Author/s:
Murguia, C;Shames, I;Farokhi, F;Nesic, D;Poor, HV

Title:
On privacy of dynamical systems: An optimal probabilistic mapping approach

Date:
2021-01-01

Citation:
Murguia, C., Shames, I., Farokhi, F., Nesic, D. & Poor, H. V. (2021). On privacy of dynamical systems: An optimal probabilistic mapping approach. IEEE Transactions on Information Forensics and Security, 16, pp.2608-2620. https://doi.org/10.1109/TIFS.2021.3055022.

Persistent Link:
https://hdl.handle.net/11343/268325

# On Privacy of Dynamical Systems: An Optimal Probabilistic Mapping Approach (Extended Preprint)

Carlos Murguia, Iman Shames, Farhad Farokhi, Dragan Nešić, and H. Vincent Poor



***Abstract*—We address the problem of maximizing privacy of stochastic dynamical systems whose state information is released through quantized sensor data. In particular, we consider the setting where information about the system state is obtained using noisy sensor measurements. This data is quantized and transmitted to a (possibly untrustworthy) remote station through a public/unsecured communication network. We aim at keeping (part of) the state of the system private; however, because the network (and/or the remote station) might be unsecure, adversaries might have access to sensor data, which can be used to estimate the system state. To prevent such adversaries from obtaining an accurate state estimate, before transmission, we randomize quantized sensor data using additive random vectors, and send the corrupted data to the remote station instead. We design the joint probability distribution of these additive vectors (over a time window) to minimize the *mutual information* (our privacy metric) between some linear function of the system state (a desired private output) and the randomized sensor data for a *desired level of distortion*–how different quantized sensor measurements and distorted data are allowed to be. We pose the problem of synthesising the joint probability distribution of the additive vectors as a convex program subject to linear constraints. Simulation experiments are presented to illustrate our privacy scheme.**

***Index Terms*—Privacy; Dynamical Systems, Quantization, Mutual Information.**

## I. Introduction

In a hyperconnected world, scientific and technological advances have led to an overwhelming amount of user data being collected and processed by hundreds of companies over public networks. Companies mine this data to provide personalized services. However, these new technologies have also led to an alarming widespread loss of privacy in society and vulnerabilities within critical infrastructure – e.g., power, water, transportation. Depending on adversaries' resources, opponents may 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 of this 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 [1], [2]. Other examples are privacy loss caused by information sharing in distributed control systems and cloud computing [3]; the use of travel data for traffic estimation in intelligent transportation systems [4]; and data collection and sharing by the Internet-of-Things (IoT) [5], which is, most of the time, done without the user's informed consent. These privacy concerns show that there is an acute need for privacy preserving mechanisms capable of handling the new privacy challenges induced by an interconnected 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 security and privacy of Cyber-Physical Systems (CPSs) – engineered systems that integrate computation, networking, and dynamic physical components – see, e.g., [6]-[27].

In most engineering applications, information about the state of systems, say $X$, is obtained through sensor measurements. For collection, this information is usually quantized, and then encoded and sent to a remote station for signal processing and decision-making purposes through communication networks. Examples of such systems are numerous: water and electricity consumption meters, traffic monitoring systems, industrial control systems, and so on. If the communication network is public/unsecured and/or the remote station is untrustworthy, adversaries might access and estimate the state of the system. To avoid an accurate state estimation, before transmission, we randomize quantized sensor data using additive random vectors and send the corrupted data to the remote station instead. These vectors are designed to hide (as much as possible) the private part of the state $S$ – a desired private output modeled as some linear function of the system state, $S = DX$, for some deterministic matrix $D$. Note, however, that it is not desired to overly distort the original sensor data. We might change the data excessively for practical purposes. Hence, when selecting the additive distorting vectors, 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* [28], $I[\tilde{S}; Z]$, between a quantized version $\tilde{S}$ of the private output $S$ and the disclosed randomized sensor data $Z = G(Y)$ (over a finite time window). Mutual information $I[V; W]$ between two jointly distributed vectors, $V$ and $W$, is a measure of the statistical

Carlos Murguia is with the Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: c.g.murguia@tue.nl.

Iman Shames, Farhad Farokhi, and Dragan Nešić are with the Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, Australia, e-mails: iman.shames@unimelb.edu.au, farhad.farokhi@unimelb.edu.au, and dnesic@unimelb.edu.au.

Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA, e-mail: poor@princeton.edu.

Manuscript received October 25, 2019; revised October 26, 2019.

dependence between $V$ and $W$ [28]. We design the joint probability distribution of the distorting additive vectors to minimize $I[\tilde{S};Z]$ (over a time window), for a *desired level of distortion* – how different quantized sensor measurements and distorted data are allowed to be. We pose the problem of synthesising the joint probability distribution of these additive vectors as a convex program subject to linear constraints.

Using additive random vectors to increase privacy is common practice. In the context of privacy of databases, a popular approach is differential privacy [17], [29], where random noise is added to the response of queries so that private information stored in the database cannot be inferred. In differential privacy, because it provides certain privacy guarantees, Laplace noise is usually used [30]. 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 noise distribution achieving maximal privacy? This question has many possible answers depending on the particular privacy and distortion metrics being considered and the system configuration [18]-[21]. 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 [1], [2], [7], [8], [12], [16], [31], [32].

In general, if the data to be kept private follows continuous probability distributions, the problem of finding the optimal additive noise to maximize privacy (even without considering distortion) is hard to solve. 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 do have a solution, it is hard to find [7]. This problem has been addressed by imposing some particular structure on the considered distributions or assuming the data to be kept private is deterministic [7], [19], [20].

The authors in [19], [20] consider deterministic input 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. In [7], the authors 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. Moreover, they prove that, in the unconstrained case, the optimal noise distribution minimizing the Fisher information is Gaussian. This observation has also been made in [33] when using mutual information as a measure of privacy.

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 stochastic non-stationary data and thus existing tools do not fit this setting. Here, we identify two possibilities for addressing our problem: 1) we might inject continuous noise to sensor measurements, then quantize the sum, and send it over the unsecured/public network; or 2), the one considered here, quantize sensor measurements, randomize quantized sensor data using additive random vectors with discrete distributions, and send the randomized data over the network. As motivated above, to address the first option, even assuming deterministic sensor data, we have to impose some particular structure on the distributions of the additive noise; and, if sensor data is stochastic, the problem becomes hard to solve (sometimes even untractable). As we prove in this manuscript, if we select the second alternative, under some mild assumptions on the system dynamics and the additive distorting vectors, we can cast the problem of finding the optimal additive vectors as a constrained convex optimization. To the best of the authors knowledge, this problem has not been considered before as it is posed it here.

## II. Notation and Preliminaries

### A. Notation

The Euclidian norm in $\mathbb{R}^n$ is denoted by $||X||$, $||X||^2 = X^\top X$, where $^\top$ denotes transposition. The $n \times n$ identity matrix is denoted by $I_n$ or simply $I$ if $n$ is clear from the context. Similarly, $n \times m$ matrices composed of only ones and only zeros are denoted by $\mathbf{1}_{n\times m}$ and $\mathbf{0}_{n\times m}$, respectively, or simply $\mathbf{1}$ and $\mathbf{0}$ when their dimensions are clear. For positive definite (semidefinite) matrices, we use the notation $P > 0$ ($P \geq 0$); moreover, $P > Q$ ($P \geq Q$) means that the matrix $P - Q$ is positive definite (semidefinite). For any two matrices $A$ and $B$, the notation $A \otimes B$ (the Kronecker product [34]) stands for the matrix composed of submatrices $A_{ij}B$ , where $A_{ij}$, $i, j = 1, ..., n$, stands for the $ij$th entry of the $n \times n$ matrix $A$. Consider a discrete random vector $X$ with alphabet $\mathcal{X} = \{x_1, \ldots, x_N\}$, $x_i \in \mathbb{R}^m$, $m, N \in \mathbb{N}$, $i \in \{1, \ldots, N\}$, and probability mass function (pmf) $p(x) = \Pr[X = x]$, $x \in \mathcal{X}$, where $\Pr[B]$ denotes probability of event $B$. We denote its probability mass function by $p(x)$ rather than $p_X(x)$ to simplify notation. Thus, $p(x)$ and $p(y)$ refer to two different random vectors, and are, in fact, different probability mass functions, $p_X(x)$ and $p_Y(y)$, respectively. For a discrete stochastic process $X(k)$ taking values from the alphabet $\mathcal{X}_k \subset \mathbb{R}^m$, we denote its probability mass function as $p_k(x(k)) = \Pr[X(k) = x(k)]$, $x(k) \in \mathcal{X}_k$. For simplicity of notation, if the alphabet of $X(k)$ is time-invariant and finite, i.e., $\mathcal{X}_k = \mathcal{X} := \{x_1, \ldots, x_N\}$, $x_i \in \mathbb{R}^m$, for some finite $m, N \in \mathbb{N}$, we write its pmf at time $k$ as $p_k(x) = \Pr[X(k) = x]$, $x \in \mathcal{X}$. We denote by "Simplex" the probability simplex defined by $\sum_{x\in\mathcal{X}} p(x) = 1$, $p(x) \geq 0$ for all $x \in \mathcal{X}$. The notation $X \sim \mathcal{N}[\mu, \Sigma^X]$ means that $X \in \mathbb{R}^n$ is a normally distributed random vector with mean $E[X] = \mu \in \mathbb{R}^n$ and covariance matrix $E[(X-\mu)(X-\mu)^T] = \Sigma^X \in \mathbb{R}^{n\times n}$, where $E[a]$ denotes the expected value of the random vector $a$. We denote independence between two random vectors, $X$ and $Y$, as $X \perp\!\!\!\perp Y$. Finite sequences of vectors are written as $X^K := (X(1)^\top, \ldots, X(K)^\top)^\top \in \mathbb{R}^{Kn}$ and $X_{K_1}^{K_2} := (X(K_1)^\top, \ldots, X(K_2)^\top)^\top \in \mathbb{R}^{(K_2-K_1)n}$ with $K_2 > K_1$, $X(i) \in \mathbb{R}^n$, and $n, K, K_1, K_2 \in \mathbb{N}$. To avoid

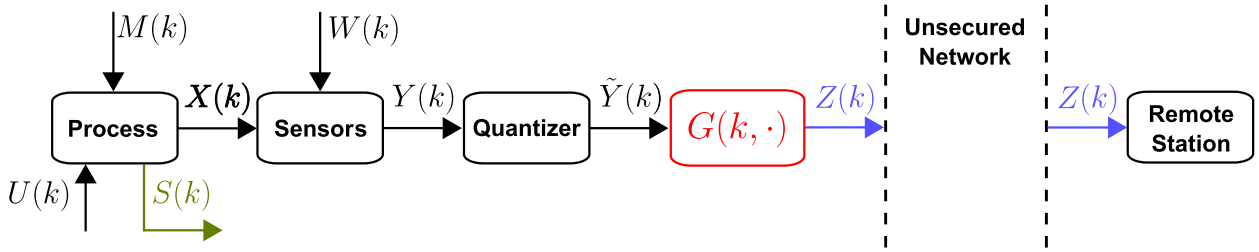


Fig. 1: System Configuration.

confusion, we denote powers of matrices as $(A)^K = A \cdots A$ ($K$ times) for $K > 0$, $(A)^0 = I$, and $(A)^K = \mathbf{0}$ for $K < 0$. Given two numbers $a$ and $b$, $b > 0$, the notation $a \bmod b$ stands for $a$ modulo $b$, and for a vector $a = (a_1, \ldots, a_n)^\top$, $a_i \in \mathbb{R}_{>0}$, $i = 1, \ldots, n$, $a \bmod b = (a_1 \bmod b, \ldots, a_n \bmod b)^\top$.

### B. Mutual Information

**Definition 1** [28] *Consider discrete random vectors, $S$ and $Z$, with joint probability mass function $p(s, z)$ and marginal probability mass functions, $p(s)$ and $p(z)$, respectively. Their mutual information $I[S; Z]$ is defined as the relative entropy between the joint distribution and the product distribution $p(s)p(z)$: $I[S; Z] := \sum_{s \in \mathcal{S}} \sum_{z \in \mathcal{Z}} p(s, z) \log \frac{p(s,z)}{p(s)p(z)}$.*

We use logarithms base 2. Then, mutual information is measured in bits [28].

**Remark 1** *Mutual information $I[S; Z]$ between two jointly distributed random vectors, $S$ and $Z$, measures the average amount of information (in bits) about $S$ contained in $Z$ (and vice versa). Then, $I[S; Z]$ can be regarded as a metric of the amount of information about $S$ that is leaked when disclosing $Z$. Mutual information has been widely used as privacy metric, not only for privacy of databases* [16],[35],[36]*, but also in a range of applications for dynamical systems* [1],[2],[22]-[24].

## III. Problem Formulation

### A. System Description, Quantization, and Stochastic Mappings

We study discrete-time stochastic systems of the form:

$$\begin{cases} X(k+1) = AX(k) + BU(k) + M(k), \\ Y(k) = CX(k) + W(k), \\ S(k) = DX(k), \end{cases} \tag{1}$$

with time-instants $k \in \mathbb{N}$, state $X \in \mathbb{R}^{n_x}$, $n_x \in \mathbb{N}$, output $Y \in \mathbb{R}^{n_y}$, $n_y \in \mathbb{N}$, *performance (private) output* $S \in \mathbb{R}^{n_s}$, $n_s \in \mathbb{N}$, stochastic disturbances $M \in \mathbb{R}^{n_x}$ and $W \in \mathbb{R}^{n_y}$, reference signal $U \in \mathbb{R}^{n_u}$, $n_u \in \mathbb{N}$, and matrices $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$, and $D \in \mathbb{R}^{n_s \times n_x}$. Matrix $D$ is full row rank. The perturbations $M(k)$ and $W(k)$ are i.i.d. multivariate Gaussian processes with $E[M(k)] = \mathbf{0}$, $E[W(k)] = \mathbf{0}$, and covariance matrices $\Sigma^M := E[M(k)M(k)^\top] \in \mathbb{R}^{n_x \times n_x}$, $\Sigma^M > 0$, and $\Sigma^W := E[W(k)W(k)^\top] \in \mathbb{R}^{n_y \times n_y}$, $\Sigma^W > 0$. The initial state $X(1)$ is assumed to be a Gaussian random vector with $E[X(1)] = \mu_1^X \in \mathbb{R}^{n_x}$ and covariance matrix $\Sigma_1^X := E[(X(1) - \mu_1^X)(X(1) - \mu_1^X)^\top] \in \mathbb{R}^{n_x \times n_x}$, $\Sigma_1^X > 0$. The processes $M(k)$, $k \in \mathbb{N}$ and $W(k)$, $k \in \mathbb{N}$, and the initial condition $X(1)$ are mutually independent. It is assumed that the matrices (vectors) $(A, B, C, D, \Sigma_1^X, \mu_1^X, \Sigma^M, \Sigma^W)$ are known, and the reference signal $U(k)$ is known and deterministic.

**Remark 2** *We introduce the notion of private outputs ($S(k) = DX(k)$) for generality. Private outputs provide freedom to choose the specific part of the state that must be kept private. For instance, if we seek randomizing mechanisms that maximize privacy of the complete state, $D = I_n$; if privacy of the measurable output is required, $D = C$; and if privacy of some other (not necessarily measurable) output, say $R(k) = FX(k)$, is required, $D = F$. For instance, let $X(k) := (P(k), V(k), A(k))^\top \in \mathbb{R}^3$ be a vector consisting on the velocity, position, and acceleration of a vehicle, respectively. Embedded sensors provide position measurements only, i.e., $C = (1, 0, 0)$. If position is to be kept private, $D = C$, if velocity is the private output, $D = (0, 1, 0)$, if acceleration must be kept private, $D = (0, 0, 1)$, and if privacy of the complete state is required, $D = I_3$.*

Sensor measurements $Y(k) \in \mathbb{R}^{n_y}$ are quantized using a *vector regular quantizer* [37] $Q_Y(Y(k), N_Y, \mathcal{C}, \mathcal{Y})$:

$$\tilde{Y}(k) = Q_Y(Y(k), N_Y, \mathcal{C}, \mathcal{Y}) := \begin{cases} y_1, & \text{if } Y(k) \in c_1, \\ \vdots & \\ y_{N_Y}, & \text{if } Y(k) \in c_{N_Y}, \end{cases} \tag{2}$$

with quantization levels $y_j \in \mathbb{R}^{n_y}$, $j = 1, 2, \ldots, N_Y$, quantization cells $c_j \subset \mathbb{R}^{n_y}$, $\bigcup_j c_j = \mathbb{R}^{n_y}$, $\bigcap_j c_j = \emptyset$, set of quantization cells $\mathcal{C} := \{c_1, \ldots, c_{N_Y}\}$, and set of quantization levels $\mathcal{Y} := \{y_1, \ldots, y_{N_Y}\}$. That is, the vector of quantized sensor measurements, $\tilde{Y}(k) = Q_Y(Y(k), N_Y, \mathcal{C}, \mathcal{Y}) \in \mathcal{Y}$, is parametrized by the quantization levels $y_j \in \mathbb{R}^{n_y}$, the quantization cells $c_j \subset \mathbb{R}^{n_y}$, $j = 1, \ldots, N_Y$, and the number of cells $N_Y \in \mathbb{N}$. Note that, if we know the multivariate probability density $f_k(y(k))$ of $Y(k)$ and the quantizer, we can always obtain the probability mass function $p_k(\tilde{y}(k))$ of $\tilde{Y}(k)$ by integrating $f_k(y(k))$ over the quantization cells $c_j$, $j = 1, \ldots, N_Y$. Moreover, the alphabet of the *discrete* multivariate random process $\tilde{Y}(k)$ is given by the set of quantization levels $\mathcal{Y}$. Because $\mathcal{Y}$ is time-invariant by construction, we write the pmf of $\tilde{Y}(k)$ as $p_k(\tilde{y})$, $\tilde{y} \in \mathcal{Y}$, i.e., $p_k(\tilde{y}) = \Pr[\tilde{Y}(k) = \tilde{y}]$ for all $\tilde{y} \in \mathcal{Y}$.

After $Y(k)$ is quantized, we pass $\tilde{Y}(k)$ through a stochastic mapping $G : \mathbb{N} \times \mathcal{Y} \to \mathcal{Y}$ characterized by the transition probabilities $p_k(z|\tilde{y}) = \Pr[Z(k) = z|\tilde{Y}(k) = \tilde{y}]$, $\tilde{y}, z \in \mathcal{Y}$, i.e., $Z(k) = G(k, \tilde{Y}(k)) \in \mathcal{Y}$, see Figure 2. The vector

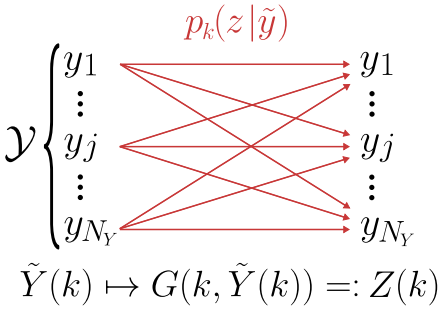


Fig. 2: Probabilistic Mapping.

$Z(k)$ is transmitted over an unsecured communication network to a (possibly untrustworthy) remote station, see Figure 1. Note that, even by passing $\tilde{Y}(k)$ through $G(k,\cdot)$ before transmission, information about the private output $S(k)$ is directly accessible through $Z(k)$ at the unsecured network. Here, we aim at finding the mapping $G(k,\cdot)$ (the transition probabilities $p_k(z|\tilde{y})$) that minimizes this information leakage. Note, however, that we do not want to make $\tilde{Y}(k)$ and $Z(k)$ overly different either. By passing $\tilde{Y}(k)$ through $G(k,\cdot)$, we might *distort* $\tilde{Y}(k)$ excessively for practical purposes. Hence, when designing the distribution $p_k(z|\tilde{y})$, we need to consider the trade-off between *privacy* and *distortion*.

**Remark 3** *The framework that we propose aims to increase privacy from the system designer point of view without assuming any particular use of the disclosed data $Z(k)$ at the remote station. The vector $Z(k)$ could be used for any real-time application that does not incur in feedback to the system dynamics. In general, applications of this sort are related to remote decision making and monitoring. Particular examples are predictive maintenance, fault-detection, state estimation, and collective decision making.*

The setting that we consider is a standard privacy-utility tradeoff problem. As we increase privacy (by distorting the original data), the utility of the data (the use of the data at the remote station) would decrease for most applications. The setting that we propose seeks the distorting mechanism that achieve the best privacy-utility tradeoff. This framework makes sense for the potential applications that we introduced in Remark 3, i.e., for remote decision making and monitoring applications. Indeed, if the performance of the application is so critical that no distortion can be tolerated; then, this scheme is not suitable for that particular application. But if there is a small amount of distortion that the application can tolerate (e.g., in fault-detection, a small increase in the false-alarm rate, or in estimation problems, a small increase in the variance of the asymptotic estimation error), then, our scheme looks for the best randomizing mechanism that yields that small amount of allowable distortion while maximizing privacy.

**Remark 4** *In the problem setting introduced above, we are assuming that either the communication network or the remote station might be untrustworthy. That is, we aim to prevent adversaries eavesdropping on the channel and those at the remote station from accurately estimating the private output $S(k)$. Standard encryption techniques would not work in this setting as, after decryption, adversaries at the remote station could accurately estimate the private output. There are applications where Partially Homomorphic Encryption (PHE) has been used to do computations over encrypted data and avoid decrypting at the remote station (e.g., control and optimization applications* [38]-[41]*). Note, however, that applications for which PHE can be used are still limited. The issue with using PHE is that the application at the remote station must work on the encrypted data. Meaning that the algorithm that we would normally use without encryption has to be redesigned to work with data in the encrypted domain (usually finite rings of integers). This is quite challenging as it is not always possible to adapt algorithms in such a way. Moreover, techniques based on PHE still suffer from insider attacks, i.e., if someone from the inside distributes the secret key for decryption, adversaries could access undistorted data directly. The scheme that we propose here avoids this by "destroying" the private information from the shared distorted data. That is, we inject uncertainty in the direction of the private output $S(k)$ while distorting the original data $\tilde{Y}(k)$ as little as possible. By doing this, we decrease the estimation performance of any adversary (independently of the estimation algorithm they use) as information about $S(k)$ is simply not there, or phrased differently, it is optimally obfuscated.*

### B. Adversarial Model

We consider worst-case adversaries that eavesdrop data at the communication network and/or the remote station. They do not only have access to all $Z(k)$, $k \in \mathcal{K}$, but also know the dynamics, quantizer, reference, and stochastically in the system, i.e., matrices $(A, B, C, D, \Sigma_1^X, \mu_1^X, \Sigma^M, \Sigma^W)$ and the deterministic reference sequence $U(k)$, $k \in \mathcal{K}$, are perfectly known by the adversary.

In practice, most of the times, actual adversaries would not have all the capabilities that we assume in this section. However, if we maximize privacy under worst-case adversaries, we ensure that adversaries with less capabilities perform even worse (or equal at most). That is, the privacy guarantees that the scheme provides under worse-case adversaries would hold for weaker ones. These guarantees provide an upper bound on the information that could actually be leaked for any adversary. Assuming worst-case adversaries is common practice in security and privacy of cyber-physical systems, see, e.g., [5], [9], [11], [16], [26], [31], [42]-[45]. The uncertainty induced by the potential limited knowledge of adversaries is usually pushed aside, and the privacy mechanism focuses on the fundamental privacy leakage from the distorted data generated by the privacy mechanism.

### C. Metrics and Problem Formulation

For given time horizon $K \in \mathbb{N}$, the aim of our privacy scheme is to make inference of the sequence of private vectors $S^K = (S(1)^\top, \dots, S(K)^\top)^\top$ from the distorted sequence $Z^K = (Z(1)^\top, \dots, Z(K)^\top)^\top$ as hard as possible without distorting $\tilde{Y}^K = (\tilde{Y}(1)^\top, \dots, \tilde{Y}(K)^\top)^\top$ excessively. As *distortion metric*, we use the stacked mean squared error:

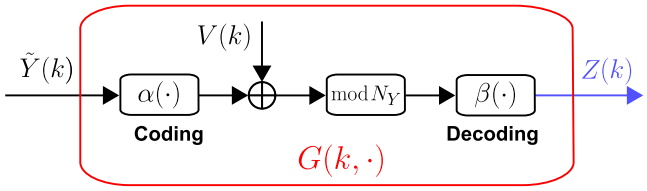


Fig. 3: Schematic diagram of the mapping $G(k,\cdot)$.

$E[||Z^K - \tilde{Y}^K||^2]$; and, as *privacy metrics*, the mutual information $I[\tilde{S}^K; Z^K]$, where $\tilde{S}^K = (\tilde{S}(1)^\top, \ldots, \tilde{S}(K)^\top)^\top$ and $\tilde{S}(i) \in \mathbb{R}^{n_s}$ denotes a quantized version of the private output $S(i)$, $i = 1, \ldots, K$. A natural question arises here: Why don't we directly use $I[S^K; Z^K]$ as privacy metric? Note that, because $X(1)$, $M(k)$, and $W(k)$ are Gaussian, entries of $S^K$ are jointly normally distributed. Moreover, $Z^K$ is a discrete random vector whose conditional distribution depends on the quantizer, $Q_Y(Y(k), N_Y, \mathcal{C}, \mathcal{Y})$, and the density of $Y(k)$. Hence, $I[S^K; Z^K]$ would denote the mutual information between continuous and discrete random vectors. This type of mutual information is not always well defined. There are some hybrid definitions available in the literature that have been used to estimate mutual information between continuous and discrete data sets [46]-[50]. However, we prefer to avoid using hybrid formulations because they lead to complicated expressions that are hard to evaluate. We work directly with the definition of mutual information introduced by Shannon for discrete random vectors (given in Definition 1) by discretizing the density of $S(k)$ to compute $I[\tilde{S}^K; Z^K]$. This formulation allows us to obtain suboptimal distorting mechanisms and, as we decrease the size of the discretization cells of $\tilde{S}(k)$ to zero, we recover $I[S^K, Z^K]$ [28, Chapter 8]. We acknowledge that optimality is sacrificed by working with approximations $I[\tilde{S}^K, Z^K]$ of $I[S^K, Z^K]$, but we gain tractability in the sense that we can always cast and solve the synthesis optimization problem when working with $I[\tilde{S}^K, Z^K]$.

Summarizing the above discussion, we aim at minimizing $I[\tilde{S}^K; Z^K]$ subject to $E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon_K$, for a desired level of distortion $\epsilon_K \in \mathbb{R}_{>0}$, using as decision variables the conditional probability mass function $p(z^K|\tilde{y}^K) = \Pr[Z^K = z^K|\tilde{Y}^K = \tilde{y}^K]$, where $z^K, \tilde{y}^K \in \mathcal{Y}^K := \mathcal{Y} \times \ldots \times \mathcal{Y}$ ($K$ times) and elements of $\mathcal{Y}^K$ belong to $\mathbb{R}^{Kn_y}$, i.e., since $\tilde{Y}(k)$ and $Z(k)$ have alphabet $\mathcal{Y}$, $k \in \{1, \ldots, K\}$, the stacked vectors $\mathcal{Y}^K$ and $Z^K$ have alphabet $\mathcal{Y}^K$.

**Remark 5** *The number of optimization variables $p(z^K|\tilde{y}^K)$ depends on the number of quantization levels $N_Y$ and the horizon $K$. Both stacked vectors $Y^K$ and $Z^K$ take values from the alphabet $\mathcal{Y}^K = \mathcal{Y} \times \ldots \times \mathcal{Y}$ ($K$ times) and because $\mathcal{Y}$ is the set of $N_Y$ quantization levels, $\mathcal{Y}^K$ has $(N_Y)^K$ elements. We aim at computing an optimal transition probability $p(z^K|\tilde{y}^K)$ from each element of the alphabet of $Y^K$ to every element of the alphabet of $Z^K$. Therefore, we have $(N_Y^K)^2$ optimization variables to minimize $I[\hat{S}^K(\tilde{Y}^K); \hat{S}^K(Z^K)]$ and $I[\tilde{S}^K; Z^K]$. This is already a large-scale optimization problem for quantizers with moderate resolution and fairly small $K$. For instance, a 3-bit quantizer ($N_Y = 8$) and a horizon of $K = 5$ requires more than a billion variables. A possible solution to this dimensionality problem is to impose some structure on $p(z^K|\tilde{y}^K)$ to reduce the number of variables. That is, we could set some of the transition probabilities to a constant known value a priori, and use the remaining ones to minimize the mutual information. In what follows, we propose a systematic way to achieve this using an independent additive random process $V(k)$. We impose this structure on the probabilistic mapping $G(k,\cdot)$ (see Figure 1) so that the number of variables in $p(z^K|\tilde{y}^K)$ is reduced from $(N_Y)^{2K}$ to $(N_V)^K$, where $N_V$ denotes the cardinality of the alphabet of additive process $V(k)$ to be designed. Indeed, there is a trade-off between the privacy level that can be achieved (large $N_V$ implies more degrees of freedom to minimize mutual information) and the complexity of the optimization problem to be solved to obtain the privacy mechanism (as increasing the number variables increases the complexity of the problem).*

The proposed probabilistic mapping $G(k,\cdot)$ consists of the following three objects: 1) a coding function $\alpha : \mathcal{Y} \to \{0, 1, \ldots, N_Y - 1\} =: \bar{\mathcal{Y}}$ that indexes each element of $\mathcal{Y}$; 2) a discrete random process $V(k)$, independent of $\tilde{Y}(k)$ for all $k \in \mathbb{N}$, with alphabet $\mathcal{V} := \{0, 1, \ldots, N_V - 1\}$, $N_V \in \mathbb{N}$, and probability mass function $q_k(v)$, $v \in \mathcal{V}$ (the pmf of $V(k)$ is denoted as $q_k(v)$ rather than $p_k(v)$ because, in what follows, we use $q_k(v)$ as optimization variables and we want to clearly distinguish $q_k(v)$ from other probability mass functions); and 3) a decoding function $\beta : \{0, 1, \ldots, N_Y - 1\} \to \mathcal{Y}$. We characterize each of these objects before introducing the mapping $G(k,\cdot)$. The indexing (coding) function $\alpha : \mathcal{Y} \to \bar{\mathcal{Y}}$ is defined as

$$\alpha(\zeta) := \begin{cases} 0, & \text{if } \zeta = y_1, \\ \vdots & \\ N_Y - 1, & \text{if } \zeta = y_{N_Y}. \end{cases} \tag{3}$$

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

$$\beta(\xi) := \begin{cases} y_1, & \text{if } \xi = 0, \\ \vdots & \\ y_{N_Y}, & \text{if } \xi = N_Y - 1. \end{cases} \tag{4}$$

Note that $\beta(\alpha(\zeta)) = \zeta$ and $\alpha(\beta(\xi)) = \xi$. We construct the mapping $G : \mathbb{N} \times \mathcal{Y} \to \mathcal{Y}$, $\tilde{Y}(k) \mapsto G(k, \tilde{Y}(k))$, combining (3) and (4) as follows

$$Z(k) = G(k, \tilde{Y}(k)) := \beta\big((\alpha(\tilde{Y}(k)) + V(k)) \bmod N_Y\big). \tag{5}$$

In Figure 3, we depict a schematic diagram of the mapping $G(k,\cdot)$. Since $\alpha(\cdot)$ and $\beta(\cdot)$ are fixed injective functions, we can only use the probability mass function $q_k(v) = \Pr[V(k) = v]$, $v \in \mathcal{V} = \{0, 1, \ldots, N_V - 1\}$, to minimize $I[\hat{S}^K(\tilde{Y}^K); \hat{S}^K(Z^K)]$ and $I[\tilde{S}^K; Z^K]$. For given time horizon $K \in \mathbb{N}$, let $V^K = (V(1)^\top, \ldots, V(K)^\top)^\top$, and denote its probability mass function as $q(v^K) = \Pr[V^K = v^K]$, where

$v^K \in \mathcal{V}^K := \mathcal{V}\times\ldots\times\mathcal{V}$ ($K$ times) and elements of $\mathcal{V}^K$ belong to $\mathbb{R}^K$, i.e., since $V(k)$ has alphabet $\mathcal{V}$, $k \in \{1,\ldots,K\}$, the stacked vector $V^K$ has alphabet $\mathcal{V}^K$. In what follows, we formally present the optimization problems we seek to address.

**Problem 1** *Given the system dynamics* (1)*, sensor quantizer* (2)*, time horizon $K \in \mathbb{N}$, desired distortion level $\epsilon_K \in \mathbb{R}_{\geq 0}$, quantized version $\tilde{S}(k)$ of $S(k)$, $k \in \{1,\ldots,K\}$, and the probabilistic mapping* (3)-(5)*, find the probability mass function $q(v^K)$ solution of the following optimization problem:*

$$\begin{cases} \min\limits_{q(v^K)} \; I[\tilde{S}^K; Z^K], \\ \quad \text{s.t. } E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon_K, \\ \qquad V^K \perp\!\!\!\perp \tilde{Y}^K, \text{ and } q(v^K) \in \text{ Simplex}. \end{cases} \tag{6}$$

### *D. On the Distortion Constraint and Linear Estimators*

Problem 1 is posed in terms of a general distortion constraint that upper bounds the second moment of $(Z^K - \tilde{Y}^K)$. As we have motivated before (see Remark 3 and the discussion below it), this general formulation seeks to avoid assuming a particular application of the transmitted data at the remote station. If a particular application was considered (e.g., state estimation, fault detection, etc.), we would have to cast the problem in terms of the performance degradation *on that particular application*. Then, for every application, we would have a different performance criteria – leading to an endless number of different formulations.

Arguably, for the class of linear systems considered in this manuscript, and most applications at the remote station (for this class), performance degradation induced by the privacy mechanism can be written (or upper bounded) in terms of the distortion on $(Z^K - \tilde{Y}^K)$. To illustrate the later, we provide a brief application example using a class of linear state estimators. Consider system (1) and the following two recursive estimators, one driven by the original quantized sensors, $\tilde{Y}(k)$, and the other driven by the distorted $Z(k)$:

$$\hat{X}_Y(k+1) = A\hat{X}_Y(k) + BU(k) + L(\tilde{Y}(k) - C\hat{X}_Y(k)), \tag{7}$$

$$\hat{X}_Z(k+1) = A\hat{X}_Z(k) + BU(k) + L(Z(k) - C\hat{X}_Z(k)), \tag{8}$$

with corresponding estimator state $\hat{X}_Y(k), \hat{X}_Z(k) \in \mathbb{R}^{n_x}$, initial condition $\hat{X}_Y(1) = \hat{X}_Z(1) = E[X(1)] = \mu_1^X$, reference signal $U(k) \in \mathbb{R}^{n_u}$ and matrices $(A, B, C)$ as in (1), and gain matrix $L \in \mathbb{R}^{n_x \times n_y}$. Matrix $L$ is designed to achieve the desired estimation performance (e.g., internal stability, disturbances attenuation, and convergence rate). Estimators (7) and (8) are identical in structure but driven by different sequences. This class of estimators is referred in the literature as Luenberger observers [51]. Here, we present a prescriptive analysis where we compare the performance between the estimators, for $k \in \mathcal{K}$, and show that the performance degradation induced by the use of the distorted $Z(k)$, instead of the original $\tilde{Y}(k)$, is upper bounded by a convex function of $E[||Z^K - \tilde{Y}^K||^2]$.

Define the estimation errors $e^Y(k) := X(k) - \hat{X}^Y(k)$ and $e^Z(k) := X(k) - \hat{X}^Z(k)$. Given the system and estimators dynamics, (1) and (7)-(8), it is easy to verify that the estimation errors evolve as

$$e_Y(k+1) = \bar{A}e_Y(k) + M(k) - L(W(k) + \delta^Y(k)),$$
$$e_Z(k+1) = \bar{A}e_Y(k) + M(k) - L(W(k) + \delta^Y(k) - \delta^Z(k)),$$

with $\bar{A} := A - LC$, quantization error $\delta^Y(k) := Y(k) - \tilde{Y}(k)$, and $\delta^Z(k) := Z(k) - \tilde{Y}(k)$. Matrix $L$ is selected such that $\rho(A - LC) < 1$, where $\rho(\cdot)$ denotes spectral radius [52]. Such a matrix $L$ exists when system (1) is *detectable* [51]. The condition $\rho(A - LC) < 1$ guarantees that, if disturbances $(M(k), W(k), \delta^Y(k), \delta^Z(k))$ are equal to zero (or converge to zero asymptotically) $\lim_{k\to\infty} e^Y(k) = \lim_{k\to\infty} e^Z(k) = 0$ (internal stability) [51]. For non-vanishing disturbances, the estimation errors do not converge to zero. They converge to a compact invariant set, if disturbances are uniformly bounded on compact sets, and to a stationary process, if disturbances are stationary random processes [51].

The performance of (7) and (8) is fully characterized by the estimation errors $e^Y(k)$ and $e^Z(k)$. Thus, to compare their performance, we introduce the auxiliary vector $\Delta(k) := E[e^Y(k) - e^Z(k)]$. Given the difference equations for $e^Y(k)$ and $e^Z(k)$ introduced above, it is easy to verify that $\Delta(k)$ evolves as $\Delta(k+1) = \bar{A}\Delta(k) + LE[\delta^Z(k)]$, with $\Delta(1) = \mathbf{0}$. Hence, the general solution of $\Delta(k)$ is given by

$$\Delta(k) = \sum_{i=0}^{k-2} \bar{A}^i L E[\delta^Z(k-i-1)], \; k > 1. \tag{9}$$

Note that $\delta^Z(k) = \mathbf{0}$, $k \in \mathbb{N}$, implies $\Delta(k) = \mathbf{0}$, $k \in \mathbb{N}$, because $\Delta(1) = \mathbf{0}$. That is, the auxiliary vector is nonzero only if there is distortion due to the privacy mechanism. The aim of this subsection is to show that an upper bound $\epsilon_K$ on $E[||Z^K - \tilde{Y}^K||^2]$ translates into an upper bound on the estimation performance degradation (i.e., into an upper bound on $||\Delta(k)||$). To accomplish this, we seek to upper bound $||\Delta^{K+1}||$ by a convex function of $E[||Z^K - \tilde{Y}^K||^2]$.

For $k \in \mathcal{K}$, equation (9) can be written in terms of $Z^K$, $\tilde{Y}^K$, and $\Delta^{K+1} = (\Delta(1)^\top, \ldots, \Delta(K+1)^\top)^\top$ as

$$\Delta(K+1) = \Theta_K(I_{K-1} \otimes L)E[Z^K - \tilde{Y}^K], \tag{10}$$

with

$$\Theta_K = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ I & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \bar{A} & I & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (\bar{A})^{K-2} & (\bar{A})^{K-3} & (\bar{A})^{K-4} & \cdots & I \end{bmatrix}.$$

It follows that

$$\begin{aligned} ||\Delta(K+1)|| &\overset{\text{(a)}}{\leq} \vartheta||E[Z^K - \tilde{Y}^K]||, \\ &\overset{\text{(b)}}{\leq} \vartheta E[||Z^K - \tilde{Y}^K||], \\ &\overset{\text{(c)}}{\leq} \vartheta E[||Z^K - \tilde{Y}^K||^2]^{1/2}, \\ &\overset{\text{(d)}}{\leq} \vartheta\sqrt{\epsilon_K}, \end{aligned}$$

where $||\cdot||$ denotes Euclidian norm for vectors and spectral norm for matrices [52], $\vartheta := ||\Theta_K(I_{K-1}\otimes L)||$, (a) follows from properties of matrix norms, (b) and (c) from Jensen's inequality [53], and (d) from (6). Therefore, by solving (6) with the distortion constraint, $E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon_K$, we enforce that the estimation performance is degraded by at most $\vartheta\sqrt{\epsilon_K}$, with $\vartheta$ being a positive constant that is independent of the privacy mechanism. Similar prescriptive analyses can be performed for other applications at the remote station (e.g., fault detection and isolation, distributed decision-making and optimization, predictive maintenance, classification, filtering, etc). By casting (6) in terms of $E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon_K$, we cover a variety of applications as their performance degradation can be written (or upper bounded) in terms of $E[||Z^K - \tilde{Y}^K||^2]$. Hence, hereafter, we focus on Problem 1 as it is casted in (6).

## IV. Solution to Problem 1

Consider the cost function $I[\tilde{S}^K; Z^K]$ for some quantized version $\tilde{S}^K$ of the stacked private output $S^K$, and the system dynamics (1). At time $k$, $S(k)$ can be written in terms of the initial condition, the stacked reference $U^{k-1}$, and the stacked disturbance $M^{k-1}$ as $S(k) = DA^{k-1}X(1) + \sum_{i=0}^{k-2} DM(k-1-i) + DU(k-1-i)$, i.e., $S(k)$ is the sum of independent Gaussian random vectors (and thus it is also Gaussian). It follows that the support of $S(k)$ is the whole $R^{n_s}$. To discretize the density of $S(k)$, we divide its support into a finite set of cells. Let $\mathcal{H} := \{h_1, \ldots, h_{N_S}\}$ denote a set of $N_S \in \mathbb{N}$ quantization cells satisfying: $h_j \subset \mathbb{R}^{n_s}$, $\bigcup_j h_j = \mathbb{R}^{n_s}$, $\bigcap_j h_j = \emptyset$. Associated with each cell $h_j$, we introduce the corresponding quantization level $s_j \in h_j \subset \mathbb{R}^{n_s}$, i.e., $s_j$ denotes a $n_s$-dimensional point in the interior of $h_j$, $j \in \{1, \ldots, N_S\}$. We collect all the quantization levels into the set $\mathcal{S} := \{s_1, \ldots, s_{N_S}\}$. Then, the quantized private output, $\tilde{S}(k) \in \mathbb{R}^{n_s}$, can be written as

$$\tilde{S}(k) = Q_S(S(k), N_S, \mathcal{H}, \mathcal{S}) := \begin{cases} s_1, & \text{if } S(k) \in h_1, \\ & \vdots \\ s_{N_S}, & \text{if } S(k) \in h_{N_S}. \end{cases} \tag{11}$$

Denote the stacked vector of quantized private outputs as $\tilde{S}^K = (\tilde{S}(1)^\top, \ldots, \tilde{S}(K)^\top)^\top \in R^{Kn_s}$. Note that if we know the joint probability density function of $S^K$, once we have fixed the quantizer $Q_S(\cdot)$, we can obtain the probability mass function $p(\tilde{s}^K) = \Pr[\tilde{S}^K = \tilde{s}^K]$, $\tilde{s}^K \in \mathcal{S}^K := \mathcal{S} \times \ldots \times \mathcal{S}$ ($K$ times) by integrating the density of $S^K$ over the cells in $\mathcal{H}$. For the discrete random vector $\tilde{S}^K$, we have a well defined mutual information $I[\tilde{S}^K; Z^K]$. By Definition 1, the cost $I[\tilde{S}^K; Z^K]$ is a function of $p(\tilde{s}^K, z^K)$, and the marginals $p(\tilde{s}^K)$ and $p(z^K)$. However, to minimize $I[\tilde{S}^K; Z^K]$, we need to write it in terms of $q(v^K)$ (our design variables). Notice that, if the joint density of $S^Y$ and $Y^K$ is a non-degenerate Gaussian, we can numerically compute $p(\tilde{s}^K|\tilde{y}^K)$ for any $\tilde{s}^K \in \mathcal{S}^K$ and $\tilde{y}^K \in \mathcal{Y}^K$; and that, because (by construction) $\tilde{S}(k)$ and $Z(k)$ are conditionally independent given $\tilde{Y}(k)$, $\tilde{S}^K$ and $Z^K$ are conditionally independent given $\tilde{Y}^K$. The latter implies that $p(\tilde{s}^K, z^K) = \sum_{\tilde{y}^K \in \mathcal{Y}^K} p(\tilde{y}^K) p(\tilde{s}^K|\tilde{y}^K) p(z^K|\tilde{y}^K)$. Then, for given $p(\tilde{y}^K)$ and $p(\tilde{s}^K|\tilde{y}^K)$, we can write the cost $I[\tilde{S}^K; Z^K]$ in terms of $p(z^K|\tilde{y}^K)$ and $q(v^K)$. In the following lemma, we write the cost function $I[\tilde{Y}^K; Z^K]$ in terms of $q(v^K)$, and prove that it is convex in $q(v^K)$. Before stating the lemma, we need some extra notation. For any $\zeta^K = (\zeta(1)^\top, \ldots, \zeta(K)^\top)^\top \in \mathcal{Y}^K$, $\zeta(i) \in \mathcal{Y}$, $i \in \{1, \ldots, K\}$, define the stacked indexing function $\bar{\alpha} : \mathcal{Y}^K \to \bar{\mathcal{Y}}^K$, with $\bar{\mathcal{Y}}^K := \bar{\mathcal{Y}} \times \cdots \times \bar{\mathcal{Y}}$ ($K$ times) and $\bar{\mathcal{Y}} = \{0, \ldots, N_Y - 1\}$, as:

$$\bar{\alpha}(\zeta^K) := (\alpha(\zeta(1)), \ldots, \alpha(\zeta(K)))^\top, \tag{12}$$

where $\alpha(\cdot)$ is the indexing function defined in (3).

**Lemma 1** *$I[\tilde{S}^K; Z^K]$ is a convex function of $q(v^K)$ for given $p(\tilde{s}^K|\tilde{y}^K)$ and $p(\tilde{y}^K)$, and is written as:*

$$I[\tilde{S}^K; Z^K] = \sum_{z^K \in \mathcal{Y}^K} \sum_{\tilde{s}^K \in \mathcal{S}^K} p(\tilde{s}^K) p(z^K|\tilde{s}^K) \log \frac{p(z^K|\tilde{s}^K)}{p(z^K)}, \tag{13a}$$

$$p(z^K|\tilde{s}^K) = \sum_{\tilde{y}^K \in \mathcal{Y}^K} p(\tilde{y}^K|\tilde{s}^K) p(z^K|\tilde{y}^K), \tag{13b}$$

$$p(z^K) = \sum_{\tilde{s}^K \in \mathcal{S}^K} \sum_{\tilde{y}^K \in \mathcal{Y}^K} p(\tilde{s}^K) p(\tilde{y}^K|\tilde{s}^K) p(z^K|\tilde{y}^K), \tag{13c}$$

$$p(z^K|\tilde{y}^K) = q\big((\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)) \bmod N_Y\big), \tag{13d}$$

*where $\bar{\alpha} : \mathcal{Y}^K \to \bar{\mathcal{Y}}^K$ denotes the stacked indexing function defined in* (12)*.*

***Proof***: See Appendix A.

By Lemma 1, $I[\tilde{S}^K; Z^K]$ is convex in our decision variables $q(v^K)$, for given $p(\tilde{s}^K)$ and $p(\tilde{y}^K|\tilde{s}^K)$. Then, if the distortion constraint, $E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon_K$, is convex in $q(v^K)$, and we know $p(\tilde{s}^K)$ and $p(\tilde{y}^K|\tilde{s}^K)$, we could minimize $I[\tilde{S}^K; Z^K]$ efficiently using off-the-shelf optimization algorithms.

**Lemma 2** *$E[||Z^K - \tilde{Y}^K||^2]$ is a linear function of $q(v^K)$ for given $p(\tilde{y}^K)$, and can be written as follows:*

$$E[||Z^K - \tilde{Y}^K||^2] = \sum_{\tilde{y}^K \in \mathcal{Y}^K} \sum_{z^K \in \mathcal{Y}^K} p(z^K, \tilde{y}^K) \big(z^K - \tilde{y}^K\big)^2, \tag{14a}$$

$$p(z^K, \tilde{y}^K) = q\big((\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)) \bmod N_Y\big) p(\tilde{y}^K). \tag{14b}$$

***Proof***: See Appendix B.

By Lemma 1 and Lemma 2, the cost, $I[\tilde{S}^K; Z^K]$, and constraint, $E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon_K$, are parametrized by $p(\tilde{s}^K)$ and $p(\tilde{y}^K|\tilde{s}^K)$. To obtain these distributions, we need to integrate the density of $S^K$ and the joint density of $(S^K, Y^K)$ (if they are not degenerate) over the quantization cells. By lifting the system dynamics (1) over $\{1, \ldots, K\}$, we can write the stacked vector $((Y^K)^\top, (S^K)^\top)^\top \in \mathbb{R}^{K(n_s+n_y)}$ as

$$\begin{aligned} \begin{bmatrix} Y^K \\ S^K \end{bmatrix} = \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} F_K X(1) + \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} T_K M^{K-1} \\ + \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} L_K U^{K-1} + \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} W^K, \end{aligned} \tag{15}$$

with $L_K := T_K(I_{K-1} \otimes B)$, $\tilde{C}_K := I_K \otimes C$, $\tilde{D}_K := I_K \otimes D$,

and

$$\begin{cases} F_K := \begin{bmatrix} I & A^\top & \dots & (A^\top)^{K-1} \end{bmatrix}^\top, \\ T_K := \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ I & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ A & I & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (A)^{K-2} & (A)^{K-3} & (A)^{K-4} & \cdots & I \end{bmatrix}, \end{cases} \tag{16}$$

**Lemma 3**

$$\begin{pmatrix} Y^K \\ S^K \end{pmatrix} \sim \mathcal{N}\left[ \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} F_K \mu_1^X + \begin{pmatrix} \tilde{C}_K \\ \tilde{D}_K \end{pmatrix} L_K U^{K-1}, \Sigma_K^{Y,S} \right],$$

*with covariance* $\Sigma_K^{Y,S} \in \mathbb{R}^{K(n_s+n_y)\times K(n_s+n_y)}$, $\Sigma_K^{Y,S} > 0$:

$$\begin{aligned} \Sigma_K^{Y,S} :=& [\begin{smallmatrix} I \\ \mathbf{0} \end{smallmatrix}](I_K \otimes \Sigma^W)[\begin{smallmatrix} I \\ \mathbf{0} \end{smallmatrix}]^\top + \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} F_K \Sigma_1^X F_K^\top \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix}^\top \\ &+ \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} T_K (I_{K-1} \otimes \Sigma^M) T_K^\top \begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix}^\top. \end{aligned} \tag{17}$$

***Proof***: See Appendix C.

To obtain the density of $S^K$, we marginalize the joint density $\mathcal{N}[\begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} F_K \mu_1^X + \begin{pmatrix} \tilde{C}_K \\ \tilde{D}_K \end{pmatrix} L_K U^{K-1}, \Sigma_K^{Y,S}]$ over $Y^K$. We give this density in the following corollary of Lemma 3.

**Corollary 1** *$S^K \sim \mathcal{N}[\tilde{D}_K F_K \mu_1^X + \tilde{D}_K L_K U^{K-1}, \Sigma_K^S]$ with covariance $\Sigma_K^S = \begin{pmatrix} \mathbf{0} \\ I_{n_s} \end{pmatrix}^\top \Sigma_K^{Y,S} \begin{pmatrix} \mathbf{0} \\ I_{n_s} \end{pmatrix} \in \mathbb{R}^{Kn_s \times Kn_s}$.*

**Remark 6** *Having the joint density of $Y^K$ and $S^K$ allows us to compute, for any set of quantization cells $\{c_l, \dots, c_i\}$, $l, \dots, i \in \{1, \dots, N_Y\}$ and $\{h_r, \dots, h_j\}$, $r, \dots, j \in \{1, \dots, N_S\}$, $\Pr[S(1) \in h_r, \dots, S(K) \in h_j]$ and $\Pr[Y(1) \in c_l, \dots, Y(K) \in c_i | S(1) \in h_r, \dots, S(K) \in h_j]$. By definition, for $\tilde{s}(1), \dots, \tilde{s}(K) \in \mathcal{S}$, and $\tilde{y}(1), \dots, \tilde{y}(K) \in \mathcal{Y}$, $p(\tilde{y}^K | \tilde{s}^K) = \Pr[\tilde{Y}(1) = \tilde{y}(1), \dots, \tilde{Y}(K) = \tilde{y}(K) | \tilde{S}(1) = \tilde{s}(1), \dots, \tilde{S}(K) = \tilde{s}(K)]$. Moreover, by construction of the quantizers, we have:*

$$\begin{aligned} &\Pr[\tilde{Y}(1) = y_l, \dots, \tilde{Y}(K) = y_i | \tilde{S}(1) = s_r, \dots, \tilde{S}(K) = s_j] \\ &= \Pr[Y(1) \in c_l, \dots, Y(K) \in c_i | S(1) \in h_r, \dots, S(K) \in h_j] \\ &= \frac{\Pr[Y(1) \in c_l, \dots, Y(K) \in c_i, S(1) \in h_r, \dots, S(K) \in h_j]}{\Pr[S(1) \in h_r, \dots, S(K) \in h_j]}, \end{aligned}$$

*for any set of quantization levels $\{y_l, \dots, y_i\}$, $l, \dots, i \in \{1, \dots, N_Y\}$ and $\{h_r, \dots, h_j\}$, $r, \dots, j \in \{1, \dots, N_S\}$. Therefore, by integrating the joint density $\mathcal{N}[\begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} F_K \mu_1^X + \begin{pmatrix} \tilde{C}_K \\ \tilde{D}_K \end{pmatrix} L_K U^{K-1}, \Sigma_K^{Y,S}]$ of $((Y^K)^\top, (S^K)^\top)^\top$ over the quantization cells, we can compute the probability mass functions $p(\tilde{y}^K | \tilde{s}^K)$ and $p(\tilde{s}^K)$.*

In what follows, we pose the nonlinear program for solving Problem 1.

**Theorem 1** *Given the system dynamics* (1)*, sensor quantizer* (2)*, private output quantizer* (11)*, probabilistic mapping* (3)-(5)*, time horizon $K \in \mathbb{N}$, desired distortion level $\epsilon_K \in \mathbb{R}_{\geq 0}$, joint probability density $\mathcal{N}[\begin{bmatrix} \tilde{C}_K \\ \tilde{D}_K \end{bmatrix} F_K \mu_1^X + \begin{pmatrix} \tilde{C}_K \\ \tilde{D}_K \end{pmatrix} L_K U^{K-1}, \Sigma_K^{Y,S}]$ of $((Y^K)^\top, (S^K)^\top)^\top$ (given in Lemma 3), and corresponding probability mass functions $p(\tilde{y}^K | \tilde{s}^K)$ and $p(\tilde{s}^K)$, the probability mass function $p(v^K)$ that minimizes $I[\tilde{S}^K; Z^K]$ subject to the distortion constraint, $E[||Z^K - \tilde{Y}^K||^2] \leq \epsilon_K$, can be found by solving the convex program in* (18)*.*

***Proof:*** The expression for the cost and its convexity follow from Lemma 1. The distortion constraint follows from Lemma 2, and the fact that, because $\tilde{S}^K$ and $Z^K$ are conditionally independent given $\tilde{Y}^K$, the joint distribution $p(z^K, \tilde{y}^K)$ can be written as $\sum_{\tilde{s}^K \in \mathcal{S}^K} p(\tilde{s}^K) p(\tilde{y}^K | \tilde{s}^K) p(z^K | \tilde{y}^K)$. ■

The input data for the optimization problem in (18) are: the upper bound on the allowable distortion $\epsilon_K$, the marginal probability mass function $p(\tilde{s}^K)$, and the transition probabilities $p(\tilde{y}^K | \tilde{s}^K)$. To obtain these probabilities, we need to integrate the density of $S^K$ and the joint density of $(S^K, Y^K)$ (if they are not degenerate) over the quantization cells. In Lemma 3 and Corollary 1, we provide closed form expressions for these densities and prove that they are not degenerate. Once we have these densities, they can be integrated in the sense of Remark 6 to obtain the required $p(\tilde{s}^K)$ and $p(\tilde{y}^K | \tilde{s}^K)$ and then solve problem (18).

Once we have the optimal joint distribution $q^*(v^K)$ solution to (18), we sample from this distribution to obtain an optimal sequence of realizations $V^K = (v^*(1), \dots, v^*(K))^\top$ that we induce to system (1) at every time-step, see Figure 1 and Figure 3. We remark that these realizations are computed a priori, i.e., before we start running the system. Real-time realizations of $Y(K)$ are not needed to compute the optimal distorting distribution. We only require the probability distributions of the processes driving the system dynamics to cast (18) and compute $q^*(v^K)$.

**Remark 7** *Note that, even though working with $N_V^K$ variables is much more manageable than working with the original $(N_Y)^{2K}$ variables (see Remark 5), having $N_V^K$ variables could still lead to large scale optimization problems as sufficiently large $N_V$ is required to have meaningful results. Note, however, that* (18) *is solved off-line, i.e., because real-time sensor realizations are not required to perform the optimization– only the distributions of the processes driving the system are needed–the optimal mappings can be computed before we start running the system. Moreover, problem* (18) *is convex and real-valued. It has a smooth cost function, bounded variables, and linear constraints. There exists many algorithms in the literature to solve this class of large-scale problems efficiently and reasonably fast, see, for instance,* [54], [55], [56].

### *A. Receding Horizon*

For the configuration given in Theorem 1, we have addressed the problem of designing optimal probabilistic mappings to maximize privacy for a finite time horizon $K$. Infinite horizon sub-optimal mappings can be designed by repeatedly solving finite time problems of the form (18) in a receding horizon fashion. With receding horizon we mean the following: for each time step, starting at the current time $k$, we solve a finite horizon problem of the form (18) over a fixed horizon $K$, i.e., over the time window $\{k, k+1, \dots, k+K-1\}$.

$$\begin{cases} \min\limits_{p(z^K|\tilde{y}^K)} \sum\limits_{\tilde{s}^K\in\mathcal{S}^K}\sum\limits_{z^K\in\mathcal{Y}^K} p(\tilde{s}^K)\sum\limits_{\tilde{y}^K\in\mathcal{Y}^K} p(\tilde{y}^K|\tilde{s}^K)p(z^K|\tilde{y}^K)\log\dfrac{\sum_{\tilde{y}^K\in\mathcal{Y}^K} p(\tilde{y}^K|\tilde{s}^K)p(z^K|\tilde{y}^K)}{\sum_{\tilde{s}^K\in\mathcal{S}^K}\sum_{\tilde{y}^K\in\mathcal{Y}^K} p(\tilde{s}^K)p(\tilde{y}^K|\tilde{s}^K)p(z^K|\tilde{y}^K)}, \\ \quad s.t. \sum\limits_{\tilde{y}^K\in\mathcal{Y}^K}\sum\limits_{z^K\in\mathcal{Y}^K}\sum\limits_{\tilde{s}^K\in\mathcal{S}^K} p(\tilde{s}^K)p(\tilde{y}^K|\tilde{s}^K)p(z^K|\tilde{y}^K)\left(z^K-\tilde{y}^K\right)^2\le\epsilon_K, \\ \qquad p(z^K|\tilde{y}^K)=q\big((\bar{\alpha}(z^K)-\bar{\alpha}(\tilde{y}^K))\bmod N_Y\big), \\ \qquad \text{and, } q(v^K)\in \textit{ Simplex.} \end{cases} \tag{18}$$

We sample from the obtained optimal probability distribution $q^*(v_k^{k+K-1})$ to obtain an optimal sequence of realizations $V_k^{k+K-1} = (v_k^*(1),\dots,v_k^*(K))^\top$. We only apply the first realization, $V(k) = v_k^*(1)$, to the system and discard the rest. Then, for the next time step, $k+1$, we solve a new finite horizon problem over $\{k+1,k+2,\dots,k+K\}$, obtain a new optimal distribution, sample from it, and again only apply the first optimal realization, $V(k+1) = v_{k+1}^*(1)$, and discard the rest. We continue shifting the prediction horizon forward to compute the sequence of receding horizon optimal probability distributions and realizations.

Note that, besides the desired distortion constant $\epsilon_K$, problem (18) only requires the probability mass functions $p(\tilde{y}^K|\tilde{s}^K)$ and $p(\tilde{s}^K)$ of $\tilde{Y}^K|\tilde{S}^K$ and $\tilde{S}^K$, respectively, as input information. Similarly, in a receding horizon formulation, for given horizon length $K$, we only need $\epsilon_K$, and the probability distributions $p(\tilde{y}_k^{k+K-1}|s_k^{k+K-1})$ and $p(s_k^{k+K-1})$, of $\tilde{Y}_k^{k+K-1}|S_k^{k+K-1}$ and $S_k^{k+K-1}$, respectively, to cast the corresponding optimization problem, where $\tilde{Y}_k^{k+K-1} = (\tilde{Y}(k)^\top,\dots,\tilde{Y}(k+K-1)^\top)^\top \in \mathcal{Y}^K$ and $\tilde{S}_k^{k+K-1} = (\tilde{S}(k)^\top,\dots,\tilde{S}(k+K-1)^\top)^\top \in \mathcal{S}^K$. To obtain these distributions, we need the density of $S_k^{k+K-1} = (S(k)^\top,\dots,S(k+K-1)^\top)^\top$, and the joint density of $Y_k^{k+K-1} = (Y(k)^\top,\dots,Y(k+K-1)^\top)^\top$ and $S_k^{k+K-1}$, and then integrate them over the quantization cells. By lifting the system dynamics (1) over $\{k,k+1,\dots,k+K-1\}$, we can write the stacked vector $((Y_k^{k+K-1})^\top,(S_k^{k+K-1})^\top)^\top \in \mathbb{R}^{K(n_s+n_y)}$, in terms of the current state $X(k)$, as follows

$$\begin{aligned}\begin{bmatrix} Y_k^{k+K-1} \\ S_k^{k+K-1}\end{bmatrix} &= \begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix} F_K X(k) + \begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix} T_K M_k^{k+K-2} \\ &\quad + \begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix} L_K U_k^{k+K-2} + \begin{bmatrix} I \\ \mathbf{0}\end{bmatrix} W_k^{k+K-1},\end{aligned} \tag{19}$$

with $L_K = T_K(I_{K-1}\otimes B)$, $\tilde{C}_K = I_K\otimes C$, $\tilde{D}_K = I_K\otimes D$, and $F_K$ and $T_K$ as defined in (16).

**Lemma 4**

$$\begin{bmatrix} Y_k^{k+K-1} \\ S_k^{k+K-1}\end{bmatrix} \sim \mathcal{N}[\begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix} F_K\mu_k^X + \begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix} L_K U_k^{k+K-2}, \Sigma_{k,K}^{Y,S}],$$

*where $\mu_k^X = E[X(k)] = A\mu_{k-1}^X + BU(k)$, positive definite covariance matrix $\Sigma_{k,K}^{Y,S}$, and $\Sigma_{k,K}^{Y,S}$ can be written in terms of $\Sigma_k^X = E[(X(k)-\mu_k^X)(X(k)-\mu_k^X)^\top]$ as follows:*

$$\begin{cases} \Sigma_k^X = A\Sigma_{k-1}^X A^\top + \Sigma^M, \\ \Sigma_{k,K}^{Y,S} = \begin{bmatrix} I \\ \mathbf{0}\end{bmatrix}(I_K\otimes\Sigma^W)\begin{bmatrix} I \\ \mathbf{0}\end{bmatrix}^\top + \begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix} F_K\Sigma_k^X F_K^\top \begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix}^\top \\ \qquad\qquad + \begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix} T_K(I_{K-1}\otimes\Sigma^M)T_K^\top\begin{bmatrix}\tilde{C}_K \\ \tilde{D}_K\end{bmatrix}^\top. \end{cases} \tag{20}$$

The proof of Lemma 4 follows similar lines as the proof of Lemma 3 and it is omitted here.

To obtain the density of $S_k^{k+K-1}$, we marginalize the joint density of $Y_k^{k+K-1}$ and $S_k^{k+K-1}$ in Lemma 4 over $Y_k^{k+K-1}$ [53].

**Corollary 2**

$$S_k^{k+K-1} \sim \mathcal{N}[\tilde{D}_K F_K\mu_k^X + \tilde{D}_K L_K U_k^{k+K-2}, \Sigma_{k,K}^S],$$

*with* $\Sigma_{k,K}^S := \begin{bmatrix}\mathbf{0} \\ I_{Kn_s}\end{bmatrix}^\top \Sigma_{k,K}^{Y,S}\begin{bmatrix}\mathbf{0} \\ I_{Kn_s}\end{bmatrix}$.

Having the joint density of $Y_k^{k+K-1}$ and $S_k^{k+K-1}$ allows us to compute, for any set of quantization cells, in the sense of Remark 6 for $S^K$ and $Y^K$, the mass functions $p(\tilde{s}_k^{k+K-1})$ and $p(\tilde{y}_k^{k+K-1}|\tilde{s}_k^{k+K-1})$ by numerically integrating this density over the quantization cells. Hereafter, we assume that $p(\tilde{s}_k^{k+K-1})$ and $p(\tilde{y}_k^{k+K-1}|\tilde{s}_k^{k+K-1})$ are known. Following a similar reasoning as in Lemma 1, the cost $I[\tilde{S}_k^{k+K-1};Z_k^{k+K-1}]$ can be written in terms of the optimization variables, $q(v_k^{k+K-1})$, and it can be proved to be convex. Moreover, following the steps in the proof of Lemma 2, it can be proved that the distortion metric $E[||Z_k^{k+K-1} - \tilde{Y}_k^{k+K-1}||^2]$ is linear in $p(z_k^{k+K-1}|\tilde{y}_k^{k+K-1})$ for given $p(\tilde{s}_k^{k+K-1})$ and $p(\tilde{y}_k^{k+K-1}|\tilde{s}_k^{k+K-1})$. The conditional distribution $p(z_k^{k+K-1}|\tilde{y}_k^{k+K-1})$ is a linear function of $q(v_k^{k+K-1})$, and can be written as $p(z_k^{k+K-1}|\tilde{y}_k^{k+K-1}) = q\big((\bar{\alpha}(z_k^{k+K-1})-\bar{\alpha}(\tilde{y}_k^{k+K-1}))\bmod N_Y\big)$. Therefore, minimizing the mutual information $I[\tilde{S}_k^{k+K-1};Z_k^{k+K-1}]$, using $q(v_k^{k+K-1})$ as optimization variables, subject to $E[||Z_k^{k+K-1} - \tilde{Y}_k^{k+K-1}||^2] \le \epsilon_K$ is a convex program. In what follows, we give the proposed receding horizon scheme.

**Receding Horizon Scheme**

**Input:** $k,K\in\mathbb{N}$, $p(\tilde{y}_k^{k+K-1}|\tilde{s}_k^{k+K-1}) = \Pr[\tilde{Y}_k^{k+K-1} = \tilde{y}_k^{k+K-1}|\tilde{S}_k^{k+K-1} = \tilde{s}_k^{k+K-1}]$, $\tilde{y}_k^{k+K-1}\in\mathcal{Y}^K$, $\tilde{s}_k^{k+K-1}\in\mathcal{S}^K$, $p(s_k^{k+K-1}) = \Pr[\tilde{S}_k^{k+K-1} = \tilde{s}_k^{k+K-1}]$, $\tilde{s}_k^{k+K-1}\in\mathcal{S}^K$, and $\epsilon_K\in\mathbb{R}_{>0}$.

$$\begin{cases} q^*(\tilde{v}_k^{k+K-1}) := \underset{q(v_k^{k+K-1})}{\arg\min}\ I[\tilde{S}_k^{k+K-1}; \tilde{Y}_k^{k+K-1}], \\ \text{s.t. } E[||Z_k^{k+K-1} - \tilde{Y}_k^{k+K-1}||^2] \le \epsilon_K, \\ p(z_k^{k+K-1}|\tilde{y}_k^{k+K-1}) \\ \qquad = q\big((\bar{\alpha}(z_k^{k+K-1}) - \bar{\alpha}(\tilde{y}_k^{k+K-1})) \bmod N_Y\big), \\ \text{and } q\left(v_k^{k+K-1}\right) \in \text{ Simplex}. \end{cases} \tag{21}$$

**Output:** $q^*(\tilde{v}_k^{k+K-1})$.

Summarizing, for given $k$, horizon $K$, and $p(\tilde{y}_k^{k+K-1}|s_k^{k+K-1})$ and $p(x_k^{k+K-1})$ corresponding to the joint density of $(Y_k^{k+K-1}, S_k^{k+K-1})$ in Lemma 4, solve problem (21) and denote the corresponding solution as $q^*\left(v_k^{k+K-1}\right)$. This optimal distribution is the output of our receding scheme at time $k$. Once the optimal distributions have been computed, we sample from $q^*\left(v_k^{k+K-1}\right)$ to obtain an optimal sequence of realizations $V_k^{k+K-1} = (v_k^*(1), \ldots, v_k^*(K))^\top$, and only apply the first realization, $V(k) = v_k^*(1)$, to the system at time $k$, and discard the rest. Then, at time $k+1$, we sample from $q^*\left(v_{k+1}^{k+K}\right)$ and only apply the first realization, $V(k+1) = v_{k+1}^*(1)$, and discard the rest. We repeat the procedure for increasing $k$.

**Remark 8** *Note that the distorting mechanisms resulting from the receding horizon scheme in* (21) *are not optimal for the complete trajectory (although they are optimal for every subinterval). This is a limitation for most applications that use receding horizon formulations (e.g., Model Predictive Control (MPC) schemes suffer from the same heuristic nature – MPC is in general not optimal). Optimality is only guaranteed for the finite horizon results in Section . We hint at how to use these results in an infinite horizon fashion using the receding horizon scheme proposed above but, again, the resulting distorting mechanisms are not guaranteed to be optimal.*

## V. Simulation Experiments

We illustrate the performance of our tools through a case study of a well stirred chemical reactor with heat exchanger. This case study has been developed over the years as a benchmark example for control systems and fault detection, see, e.g., [57]-[59] and references therein. The state, inputs, and output of the reactor are:

$$\begin{cases} X(t) = \begin{pmatrix} C_0 \\ T_0 \\ T_w \\ T_m \end{pmatrix}, U(t) = \begin{pmatrix} C_u \\ T_u \\ T_{w,u} \end{pmatrix}, Y(t) = T_0, \end{cases}$$

where

$$\begin{cases} C_0 & \text{: Concentration of the chemical product,} \\ T_0 & \text{: Temperature of the product,} \\ T_w & \text{: Temperature of the jacket water of heat exchanger,} \\ T_m & \text{: Coolant temperature,} \\ C_u & \text{: Inlet concentration of reactant,} \\ T_u & \text{: Inlet temperature,} \\ T_{w,u} & \text{: Coolant water inlet temperature.} \end{cases}$$

We use the discretized dynamics of the reactor introduced in [59]. The discrete-time dynamics is of the form (1) with matrices $A, B, C$ as follows

$$\begin{cases} A = \begin{bmatrix} 0.8353 & 0 & 0 & 0 \\ 0 & 0.8324 & 0 & 0.0031 \\ 0 & 0.0001 & 0.1633 & 0 \\ 0 & 0.0001 & 0.1633 & 0 \end{bmatrix} \\ (B|C^\top) = \left[\begin{array}{ccc|c} 0.0458 & 0 & 0 & 1 \\ 0 & 0.0457 & 0 & 1 \\ 0 & 0 & 0.0231 & 0 \\ 0 & 0.0007 & 0.0006 & 0 \end{array}\right] \end{cases} \tag{22}$$

The original model in [57] does not consider sensor/system noise and reference signals, we include some arbitrary noise and references for our simulation experiments. We consider system and sensor noise with covariance matrices $\Sigma^M = \text{diag}[0.1, 0.2, 0.3, 0.4]$ and $\Sigma^W = 0.1$, respectively, normally distributed initial condition $X(1) \sim \mathcal{N}[(6.94; 13.76; 1; 1)^\top, I_4]$, and reference signal $U(k) = (50\cos[0.5k]^2, 50\tanh[3k], -70\sin[0.1k])^\top$. As private output, we use the concentration of the chemical product; then, the matrix $D$ in (1) is given by the full row rank matrix $D = (1, 0, 0, 0)$. The output of the system is the temperature of the product $T_0$, which could be monitored, e.g., for quality/safety reasons. Then, the aim of the privacy scheme is to hide the concentration of the reactant as much as possible without distorting temperature measurements excessively.

We first consider the receding horizon formulation of Problem 1, i.e., minimizing $I[\tilde{S}_k^{k+K-1}; Z_k^{k+K-1}]$ for some quantized version $\tilde{S}(k)$ of the private output $S(k)$, over sliding windows of the form $\{k, k+1, \ldots, k+K-1\}$, $k \in \mathbb{N}$. For these experiments, we use a 3-bit quantizer for $Y(k)$ ($N_Y = 8$) with levels

$$\begin{aligned} \mathcal{Y} &= \{y_1, \ldots, y_8\} \\ &= \{18.38, 19.04, 19.71, 20.37, 21.04, 21.70, 22.36, 23.03\}, \end{aligned}$$

and corresponding quantization cells

$$\begin{aligned} \mathcal{C} = \{&(-\infty, 18.71], (18.71, 19.38], (19.38, 20.04], (20.04, 20.70], \\ &(20.70, 21.37], (21.37, 22.03], (22.03, 22.70], (22.70, \infty)\}. \end{aligned}$$

For the private output $S(k)$, we use a binary quantizer ($N_S = 2$) with levels $\mathcal{S} = \{s_1, s_2\} = \{6.20, 7.68\}$, and cells $\mathcal{H} = \{h_1, h_2\} = \{(-\infty, 6.94], (6.94, \infty)\}$. These quantizers were selected based on system trajectories to avoid having always saturated $\tilde{Y}(k)$ and $\tilde{S}(k)$, see Figure 4. Then, for horizon $K = 3$ and $k \in \mathbb{N}$, the stacked quantized output $\tilde{Y}_k^{k+K-1}$ (and thus also $Z_k^{k+K-1}$) and the stacked private output $\tilde{S}_k^{k+K-1}$ have alphabets $\mathcal{Y}^3$ (with $(N_Y)^K = 512$ elements) and $\mathcal{S}^3$ (with $(N_S)^K = 8$ elements) given by

$$\mathcal{Y}^3 = \left\{ \begin{bmatrix} y_1 \\ y_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ y_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} y_3 \\ y_1 \\ y_1 \end{bmatrix}, \ldots, \begin{bmatrix} y_6 \\ y_8 \\ y_8 \end{bmatrix}, \begin{bmatrix} y_7 \\ y_8 \\ y_8 \end{bmatrix}, \begin{bmatrix} y_8 \\ y_8 \\ y_8 \end{bmatrix} \right\}, \tag{23}$$

$$\mathcal{S}^3 = \left\{ \begin{bmatrix} s_1 \\ s_1 \\ s_1 \end{bmatrix}, \begin{bmatrix} s_2 \\ s_1 \\ s_1 \end{bmatrix}, \begin{bmatrix} s_1 \\ s_2 \\ s_1 \end{bmatrix}, \begin{bmatrix} s_2 \\ s_2 \\ s_1 \end{bmatrix}, \ldots, \begin{bmatrix} s_1 \\ s_2 \\ s_2 \end{bmatrix}, \begin{bmatrix} s_2 \\ s_2 \\ s_2 \end{bmatrix} \right\}, \tag{24}$$

with $y_i$, $i = 1, 2, \ldots, 8$, and $s_j$, $j \in \{1, 2\}$, as introduced above. We let the distorting random process $V(k)$

(see (3)-(5)) have an alphabet with $N_V = 5$ elements, i.e., $\mathcal{V} = \{0,1,2,3,4\}$. Then, for $K = 3$, the stacked vector $V_k^{k+2}$ has alphabet $\mathcal{V}^3$ with $(N_V)^3 = 125$ elements. To cast problem (21), we need the probability mass functions $p(\tilde{y}_k^{k+2}|\tilde{s}_k^{k+2}) = p(\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2})/p(\tilde{s}_k^{k+2})$ and $p(\tilde{s}_k^{k+2}) = \sum_{\tilde{s}_k^{k+2} \in \mathcal{S}^3} p(\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2})$, $\tilde{y}_k^{k+2} \in \mathcal{Y}^3$, $\tilde{s}_k^{k+2} \in \mathcal{S}^3$. These distributions are fully characterized by the joint pmf, $p(\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2})$, $\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2} \in \mathcal{Y}^3 \times \mathcal{S}^3$, where the alphabet $\mathcal{Y}^3 \times \mathcal{S}^3$ is given by

$$\mathcal{Y}^3 \times \mathcal{S}^3 = \left\{ \begin{bmatrix} y_1 \\ y_1 \\ y_1 \\ s_1 \\ s_1 \\ s_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ y_1 \\ y_1 \\ s_1 \\ s_1 \\ s_1 \end{bmatrix}, \ldots, \begin{bmatrix} y_7 \\ y_8 \\ y_8 \\ s_1 \\ s_1 \\ s_1 \end{bmatrix}, \begin{bmatrix} y_8 \\ y_8 \\ y_8 \\ s_1 \\ s_1 \\ s_1 \end{bmatrix}, \right. \tag{25}$$

$$\left. \begin{bmatrix} y_1 \\ y_1 \\ y_1 \\ s_2 \\ s_1 \\ s_1 \end{bmatrix}, \ldots, \begin{bmatrix} y_8 \\ y_8 \\ y_8 \\ s_2 \\ s_1 \\ s_1 \end{bmatrix}, \ldots, \begin{bmatrix} y_8 \\ y_8 \\ y_8 \\ s_3 \\ s_3 \\ s_3 \end{bmatrix} \right\}, \tag{26}$$

with $|\mathcal{Y}^3 \times \mathcal{S}^3| = (N_Y N_S)^3 = 4096$. We integrate (in the sense of Remark 6) the joint density of $Y_k^{k+2}$ and $S_k^{k+2}$ (given in Lemma 4) over the quantization cells, $\mathcal{C}$ and $\mathcal{H}$, to obtain $p(\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2})$ (and thus also $p(\tilde{s}_k^{k+2})$ and $p(\tilde{y}_k^{k+2}|\tilde{s}_k^{k+2})$) for $k \in \{1, \ldots, 25\}$. In Figure 5(a), we show the joint probability mass function $p(\tilde{y}_1^3, \tilde{s}_1^3)$, $\tilde{y}_1^3, \tilde{s}_1^3 \in \mathcal{Y}^3 \times \mathcal{S}^3$. The mass points are indexed following the ordering logic in (26). Figure 5(b) depicts the joint pmf $p(\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2})$, $k = 1, \ldots, 25$, for the first 256 mass points of the alphabet $\mathcal{Y}^3 \times \mathcal{S}^3$ (indexed as in (26)); and, in Figure 5(c), we show zooms of the probabilities of the first six mass points for increasing $k$. Note that there is a lot of variability in the probabilities of the mass points as $k$ increases.

Next, for given $p(\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2})$, we show results of the receding horizon scheme (21) for horizon $K = 3$, $k = 1, 2, \ldots, 25$, and three different levels of distortion, $\epsilon_K = \infty, 7, 3$, where $\epsilon_K = \infty$ means that the optimization problems in (21) are solved without considering the distortion constraint. In Figure 6, we show the optimal distorting distribution $q^*(v_1^2)$ for $\epsilon_K = \infty$ and, in Figure 7, the evolution (of the first sixteen mass points) of the optimal distribution $q^*(v_k^{k+2})$ solution to (21) for increasing $k$ and $\epsilon_K = \infty, 7, 2$. The mass points are indexed following the ordering logic in (26). Note that there is a lot of probability variability both among mass points and in time (as $k$ grows). The optimal distorting distributions follow some nontrivial patterns that depend on the dynamics, quantizer, and desired distortion level. Further note that even the unconstraint formulation ($\epsilon_K = \infty$) does not lead to uniform distributions; and, as $\epsilon_K \to 0$, the sequence of optimal distributions $q^*(v_k^{k+2})$ concentrates most of its probability at the first mass point (the zero vector). This is what one would expect as the zero vector leads to no distortion. For $\epsilon_K = \infty, 7, 2$, Figure 8 depicts the evolution of the optimal cost $I[\tilde{S}_k^{k+K-1}; Z_k^{k+K-1}]$ and the mutual information $I[\tilde{S}_k^{k+K-1}; \tilde{Y}_k^{k+K-1}]$. As expected, $I[\tilde{S}_k^{k+K-1}; Z_k^{k+K-1}]$ decreases for increasing $\epsilon_K$ uniformly in $k$. The mutual information $I[\tilde{S}_k^{k+K-1}; \tilde{Y}_k^{k+K-1}]$ in Figure 8(b) characterizes the disclosed information if no privacy preserving mapping was in place. Finally, in Figure 9, we show the joint probability distributions $p(z_1^3, \tilde{s}_1^3)$ and $p(\tilde{y}_1^3, \tilde{s}_1^3)$ for $\epsilon_K = \infty$ and $\epsilon_K = 2$. Note that, as one would expect, $p(z_1^3, \tilde{s}_1^3) \to p(\tilde{y}_1^3, \tilde{s}_1^3)$ as $\epsilon_K \to 0$. That is, as we allow for less distortion, we have less freedom to reshape $p(\tilde{y}_1^3, \tilde{s}_1^3)$ by passing $\tilde{Y}_1^3$ through $p(\tilde{y}_1^3, \tilde{z}_1^3)$ before transmission.

We remark that all the above computations were performed on a PC, Intel 2.70 GHz, in Matlab 2015b (using the parallel computing toolbox with four cores). In Figure 6, we show the optimal distribution $q^*(v_1^2)$, i.e., we solve the optimization problem in (21) once (for $K = 3$ and $k = 1$) using 125 variables ($N_V = 5$) and $\epsilon_3 = \infty$ (unconstraint case). This optimization took 8.7 seconds to be performed. The same optimization but with $\epsilon_3 = 8$ (considering the distortion constraint) took 14.6 seconds. In Figure 10, for $\epsilon_3 = 8$, we show the evolution of the computation time as the cardinality $N_V = |\mathcal{V}|$ of the alphabet of the additive process $V(k)$ increases (i.e., as we increase the number of variables $N_V^K$). The computation time grows exponentially with the number of variables. Note, however, that the number of variables is independent of the size of the quantizers. We fix $|V|$ a priori and then solve (21) for given quantizers. What changes with the size of the quantizers is the structure of the cost function as finer quantizers lead to increasingly involved expressions in (21). Finally, in Figure 7(a), we show the evolution of the optimal $q^*(v_k^{k+K-1})$ for $K = 3$ and $k = 1, 2, \ldots, 16$. That is, we solve the optimization problem in (21) sixteen times following the receding horizon formulation. It took 160.5 seconds to perform these sixteen optimizations.

## VI. Conclusion

For a class of Cyber-Physical-Systems (CPSs), we have presented a detailed mathematical framework built around systems theory, 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 private part of the system 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 private output as much as possible) while satisfying the distortion constraint on the original sensor data. Our tools are capable of dealing with the dynamic and non-stationary nature of CPSs at the price of having to solve some medium to large scale optimization problems. We have presented extensive simulation experiments to show the performance of out tools. Note that we have found some nontrivial distorting probability distributions that highly depend on the system dynamics, quantizer, and desired distortion level.

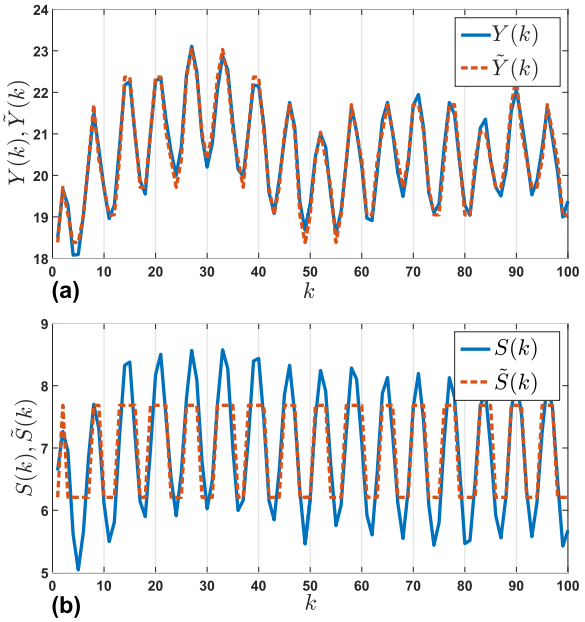


Fig. 4: (a) Sample path of the output $Y(k)$ and its quantized version $\tilde{Y}(k)$ (3-bit quantizer); and (b) Sample path of the private output $S(k)$ and its quantized version $\tilde{S}(k)$ (1-bit quantizer).


## ACKNOWLEDGMENT

This work was supported by the Australian Research Council (ARC) under the Project DP170104099; the NATO Science for Peace and Security (SPS) PROGRAMME under the project SPS.SFP G5479; and a Schmidt Data-X Grant from the Princeton Center for Statistics and Machine Learning.


## APPENDICES

### A. *Proof of Lemma 1*

The expression on the right-hand side of (13a) follows by inspection of Definition 1, and the fact that $p(\tilde{s}^K, z^K) = p(\tilde{s}^K)p(z^K|\tilde{s}^K)$ (chain rule). By [28, Theorem 2.7.4], cost (13a) is convex in $p(z^K|\tilde{s}^K)$ for given $p(\tilde{s}^K)$. However, our optimization variables are $q(v^K)$ and not $p(z^K|\tilde{s}^K)$. By construction, $\tilde{S}^K$ and $Z^K$ are conditionally independent given $\tilde{Y}^K$. That is, given the information flow in the system (see Figure 1), all the information that $Z^K$ might carry about $\tilde{S}^K$ is carried by $\tilde{Y}^K$ because $Z^K = G(\tilde{Y}^K)$ is a random function of $\tilde{Y}^K$ only. The later implies that given $\tilde{Y}^K$, $\tilde{S}^K$ and $Z^K$ are conditionally independent (see, e.g., [28, Sections 2.8 and 4.1] for details). It follows that $p(\tilde{s}^K, \tilde{y}^K, z^K) = p(\tilde{s}^K)p(\tilde{y}^K|\tilde{s}^K)p(z^K|\tilde{y}^K)$. Using this expression for $p(\tilde{s}^K, \tilde{y}^K, z^K)$, we can write (13b) and (13c) by conditioning and marginalizing the joint distribution. It follows that, by combining (13a)-(13c), we can write $I[\tilde{S}^K; Z^K]$ in terms of $p(z^K|\tilde{y}^K)$, $p(\tilde{y}^K|\tilde{s}^K)$, and $p(\tilde{s}^K)$, i.e., the cost is a function of $p(z^K|\tilde{y}^K)$ and it is parametrized by $p(\tilde{y}^K|\tilde{s}^K)$ and $p(\tilde{s}^K)$. Convexity with respect to $p(z^K|\tilde{y}^K)$ follows from convexity with respect to $p(z^K|\tilde{s}^K)$ because $p(z^K|\tilde{s}^K)$ is just a linear combination of $p(z^K|\tilde{y}^K)$, see (13b), and convexity is preserved under affine transformations [54]. By definition,

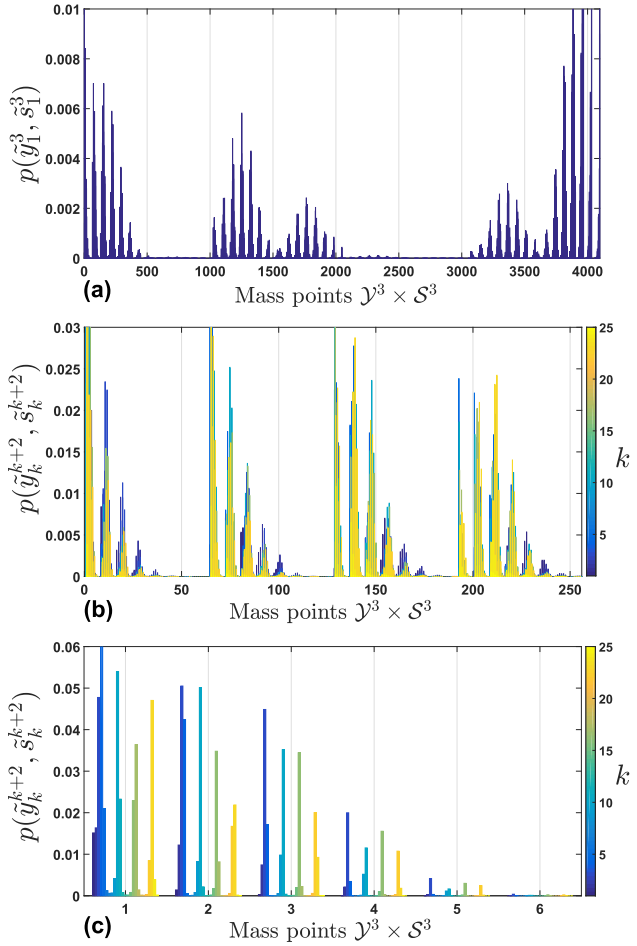


Fig. 5: (a) Joint pmf $p(\tilde{y}_1^3, \tilde{s}_1^3)$, $\tilde{y}_1^3, \tilde{s}_1^3 \in \mathcal{Y}^3 \times \mathcal{S}^3$; (b) Joint pmf $p(\tilde{y}_k^{k+2}, \tilde{s}_k^{k+2})$, $k = 1, \ldots, 25$, for the first 256 mass points; and (c) Zoom of the probabilities of the first six mass points for increasing $k$ .

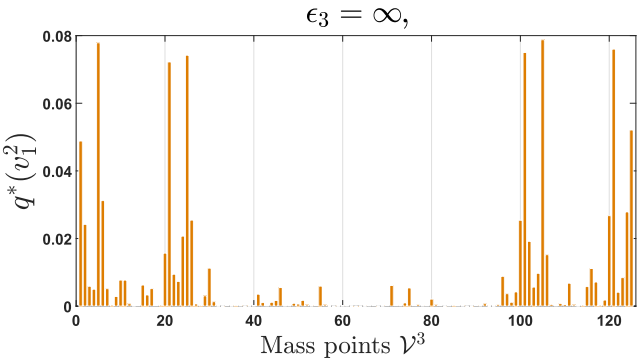


Fig. 6: Optimal distorting distribution $q^*(v_1^2)$ for $\epsilon_K = \infty$.

$p(z^K, \tilde{y}^K) = \Pr[Z^K = z^K, \tilde{Y}^K = \tilde{y}^K]$, $z^K, \tilde{y}^K \in \mathcal{Y}^K$. Note that

$$\begin{aligned}&\Pr[Z^K = z^K, \tilde{Y}^K = \tilde{y}^K]\\ &= \Pr[Z(1) = z(1), \ldots, Z(K) = z(K),\\ &\qquad \tilde{Y}(1) = \tilde{y}(1), \ldots, \tilde{Y}(K) = \tilde{y}(K)],\end{aligned}$$

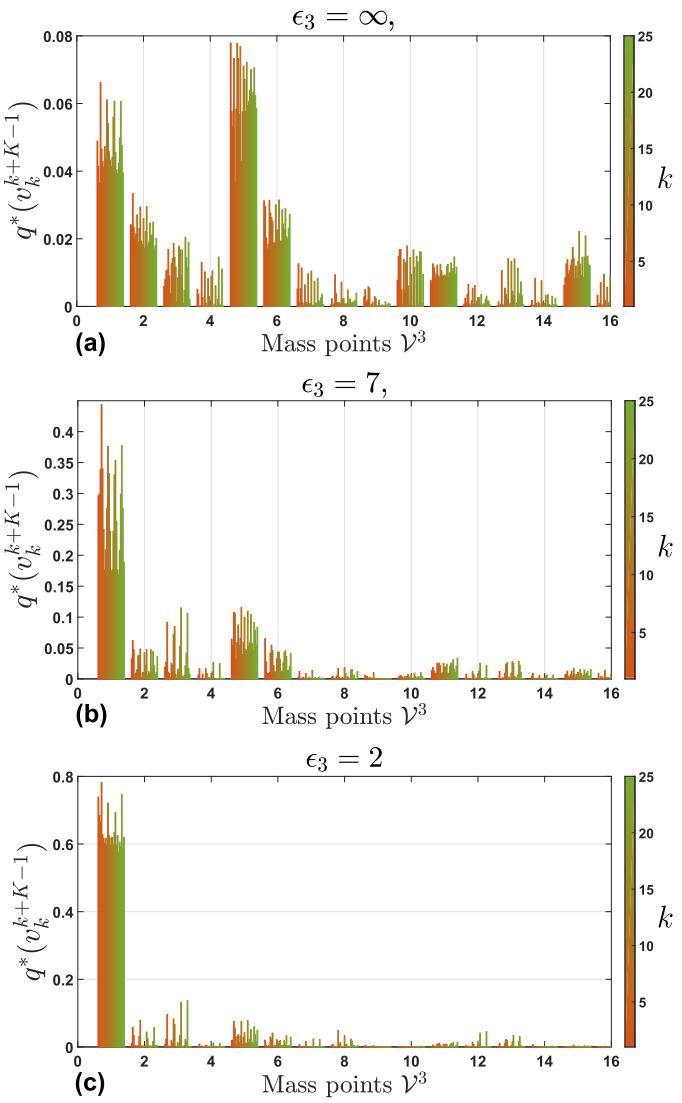


Fig. 7: Evolution of the optimal probability distribution $q^*(v_k^{k+2})$ solution to (21) for increasing $k$ and different distortion upper bounds $\epsilon_K$.

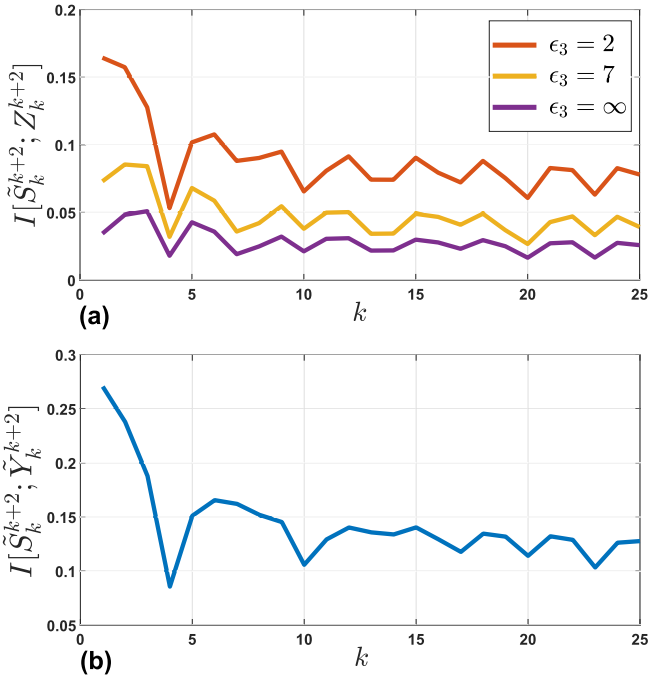


Fig. 8: For $K = 3$ and increasing $k$: **(a)** Evolution of the optimal cost $I[\tilde{S}_k^{k+K-1}; Z_k^{k+K-1}]$; and **(b)** Mutual information between $\tilde{S}_k^{k+K-1}$ and $\tilde{Y}_k^{k+K-1}$, $I[\tilde{S}_k^{k+K-1}; \tilde{Y}_k^{k+K-1}]$ (information disclosed if no privacy preserving mapping was in place).

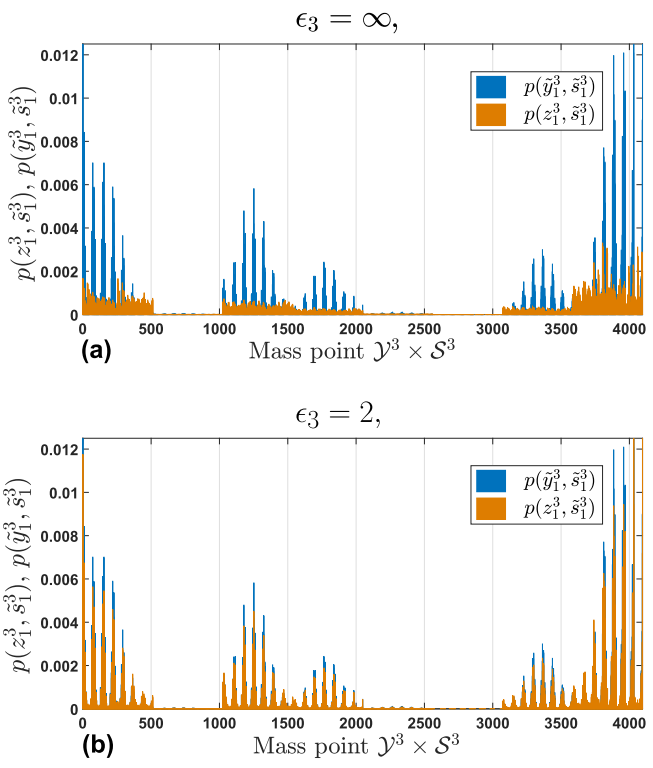


Fig. 9: For $K = 3$: **(a)** Joint probability distributions $p(z_1^3, \tilde{s}_1^3)$ and $p(\tilde{y}_1^3, \tilde{s}_1^3)$ for $\epsilon_K = \infty$; and **(b)** $p(z_1^3, \tilde{s}_1^3)$ and $p(\tilde{y}_1^3, \tilde{s}_1^3)$ for $\epsilon_K = 2$. Note that as $\epsilon_K \to 0$, $p(z_1^3, \tilde{s}_1^3) \to p(\tilde{y}_1^3, \tilde{s}_1^3)$.

for $z(i), \tilde{y}(j) \in \mathcal{Y}$, $i, j \in \{1, \dots, K\}$. Using (5), we can further expand $\Pr[Z^K = z^K, \tilde{Y}^K = \tilde{y}^K]$ as follows

$$
\begin{aligned}
&\Pr[Z^K = z^K, \tilde{Y}^K = \tilde{y}^K] \\
&= \Pr[\beta\big((\alpha(\tilde{Y}(1)) + V(1)) \bmod N_Y\big) = z(1), \dots, \\
&\qquad\qquad \tilde{Y}(1) = \tilde{y}(1), \dots, \tilde{Y}(K) = \tilde{y}(K)] \\
&= \Pr[V(1) = \big(\alpha(z(1)) - \alpha(\tilde{y}(1))\big) \bmod N_Y, \dots, \\
&\qquad\qquad \tilde{Y}(1) = \tilde{y}(1), \dots, \tilde{Y}(K) = \tilde{y}(K)] \\
&\overset{(a)}{=} \Pr[V(1) = \big(\alpha(z(1)) - \alpha(\tilde{y}(1))\big) \bmod N_Y, \dots, \\
&\quad V(K) = \big(\alpha(z(K)) - \alpha(\tilde{y}(K))\big) \bmod N_Y] \Pr[\tilde{Y}^K = \tilde{y}^K] \\
&\overset{(b)}{=} \Pr[V^K = \big(\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)\big) \bmod N_Y] \Pr[\tilde{Y}^K = \tilde{y}^K] \\
&\overset{(c)}{=} q\big(\big(\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)\big) \bmod N_Y\big) p(\tilde{y}^K),
\end{aligned}
$$

where (a) follows from independence between $V^K$ and $\tilde{Y}^K$, (b) from the definition of $\bar{\alpha}(\cdot)$ in (12), and (c) by con-

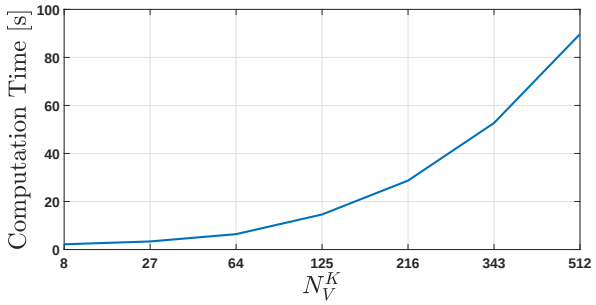


Fig. 10: Computation time for problem (21), with $K = 3$ and $k = 1$, as the number of variables, $N_V^K$, increases.

struction of $q(v^K)$ since $q(v^K) = \Pr[V^K = v^k]$, $v^k \in \mathcal{V}^K$. It follows that $p(z^K|\tilde{y}^K) = p(z^K, \tilde{y}^K)/p(\tilde{y}^K) = q\big(\big(\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)\big) \bmod N_Y\big)$ and thus (13d) holds true. It remains to prove that $I[\tilde{S}^K; Z^K]$ is convex in $q(v^K)$ for given $p(\tilde{y}^K)$ and $p(\tilde{s}^K|\tilde{y}^K)$. We have concluded convexity of $I[\tilde{S}^K; Z^K]$ with respect to $p(z^K|\tilde{y}^K)$ above. Hence, because $p(z^K|\tilde{y}^K) = q\big(\big(\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)\big) \bmod N_Y\big)$ and $q\big(\big(\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)\big) \bmod N_Y\big)$ is a linear transformation of $q(v^K)$ (note that $q\big(\big(\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)\big) \bmod N_Y\big) = q(v^K)$ for $\big(\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)\big) \bmod N_Y = v^K$ and zero otherwise), the cost $I[\tilde{S}^K; Z^K]$ is convex in $q(v^K)$ because convexity is preserved under affine transformations. ■

### B. Proof of Lemma 2

Define the function $d(Z^K, \tilde{Y}^K) := ||Z^K - \tilde{Y}^K||^2$. The function $d(Z^K, \tilde{Y}^K)$ is a deterministic function of two jointly distributed random vectors, $Z^K$ and $\tilde{Y}^K$, with joint distribution $p(z^K, \tilde{y}^K)$. The joint distribution can be written as $p(z^K, \tilde{y}^K) = p(z^K|\tilde{y}^K)p(\tilde{y}^K)$ (chain rule), and, by (13d), the conditional probability distribution is given by $p(z^K|\tilde{y}^K) = q\big((\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)) \bmod N_Y$ (see the proof of Lemma 1 for details). That is, $p(z^K|\tilde{y}^K)$ is a linear transformation of $q(v^K)$. Therefore, see, e.g., [53] for details, $E[d(Z^K, \tilde{Y}^K)] = \sum_{\tilde{y}^K \in \mathcal{Y}^K} \sum_{z^K \in \mathcal{Y}^K} p(z^K, \tilde{y}^K) d(z^K, \tilde{y}^K)$, and because the joint distribution can be written as $p(z^K, \tilde{y}^K) = q\big((\bar{\alpha}(z^K) - \bar{\alpha}(\tilde{y}^K)) \bmod N_Y\big)p(\tilde{y}^K)$, the expected distortion $E[d(Z^K, \tilde{Y}^K)]$ is given by (14), and it is linear in $q(v^K)$ for given $p(\tilde{y}^K)$. ■

### C. Proof of Lemma 3

To simplify notation, we introduce the stacked vector $\Theta^K := ((Y^K)^\top, (S^K)^\top)^\top$. By assumption, the initial condition $X(1)$, and the processes, $M(k)$ and $W(k)$, $k \in \mathbb{N}$, are mutually independent, and $X(1) \sim \mathcal{N}[\mu_1^X, \Sigma_1^X]$, $M(k) \sim \mathcal{N}[\mathbf{0}, \Sigma^M]$, and $W(k) \sim \mathcal{N}[\mathbf{0}, \Sigma^W]$ for some positive definite covariance matrices $\Sigma_1^X$, $\Sigma^M$, and $\Sigma^W$. Then, see [53] for details, we have $L_1 X(1) \sim \mathcal{N}[L_1\mu_1^X, L_1\Sigma_1^X L_1^\top]$, $L_2 M^{K-1} \sim \mathcal{N}[\mathbf{0}, L_2(I_{K-1} \otimes \Sigma^M)L_2^\top]$, and $L_3 W^K \sim \mathcal{N}[\mathbf{0}, L_3(I_K \otimes \Sigma^W)L_3^\top]$, for any deterministic matrices $L_j$, $j = 1, 2, 3$, of appropriate dimensions. It follows that $\Theta^K = ((Y^K)^\top, (S^K)^\top)^\top$ given in (15) is the sum of a deterministic vector, $[\,\tilde{C}_K^\top \;\; \tilde{D}_K^\top\,]^\top L_K U^{K-1}$, and three independent normally distributed vectors. Therefore, $\Theta^K$ follows a multivariate normal distribution with $E[\Theta^K] = [\,\tilde{C}_K^\top \;\; \tilde{D}_K^\top\,]^\top F_K \mu_1^X + [\,\tilde{C}_K^\top \;\; \tilde{D}_K^\top\,]^\top L_K U^{K-1}$. By inspection, using the expression of $\Theta^K$ in (15), mutual independence among $X(1)$, $M(k)$, and $W(k)$, $k \in \mathbb{N}$, and the definition of $\Sigma_1^X$, $\Sigma_1^X := E[(X(1) - \mu_1^X)(X(1) - \mu_1^X)^\top]$, it can be verified that the covariance matrix of $\Theta^K$, $E[(\Theta^K - E[\Theta^K])(\Theta^K - E[\Theta^K])^\top]$, is given by $\Sigma_K^{Y,S}$ in (17). It remains to prove that the distribution of $\Theta^K$ is not degenerate, i.e., $\Sigma_K^{Y,S} > 0$. Note that $\Sigma_K^{Y,S}$ in (17) can be written as

$$\Sigma_K^{Y,S} = \begin{bmatrix} \tilde{C}_K Q \tilde{C}_K^\top + (I_K \otimes \Sigma^W) & \tilde{C}_K Q \tilde{D}_K^\top \\ \tilde{D}_K Q \tilde{C}_K^\top & \tilde{D}_K Q \tilde{D}_K^\top \end{bmatrix}, \qquad (27)$$

with $Q := F_K \Sigma_1^X F_K^\top + T_K(I_{K-1} \otimes \Sigma^M)T_K^\top$. A necessary condition for the block matrix $\Sigma_K^{Y,S}$ in (27) to be positive definite is that the diagonal blocks are positive definite [52]. The left-upper block is positive definite because $\Sigma^W > 0$ (which implies $(I_K \otimes \Sigma^W) > 0$ [52]). The right-lower block is positive definite if $\tilde{D}_K$ is full row rank and $Q$ is positive definite. Because $D$ is full row rank by assumption, matrix $\tilde{D}_K := (I_K \otimes D)$ is also full row rank [60, Theorem 4.2.15]. Note that $Q$ can be factored as follows

$$Q = \underbrace{\begin{bmatrix} F_K & T_K \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} \Sigma_1^X & \mathbf{0} \\ \mathbf{0} & I_{K-1} \otimes \Sigma^M \end{bmatrix}}_{Q'} \begin{bmatrix} F_K & T_K \end{bmatrix}^\top.$$

That is, $Q$ is a linear transformation of the block diagonal matrix $Q'$ above. By inspection, it can be verified that matrix $P = [F_K \; T_K]$, see (16), is lower triangular with identity matrices on the diagonal; thus, $P$ is invertible. It follows that $Q = PQ'P^\top$ is a congruence transformation of $Q'$ [61]. The later implies that $Q$ and $Q'$ have the same signature [61] (equal number of positive and nonpositive eigenvalues); hence, $Q$ is positive definite if and only if the block diagonal matrices of $Q'$ are positive definite. Matrices $\Sigma_1^X$ and $\Sigma^M$ are positive definite by assumption (which implies $(I_{K-1} \otimes \Sigma^M) > 0$), and thus we can conclude that $Q > 0$, which implies $\tilde{D}_K Q \tilde{D}_K^\top > 0$ because $\tilde{D}_K$ is full row rank. Necessary and sufficient conditions for $\Sigma_K^{Y,S} > 0$ are $\tilde{D}_K Q \tilde{D}_K^\top > 0$ (which we have already proved) and that the Schur complement of block $\tilde{D}_K Q \tilde{D}_K^\top$ of $\Sigma_K^{Y,S}$, denoted as $\Sigma_K^{Y,S}/\tilde{D}_K Q \tilde{D}_K^\top$, is positive definite [62, Theorem 1.12]. This Schur complement is given by

$$\begin{aligned} &\Sigma_K^{Y,S}/\tilde{D}_K Q \tilde{D}_K^\top \\ &= (I_K \otimes \Sigma^W) + \tilde{C}_K\big(Q - Q\tilde{D}_K^\top(\tilde{D}_K Q \tilde{N}_K^\top)^{-1}\tilde{N}_K Q\big)\tilde{C}_K^\top. \end{aligned}$$

Since matrix $(I_K \otimes \Sigma^W)$ is positive definite by construction, a sufficient condition for $\Sigma_K^{Y,S}/\tilde{D}_K Q \tilde{D}_K^\top > 0$ is

$$Q'' := Q - Q\tilde{D}_K^\top(\tilde{D}_K Q \tilde{N}_K^\top)^{-1}\tilde{N}_K Q > 0.$$

Regarding $Q''$ as the Schur complement of a higher dimensional matrix $Q'''$, we can conclude that:

$$Q'' \geq 0 \iff Q''' := \begin{bmatrix} Q \\ \tilde{D}_K Q \end{bmatrix} Q^{-1} \begin{bmatrix} Q & Q\tilde{D}_K^\top \end{bmatrix} \geq 0,$$

which is trivially true because $Q^{-1}$ is positive definite since $Q > 0$. Hence, $\tilde{D}_K Q \tilde{D}_K^\top$ and $\Sigma_K^{Y,S}/\tilde{D}_K Q \tilde{D}_K^\top$ are both positive definite, and thus $\Sigma_K^{Y,S} > 0$. ■

## REFERENCES


[1] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor, "Smart meter privacy: A utility-privacy framework," in *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 190–195.

[2] O. Tan, D. Gündüz, and H. V. Poor, "Increasing smart meter privacy through energy harvesting and storage devices," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 1331–1341, 2013.

[3] Z. Huang, Y. Wang, S. Mitra, and G. E. Dullerud, "On the cost of differential privacy in distributed control systems," in *Proceedings of the 3rd International Conference on High Confidence Networked Systems*, 2014, pp. 105–114.

[4] and M. Gruteser, , and A. Alrabady, "Enhancing security and privacy in traffic-monitoring systems," *IEEE Pervasive Computing*, vol. 5, pp. 38–46, 2006.

[5] R. H. Weber, "Internet of things - new security and privacy challenges," *Computer Law and Security Review*, vol. 26, pp. 23–30, 2010.

[6] V. Poor, *Privacy in Networks of Interacting Agents. Chapter 19 in Emerging Applications of Control and System Theory: A Festschrift in Honor of Mathukumalli Vidyasagar*. Springer Berlin Heidelberg, 2018.

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

[8] F. Farokhi, H. Sandberg, I. Shames, and M. Cantoni, "Quadratic Gaussian privacy games," in *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 4505–4510.

[9] J. Le Ny and M. Mohammady, "Differentially private mimo filtering for event streams," *IEEE Transactions on Automatic Control*, vol. 63, pp. 145–157, 2018.

[10] Y. Kawano and M. Cao, "Design of privacy-preserving dynamic controllers," technical report, arXiv:1901.07384 [eess.SY]. https://arxiv.org/abs/1901.07384v3.

[11] N. Monshizadeh and P. Tabuada, "Plausible deniability as a notion of privacy," in *Proceedings of the IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 1710–1715.

[12] F. Farokhi and G. Nair, "Privacy-constrained communication," *IFAC-PapersOnLine*, vol. 49, pp. 43 – 48, 2016.

[13] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding sensor outputs for injection attacks detection," in *Proceedings of the 53rd Annual Conference on Decision and Control (CDC)*, 2014, pp. 5776–5781.

[14] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, pp. 1355–1387, 1975.

[15] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," in *Advances in Cryptology*, T. Beth, N. Cot, and I. Ingemarsson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1985, pp. 33–50.

[16] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 1401–1408.

[17] J. L. Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, pp. 341–354, 2014.

[18] S. Han, U. Topcu, and G. J. Pappas, "Differentially private convex optimization with piecewise affine objectives," in *Proceedings of the 53rd IEEE Conference on Decision and Control*, 2014.

[19] J. Soria-Comas and J. Domingo-Ferrer, "Optimal data-independent noise for differential privacy," *Information Sciences*, vol. 250, pp. 200 – 214, 2013.

[20] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy," in *Proceedings of the IEEE International Symposium on Information Theory*, 2014, pp. 2371–2375.

[21] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud, "Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems," in *Proceedings of the 53rd IEEE Conference on Decision and Control*, 2014, pp. 2130–2135.

[22] E. Nekouei, T. Tanaka, M. Skoglund, and K. H. Johansson, "Information-theoretic approaches to privacy in estimation and control," *Annual Reviews in Control*, vol. 47, pp. 412 – 422, 2019.

[23] T. T. Ali Reza Pedram and M. Hale, "Bidirectional information flow and the roles of privacy masks in cloud-based control," technical report, arXiv:1905.07459 [cs.IT]. https://arxiv.org/abs/1905.07459.

[24] H. S. Takashi Tanaka, Mikael Skoglund and K. H. Johansson, "Directed information as privacy measure in cloud-based control," technical report, arXiv:1705.02802 [math.OC]. https://arxiv.org/abs/1705.02802.

[25] F. F. N. D. Murguia C., Shames I., *Information-Theoretic Privacy Through Chaos Synchronization and Optimal Additive Noise. In: Farokhi F. (eds) Privacy in Dynamical Systems*. Singapore: Springer, 2020.

[26] F. Farokhi, *Privacy in Dynamical Systems*. Springer Singapore, 2019.

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

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

[29] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–19.

[30] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, pp. 211–407, 2014.

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

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

[33] E. Akyol, C. Langbort, and T. Basar, "Privacy constrained information processing," in *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 4511–4516.

[34] B. Bollobas, *Modern Graph Theory*. New York: Springer-Verlag, 1998.

[35] S. Salamatian, A. Zhang, F. d. P. Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft, "How to hide the elephant- or the donkey- in the room: Practical privacy against statistical inference for large data," in *Proceedings of the IEEE Global Conference on Signal and Information Processing*, 2013, pp. 269–272.

[36] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "A theory of utility and privacy of data sources," in *Proceedings of the IEEE International Symposium on Information Theory*, 2010, pp. 2642–2646.

[37] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Boston, MA, USA: Kluwer, 1992.

[38] F. Farokhi, I. Shames, and N. Batterham, "Secure and private control using semi-homomorphic encryption," *Control Engineering Practice*, vol. 67, pp. 13 – 20, 2017.

[39] J. Kim, C. Lee, H. Shim, J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Encrypting controller using fully homomorphic encryption for security of cyber-physical systems." *IFAC-PapersOnLine*, vol. 49, pp. 175 – 180, 2016.

[40] Y. Shoukry, K. Gatsis, A. Alanwar, G. J. Pappas, S. A. Seshia, M. Srivastava, and P. Tabuada, "Privacy-aware quadratic optimization using partially homomorphic encryption," in *Proceedings of the IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 5053–5058.

[41] C. Murguia, F. Farokhi, and I. Shames, "Secure and private implementation of dynamic controllers using semi-homomorphic encryption," *IEEE Transactions on Automatic Control*, 2020.

[42] O. Tan, D. Gündüz, and H. V. Poor, "Increasing smart meter privacy through energy harvesting and storage devices," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 1331–1341, 2013.

[43] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, 2013.

[44] Y. Mo and B. Sinopoli, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Transactions on Automatic Control*, vol. 61, pp. 2618–2624, 2016.

[45] A. Teixeira, I. Shames, H. Sandberg, and H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135 – 148, 2015.

[46] K. Torkkola and W. M. Campbell, "Mutual information in learning feature transformations," in *Proceedings of the 17th International Conference on Machine Learning*, 2000, pp. 1015–1022.

[47] W. Gao, S. Kannan, S. Oh, and P. Viswanath, "Estimating mutual information for discrete-continuous mixtures," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 5986–5997.

[48] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLOS ONE*, vol. 9, pp. 1–5, 2014.

[49] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-

redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.
[50] Q. Hu, L. Zhang, D. Zhang, W. Pan, S. An, and W. Pedrycz, "Measuring relevance between discrete and continuous features based on neighborhood mutual information," *Expert Systems with Applications*, pp. 10 737 – 10 750, 2011.
[51] K. J. Aström and B. Wittenmark, *Computer-controlled Systems (3rd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.
[52] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge University Press, 2012.
[53] M. Ross, *Introduction to Probability Models, Ninth Edition*. Orlando, FL, USA: Academic Press, Inc., 2006.
[54] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
[55] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.
[56] D. Bertsekas, *Nonlinear Programming*. Michigan, USA: Athena Scientific, 1999.
[57] J. Chen and R. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Hull, UK: Springer Publishing Company, Incorporated, 2012.
[58] K. Watanabe and D. M. Himmelblau, "Fault diagnosis in nonlinear chemical processes. part ii. application to a chemical reactor," *AIChE Journal*, vol. 29, 1983.
[59] C. Murguia and J. Ruths, "On model-based detectors for linear time-invariant stochastic systems under sensor attacks," *IET Control Theory Applications*, vol. 13, pp. 1051–1061, 2019.
[60] R. A. Horn, *Topics in Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.
[61] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. Studies in Applied Mathematics. Philadelphia, PA: SIAM, 1994, vol. 15.
[62] F. Zhang, *The Schur Complement and its Applications*. New York: Springer, 2005, vol. 4.