $Y \in \mathbb{R}^{n_y}$, for some (stochastic or deterministic) mapping $h : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ characterized by the transition probabilities $p_{Y|X}(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, where $\mathcal{Y} = \{y_1,\dots,y_{N_Y}\}$, $y_i \in \mathbb{R}^{n_y}$, $N_Y, n_y \in \mathbb{N}$. Our privacy scheme discloses $Z = G(Y)$, for some stochastic mapping $G : \mathcal{Y} \to \mathcal{Y}$, instead of Y, so that, when releasing Z, the individual entries of X are "hidden". The mapping $G(\cdot)$ is characterized by the transition probabilities $p_{Z|Y}(z|y) = \Pr[Z = z|Y = y]$, $y, z \in \mathcal{Y}$, i.e., $Z = G(Y) \in \mathcal{Y}$. Realizations of the vector Z are transmitted over a public (unsecured) communication channel to a remote station, see Figure 1. Note that even by passing Y through $G(\cdot)$ before transmission, information about X is directly accessible through the public channel. Here, we aim at finding the mapping $G(\cdot)$ (the transition probabilities $p_{Z|Y}(z|y)$) that minimizes this information leakage. Note, however, that we do not want to make Y and Z overly different. By passing Y through $G(\cdot)$, we might *distort* it excessively for practical purposes. Hence, when designing the distribution $p_{Z|Y}(z|y)$, we need to consider the trade-off between *privacy* and *distortion*. As *distortion metric*, we use the mean squared error: $E[||Z - Y||^2]$. Let $\hat{X}(Y)$ and $\hat{X}(Z)$ denote estimates of X through Y and Z, respectively. As privacy metric, we use the mutual information $I[\hat{X}(Y);\hat{X}(Z)]$ for given pair of estimators $\hat{X}(Y)$ and $\hat{X}(Z)$. Hence, we aim at minimizing $I[\hat{X}(Y);\hat{X}(Z)]$ subject to $E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon$, for a desired level of distortion $\epsilon \in \mathbb{R}_{>0}$, using as decision variables the conditional probability mass function $p_{Z|Y}(z|y) = \Pr[Z = z|Y = y]$, $y, z \in \mathcal{Y}$.

Remark 1. The number of optimization variables $p_{Z|Y}(z|y)$ depends on the number of mass points, N_Y, in $\mathcal{Y}$. We aim at computing an optimal transition probability $p_{Z|Y}(z|y)$ from each element of the alphabet of Y to every element of the alphabet of Z, and because Y and Z take values from the alphabet $\mathcal{Y}$, we have $(N_Y)^2$ optimization variables to minimize $I[\hat{X}(Y);\hat{X}(Z)]$. That is, the number of variables grows quadratically with the cardinality of $\mathcal{Y}$. Thus, for alphabets with a large number of elements, the number of variables could lead to computationally untractable optimization problems. A solution to this dimensionality issue is to impose some structure on $p_{Z|Y}(z|y)$ to reduce the variables. We propose a systematic way to achieve this using additive random vectors. We impose structure to the probabilistic mapping $G(\cdot)$ (see Figure 1) so that the number of variables in $p_{Z|Y}(z|y)$ is reduced to N_Y.

The proposed $G(\cdot)$ consists of the following three objects: 1) a coding function $\alpha : \mathcal{Y} \to \{0,1,\dots,N_Y-1\} =: \bar{\mathcal{Y}}$ that indexes each element of $\mathcal{Y}$; 2) a discrete random vector V, independent of Y, with alphabet $\mathcal{V} := \{0,1,\dots,N_V-1\}$, $N_V \in \mathbb{N}$, and probability mass function $p_V(v)$, $v \in \mathcal{V}$; and 3) a decoding function $\beta : \bar{\mathcal{Y}} \to \mathcal{Y}$. We characterize each of these objects before introducing the mapping $G(\cdot)$. The indexing (coding) function $\alpha : \mathcal{Y} \to \bar{\mathcal{Y}}$ is defined as

$$\alpha(\zeta) := \begin{cases} 0, & \text{if } \zeta = y_1, \\ \vdots & \\ N_Y - 1, & \text{if } \zeta = y_{N_Y}. \end{cases} \tag{1}$$

For given $Y \in \mathcal{Y}$ and corresponding $\alpha(Y) \in \bar{\mathcal{Y}}$, we add a realization of the process $V \in \{0,1,\dots,N_V-1\}$ to randomize $\alpha(Y)$, and project the sum onto the ring $\{0,1,\dots,N_Y-1\}$, i.e., $(\alpha(Y)+V) \bmod N_Y \in \mathcal{Y}$, where $\bmod N_Y$ denotes modulo N_Y. We project $\alpha(Y)+V$ onto $\mathcal{Y}$ to ensure that Z has the same alphabet as Y. Then, we decode the sum using the function $\beta : \bar{\mathcal{Y}} \to \mathcal{Y}$ defined as

$$\beta(\xi) := \begin{cases} y_1, & \text{if } \xi = 0, \\ \vdots & \\ y_{N_Y}, & \text{if } \xi = N_Y - 1. \end{cases} \tag{2}$$

Note that $\beta(\alpha(\zeta)) = \zeta$ and $\alpha(\beta(\xi)) = \xi$. We construct the mapping $G : \mathcal{Y} \to \mathcal{Y}$, $Y \mapsto G(Y)$, combining (1) and (2):

$$G(Y) := \beta\big((\alpha(Y) + V) \bmod N_Y\big). \tag{3}$$

Since $\alpha(\cdot)$ and $\beta(\cdot)$ are fixed injective functions, we can only use the probability mass function $p_V(v)$, $v \in \mathcal{V}$, to minimize $I[\hat{X}(Y);\hat{X}(Z)]$. In what follows, we formally present the optimization problem we seek to address.

Problem 1. Given the probability distribution $p_{X,Y}(x,y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, desired distortion level $\epsilon \in \mathbb{R}_{\geq 0}$, a pair of estimators $\hat{X}(Y)$ and $\hat{X}(Z)$, and the probabilistic mapping (1)-(3), find the probability mass function $p_V(v)$ solution of the optimization problem:

$$\begin{cases} \min\limits_{p_V(v)} \; I[\hat{X}(Y);\hat{X}(Z)], \\ \quad s.t. \; E[||Z - Y||^2] \leq \epsilon, \\ \qquad V \perp\!\!\!\perp Y, \; \textit{and } p_V(v) \in \; \textit{Simplex}. \end{cases} \tag{4}$$

Note that in Problem 1, the optimization is performed for a given pair of estimators. We could select a particular pair of estimators and pose the problem based on this pair. However, it is not realistic to assume that we know the estimator that an opponent would use. We can, however, look for an upper bound on $I[\hat{X}(Y);\hat{X}(Z)]$ that holds for any pair of estimators, $\hat{X}(Y)$ and $\hat{X}(Z)$, and then minimize this upper bound over $p_V(v)$ subject to the distortion constraint $E[||Z - Y||^2] \leq \epsilon$, at the price of suboptimality of the solution.

4. SUBOPTIMAL SOLUTION TO PROBLEM 1

In this section, we provide a detailed formulation of a relaxed version of Problem 1. We relax the problem by working with an upper bound on the cost $I[\hat{X}(Y);\hat{X}(Z)]$ obtained using the data processing inequality.

Proposition 1. Let $\hat{X}(Z) = h_Z(Z)$ and $\hat{X}(Y) = h_Y(Y)$ be estimates of X through Z and Y, respectively, for some deterministic functions $h_Z, h_Y : \mathbb{R}^{n_y} \to \mathbb{R}^{n_x}$. For any pair of functions $h_Z(\cdot)$ and $h_Y(\cdot)$, $I[\hat{X}(Y);\hat{X}(Z)] \leq I[Y;Z]$.

Proof: The assertion follows from the *data processing inequality*, (Cover and Thomas, 1991).

Remark 2. Proposition 1 has an interesting interpretation: for any pair of estimators $\hat{X}(Z) = h_Z(Z)$ and $\hat{X}(Y) = h_Y(Y)$ that can be constructed using Y and Z, respectively; the mutual information between them is always upper bounded by $I[Y;Z]$ independently of the choice of estimators. This implies that by minimizing $I[Y;Z]$, we are effectively decreasing the information $I[\hat{X}(Y);\hat{X}(Z)]$. Indeed, the tightness of this bound depends on the particular choice of estimators.

In the following lemma, we write the cost function $I[Y;Z]$ in terms of the optimization variables $p_V(v)$, and prove that $I[Y;Z]$ is convex in $p_V(v)$.

Lemma 1. $I[Y;Z]$ is a convex function of $p_V(v)$ for given $p_Y(y)$, and can be written compactly as

$$I[Y;Z] = \sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Y}} p_{Z|Y}(z|y)p_Y(y)\log\frac{p_{Z|Y}(z|y)}{p_Z(z)}, \quad (5a)$$

$$p_{Z|Y}(z|y) = p_V\big((\alpha(z)-\alpha(y)) \bmod N_Y\big), \quad (5b)$$

$$p_Z(z) = \sum_{y\in\mathcal{Y}} p_{Z|Y}(z|y)p_Y(y). \quad (5c)$$

Proof: The expression on the right-hand side of (5a) follows by inspection of Definition 1, and the fact that the joint and marginal distributions can be written as $p_{Z,Y}(z,y) = p_{Z|Y}(z|y)p_Y(y)$ (chain rule) and $p_Z(z) = \sum_{y\in\mathcal{Y}} p_{Z,Y}(z,y) = \sum_{y\in\mathcal{Y}} p_{Z|Y}(z|y)p_Y(y)$ (marginalization), respectively. Convexity of (5a) with respect to $p_{Z|Y}(z|y)$, for given $p_Y(y)$, follows from (Cover and Thomas, 1991, Theorem 2.7.4). However, our optimization variables are $p_V(v)$ and not $p_{Z|Y}(z|y)$. By definition, $p_{Z,Y}(z,y) = \Pr[Z=z, Y=y]$, $z,y\in\mathcal{Y}$. Using (3), we can expand $\Pr[Z=z,Y=y]$ as follows

$$\begin{aligned}
&\Pr[Z=z,Y=y]\\
&= \Pr[\beta\big((\alpha(\tilde{Y})+V) \bmod N_Y\big)=z, Y=y]\\
&= \Pr[V=\big(\alpha(z)-\alpha(y)\big) \bmod N_Y, Y=y]\\
&\overset{(a)}{=} \Pr[V=\big(\alpha(z)-\alpha(y)\big) \bmod N_Y]\ \Pr[Y=y]\\
&\overset{(b)}{=} p_V\big(\big(\alpha(z)-\alpha(y)\big) \bmod N_Y\big)p_Y(y),
\end{aligned}$$

where (a) follows from independence between V and Y and (b) by construction of $p_V(v)$ since $p_V(v) = \Pr[V=v]$, $v\in\mathcal{V}$. It follows that $p_{Z|Y}(z|y) = p_{Z,Y}(z,y)/p_Y(y) = p_V\big(\big(\alpha(z)-\alpha(y)\big) \bmod N_Y\big)$ and thus (5b) holds true. It remains to prove that $I[Y;Z]$ is convex in $p_V(v)$ for given $p_Y(y)$. We have concluded convexity of $I[Y;Z]$ with respect to $p_{Z|Y}(z|y)$ above. Hence, because $p_{Z|Y}(z|y) = p_V\big(\big(\alpha(z)-\alpha(y)\big) \bmod N_Y\big)$ and $p_V\big(\big(\alpha(z)-\alpha(y)\big) \bmod N_Y\big)$ is a linear transformation of $p_V(v)$ (note that $p_V\big(\big(\alpha(z)-\alpha(y)\big) \bmod N_Y\big) = p_V(v)$ for $\big(\alpha(z)-\alpha(y)\big) \bmod N_Y = v$ and zero otherwise), the cost $I[Y;Z]$ is convex in $p_V(v)$ because convexity is preserved under affine transformations, (Boyd and Vandenberghe, 2004). ■

In light of Proposition 1 and Remark 2, from now on, we focus on minimizing the upper bound $I[Y;Z]$. By Lemma 1, $I[Y;Z]$ is convex in our decision variables $p_V(v)$ for given $p_Y(y)$. Then, if the distortion constraint, $E[||Z-Y||^2] \leq \epsilon$, is convex in $p_V(v)$, we could minimize $I[Y;Z]$ efficiently using off-the-shelf optimization algorithms.

Lemma 2. $E[||Z-Y||^2]$ is a linear function of $p_V(v)$ for given $p_Y(y)$, and can be written compactly as

$$E[||Z-Y||^2] = \sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Y}} p_{Z,Y}(z,y)\,(z-y)^2\,, \quad (6a)$$

$$p_{Z,Y}(z,y) = p_V\big((\alpha(z)-\alpha(y)) \bmod N_Y\big)p_Y(y). \quad (6b)$$

Proof: Define $d(Z,Y) := ||Z-Y||^2$. The function $d(Z,Y)$ is a deterministic function of two jointly distributed random vectors, Z and Y, with joint distribution $p_{Z,Y}(z,y) = p_{Z|Y}(z|y)p_Y(y)$. Therefore, see, e.g., (Ross, 2006) for details, $E[d(Z,Y)] = \sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Y}} p_{Z,Y}(z,y)d(z,y)$, and, by (5b), $p_{Z|Y}(z|y) = p_V\big((\alpha(z)-\alpha(y)) \bmod N_Y\big)$ (see the proof of Lemma 1 for details). It follows that $E[d(Z,Y)]$ is given by (6) and because $p_{Z|Y}(z|y)$ is a linear transformation of $p_V(v)$, $E[d(Z,Y)]$ is linear in $p_V(v)$ for given $p_Y(y)$. ■

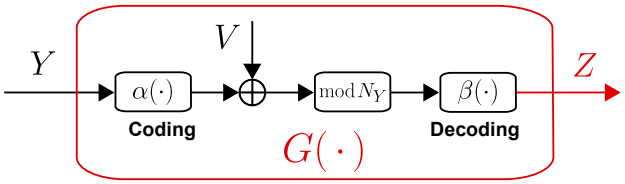

Fig. 2. Schematic diagram of the mapping $G(k,\cdot)$.

Hence, by Lemma 1 and Lemma 2, given the probabilities $p_Y(y)$ of Y, we can numerically minimize $I[Y;Z]$ constrained to $E[||Z-Y||^2] \leq \epsilon$. In the following theorem, we summarize the discussion of this section.

Theorem 1. Given $p_Y(y)$, desired distortion level $\epsilon\in\mathbb{R}_{\geq 0}$, and the mapping (1)-(3), the probability mass function $p_V(v)$ that minimizes $I[Y;Z]$ subject to the distortion constraint, $E[||Z-Y||^2] \leq \epsilon$, can be found by solving the following convex program:

$$\begin{cases}
\min\limits_{p_V(v)} \displaystyle\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Y}} p_{Z|Y}(z|y)p_Y(y)\log\frac{p_{Z|Y}(z|y)}{p_Z(z)},\\
\quad s.t. \displaystyle\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Y}} p_{Z|Y}(z|y)p_Y(y)(z,y)\,(z-y)^2 \leq \epsilon,\\
\qquad p_{Z|Y}(z|y) = p_V\big((\alpha(z)-\alpha(y)) \bmod N_Y\big),\\
\qquad p_Z(z) = \displaystyle\sum_{y\in\mathcal{Y}} p_{Z|Y}(z|y)p_Y(y),\\
\qquad \text{and, } p_V(v) \in \text{ Simplex }.
\end{cases} \quad (7)$$

Proof: *Theorem 1 follows from Lemma 1 and Lemma 2.* ■

5. SIMULATIONS

Consider sensors $Y\in\mathcal{Y}\subset\mathbb{R}^4$ on the alphabet:

$$\mathcal{Y} = \left\{\begin{bmatrix}y_1\\y_1\\y_1\\y_1\end{bmatrix}, \begin{bmatrix}y_2\\y_1\\y_1\\y_1\end{bmatrix}, \begin{bmatrix}y_1\\y_2\\y_1\\y_1\end{bmatrix}, \begin{bmatrix}y_2\\y_1\\y_2\\y_1\end{bmatrix}, \ldots, \begin{bmatrix}y_1\\y_2\\y_2\\y_2\end{bmatrix}, \begin{bmatrix}y_2\\y_2\\y_2\\y_2\end{bmatrix}\right\}, \quad (8)$$

with $y_1 = 13.42$, $y_2 = 14.03$, and $|\mathcal{Y}| = 16$. For all the subsequent figures, we index the mass points of the alphabets $\mathcal{Y}$ and $\mathcal{V}$ following the ordering logic in (8). For instance, for $\mathcal{Y}$ in (8), $(y_1,y_1,y_1,y_1)^\top$ is indexed as 1, $(y_2,y_1,y_1,y_1)^\top$ as 2, $(y_2,y_2,y_2,y_2)^\top$ as 16, and so forth. The probability mass function, $p_Y(y)$, of Y is depicted in Figure 3. We let the distorting random vector V (see (1)-(3)) have an alphabet $\mathcal{V}$ with $|\mathcal{V}| = 16$ given by

$$\mathcal{V} = \left\{\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \ldots, \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right\}. \quad (9)$$

We consider the distortion bounds $\epsilon = \infty, 0.5, 0.2$. The bound $\epsilon = \infty$ means that the optimization problem in (7) is solved without considering the distortion constraint. In Figure 4, we show the evolution of the optimal probability distribution $p_V^*(v)$ solution to (7) for $\epsilon = \infty, 0.5, 0.2$.

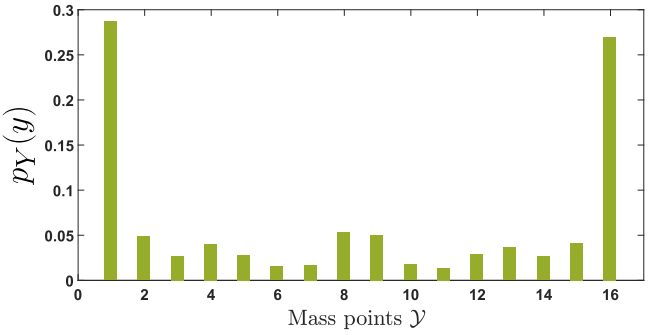

Fig. 3. Probability mass function $p_Y(y)$, $y \in \mathcal{Y}$.

Note that the optimal $p_V^*(v)$ in Figure 4(a) (for $\epsilon = \infty$) is uniform. Uniform $p_V^*(v)$ makes the input, Y, and the output, Z, completely independent at the price of large distortion. On the other hand, see Figure 4(c), as $\epsilon \to 0$, the optimal distribution $p_V^*(v)$ concentrates most of its probability at the first mass point (the zero vector, see (9)). This is intuitive as the zero vector leads to no distortion. Therefore, $p_V^*(v)$ varies from a uniform distribution (no distortion constraint) and a single mass point at the zero vector (no distortion allowed). In Figure 4(b), we show the optimal distribution for finite and larger than zero ϵ. In this case, the optimal distribution follows some nontrivial pattern that depends on $p_Y(y)$. The resulting optimal cost $I^*[Y;Z]$ is given by $I^*[Y;Z] = 0, 0.20, 1.01$ bits, for $\epsilon = \infty, 0.5, 0.2$, respectively. Compare these costs against $I[Y;Y] = H[Y] = 2.2$ bits (self-information), where $H[\cdot]$ denotes entropy (Cover and Thomas, 1991). The entropy $H[Y]$ characterizes the information that would be disclosed if no privacy preserving mapping was in place. Note that for $\epsilon = \infty$ (uniform optimal distribution in Figure 4(a)) the cost $I[Y;Z]$ is zero (because Y and Z are independent in this case). As $\epsilon \to 0$ (no distortion allowed), $I[Y;Z] \to H[Y]$.

6. CONCLUSIONS

We have presented a mathematical framework built around information theory and convex optimization to deal with privacy problems raised by the use of public/unsecured communication networks to transmit sensor data. In particular, to prevent adversaries from obtaining an accurate estimate of the state, we have provided tools (in terms of convex programs) to optimally randomize (via some probabilistic mappings) sensor data before transmission for a desired level of distortion. That is, given a maximum level of distortion tolerated by a particular application, we give tools to synthesize probabilistic mappings that maximize privacy (in the sense of hiding the state as much as possible) while satisfying the distortion constraint on the original sensor data. We have presented simulation experiments to show the performance of our tools. Note that we have found some nontrivial distorting probability distributions that highly depend on the probability distribution of the sensor data and the desired distortion level.

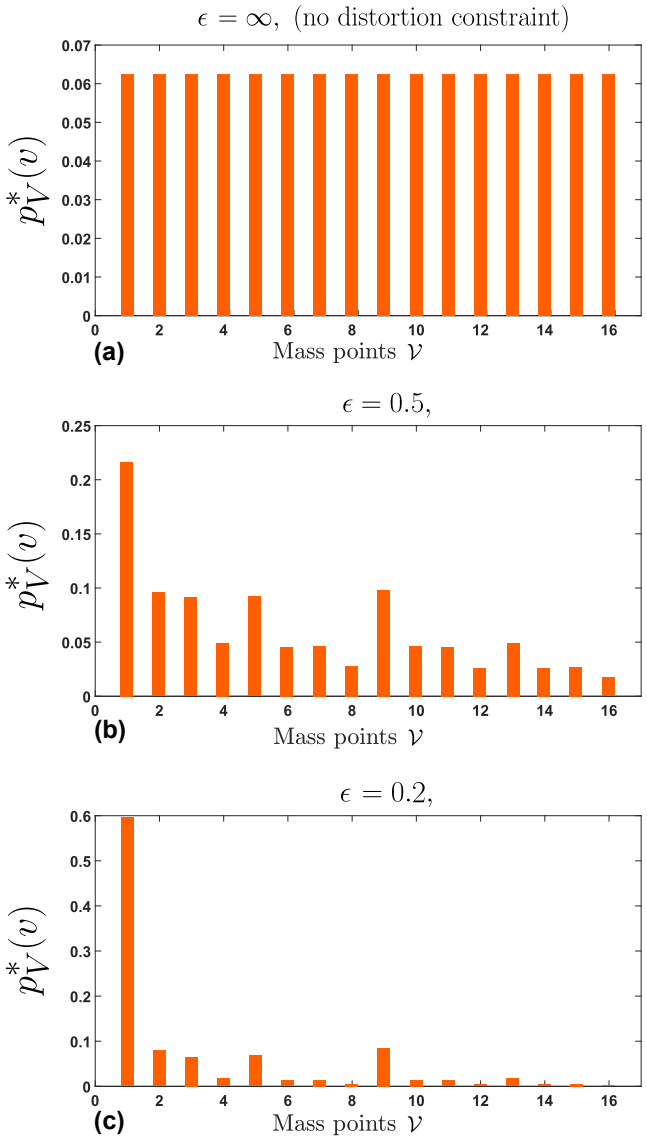

Fig. 4. Optimal probability distribution $p_V^*(v)$ solution to (7) for different distortion upper bounds ϵ.

REFERENCES

Ahmed, C.M., Murguia, C., and Ruths, J. (2017). Model-based attack detection scheme for smart water distribution networks. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, ASIA CCS '17, 101–113.

Belmega, E.V., Sankar, L., and Poor, H.V. (2015). Enabling data exchange in two-agent interactive systems under privacy constraints. *IEEE Journal of Selected Topics in Signal Processing*, 9, 1285–1297.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press, New York, NY, USA.

Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA.

du Pin Calmon, F. and Fawaz, N. (2012). Privacy against statistical inference. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 1401–1408.

Dwork, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*, 1–19. Springer Berlin Heidelberg, Berlin, Heidelberg.

Dwork, C. and Roth, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9, 211–407.

Farokhi, F. and Sandberg, H. (2017). Optimal privacy-preserving policy using constrained additive noise to minimize the fisher information. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*.

Farokhi, F., Sandberg, H., Shames, I., and Cantoni, M. (2015). Quadratic Gaussian privacy games. In *2015 54th IEEE Conference on Decision and Control (CDC)*, 4505–4510.

Farokhi, F. and Nair, G. (2016). Privacy-constrained communication. *IFAC-PapersOnLine*, 49, 43 – 48.

Geng, Q. and Viswanath, P. (2014). The optimal mechanism in differential privacy. In *2014 IEEE International Symposium on Information Theory*, 2371–2375.

Han, S., Topcu, U., and Pappas, G.J. (2014). Differentially private convex optimization with piecewise affine objectives. In *53rd IEEE Conference on Decision and Control*.

Hashemi, N. and Ruths, J. (2019). Generalized chi-squared detector for lti systems with non-gaussian noise. In *2019 American Control Conference (ACC)*.

Hashemi, N., German, E.V., Ramirez, J.P., and Ruths, J. (2019). Filtering approaches for dealing with noise in anomaly detection. In *arXiv:1909.01477*.

Hashemi, N. and Ruths, J. (2019). Gain design via lmis to minimize the impact of stealthy attack. In *arXiv:1909.12452*.

Hashemil, N., Murguia, C., and Ruths, J. (2018). A comparison of stealthy sensor attacks on control systems. In *proceedings of the American Control Conference (ACC), 2018*.

Kafash, S.H., Giraldo, J., Murguia, C., Cardenas, A.A., and Ruths, J. (2018). Constraining attacker capabilities through actuator saturation. In *proceedings of the American Control Conference (ACC), 2018*.

Miao, F., Zhu, Q., Pajic, M., and Pappas, G.J. (2014). Coding sensor outputs for injection attacks detection. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, 5776–5781.

Murguia, C. and Ruths, J. (2016a). Characterization of a cusum model-based sensor attack detector. In *proceedings of the 55th IEEE Conference on Decision and Control (CDC)*.

Murguia, C. and Ruths, J. (2016b). Cusum and chi-squared attack detection of compromised sensors. In *proceedings of the IEEE Multi-Conference on Systems and Control (MSC)*.

Murguia, C. and Ruths, J. (2018). On reachable sets of hidden cps sensor attacks. In *proceedings of the American Control Conference* (ACC).

Murguia, C., Shames, I., Farokhi, F., and Nešić, D. (2018). On privacy of quantized sensor measurements through additive noise. In *proceedings of the 57th IEEE Conference on Decision and Control (CDC)*.

Murguia, C., van de Wouw, N., and Ruths, J. (2016). Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools. In *proceedings of the IFAC World Congress*.

Ny, J.L. and Pappas, G.J. (2014). Differentially private filtering. *IEEE Transactions on Automatic Control*, 59, 341–354.

Ozarow, L.H. and Wyner, A.D. (1985). Wire-tap channel ii. In T. Beth, N. Cot, and I. Ingemarsson (eds.), *Advances in Cryptology*, 33–50. Springer Berlin Heidelberg, Berlin, Heidelberg.

Pasqualetti, F., Dorfler, F., and Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58, 2715–2729.

Rajagopalan, S.R., Sankar, L., Mohajer, S., and Poor, H.V. (2011). Smart meter privacy: A utility-privacy framework. In *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 190–195.

Renganathan, V., Hashemi, N., Ruths, J., and Summers, T. (2019). Distributionally robust tuning of anomaly detectors in cyber-physical systems with stealthy attacks. In *arXiv:1909.12506*.

Ross, M. (2006). *Introduction to Probability Models, Ninth Edition*. Academic Press, Inc., Orlando, FL, USA.

Rothstein Morris, E., Murguia, C., and Ochoa, M. (2017). Design-time quantification of integrity in cyber-physical-systems. In *Proceedings of the 2017 ACM SIGSAC Workshop on Programming Languages and Analysis for Security*.

Salamatian, S., Zhang, A., du Pin Calmon, F., Bhamidipati, S., Fawaz, N., Kveton, B., Oliveira, P., and Taft, N. (2015). Managing your private and public data: Bringing down inference attacks against your privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9, 1240–1255.

Soria-Comas, J. and Domingo-Ferrer, J. (2013). Optimal data-independent noise for differential privacy. *Information Sciences*, 250, 200 – 214.

Sultangazin, A. and Tabuada, P. (2019). Symmetries and isomorphisms for privacy in control over the cloud. *arXiv:1906.07460*.

Tan, O., Gunduz, D., and Poor, H.V. (2013). Increasing smart meter privacy through energy harvesting and storage devices. *IEEE Journal on Selected Areas in Communications*, 31, 1331–1341.

Wang, Y., Huang, Z., Mitra, S., and Dullerud, G.E. (2014). Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems. In *53rd IEEE Conference on Decision and Control*, 2130–2135.

Weber, R.H. (2010). Internet of things - new security and privacy challenges. *Computer Law and Security Review*, 26, 23–30.

Wyner, A.D. (1975). The wire-tap channel. *The Bell System Technical Journal*, 54, 1355–1387.

Yang, T., Murguia, C., Kuijper, M., and Nešić, D. (2018a). A multi-observer approach for attack detection and isolation of discrete-time nonlinear systems. In *2018 Australian New Zealand Control Conference (ANZCC)*.

Yang, T., Murguia, C., Kuijper, M., and Nešić, D. (2018b). A robust circle-criterion observer-based estimator for discrete-time nonlinear systems in the presence of sensor attacks. In *2018 IEEE Conference on Decision and Control (CDC)*.

Yang, T., Murguia, C., Kuijper, M., and Nešić, D. (2019). An unknown input multi-observer approach for estimation, attack isolation, and control of lti systems under actuator attacks. In *2019 18th European Control Conference (ECC)*.