2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Infinite Horizon Privacy in Networked Control Systems: Utility/Privacy Tradeoffs and Design Tools

Haleh Hayati, Nathan van de Wouw, Carlos Murguia

Abstract— **We address the problem of synthesizing distorting mechanisms that maximize infinite horizon privacy for Networked Control Systems (NCSs). We consider stochastic LTI systems where information about the system state is obtained through noisy sensor measurements and transmitted to a (possibly adversarial) remote station via unsecured/public communication networks to compute control actions (a remote LQR controller). Because the network/station is untrustworthy, adversaries might access sensors and control data, and estimate the system state. To mitigate this risk, we pass sensor and control data through distorting (privacy-preserving) mechanisms before transmission and send the distorted data through the communication network. These mechanisms consist of a linear coordinate transformation and additive-dependent Gaussian vectors. We formulate the synthesis of the distorting mechanisms as a convex program where we minimize the infinite horizon mutual information (our privacy metric) between the system state and its optimal estimate at the remote station for a desired upper bound on the control performance (LQR cost) degradation induced by the distortion mechanism.**

I. Introduction

Recently, control systems have become increasingly distributed and networked. Networked Control Systems (NCSs) involve closing control loops over real-time communication networks. This allows controllers, sensors, and actuators to be connected through multipurpose networks, providing benefits such as increased system flexibility and ease of installation and maintenance. However, when estimation/control tasks in NCSs are performed by third parties, information sharing might result in private information leakage [1]- [4].

In NCSs, information about the plant state is obtained through sensor measurements and sent through communication networks to a remote station to perform computations, e.g., estimation or control. Shared information is correlated with private variables that carry sensitive information, e.g., the state itself (as it can reveal private system trajectories or it could be used to launch state-dependent attacks [5]), and references (as they can reveal manufactured products specs, tracked trajectories, and visited locations). If communication networks and/or the remote station are untrustworthy, adversaries might access the sensor and actuator signals and estimate the system state. To avoid this, we randomize the disclosed data before transmission using additive-dependent Gaussian random vectors and transmit the distorted data.

Using additive random noise is a common practice to enforce sensitive data privacy. For privacy of databases, a popular approach is Differential Privacy (DP) [6], where random noise is added to the queries' response to avoid leaking the database's private information. DP has also been applied to estimation and control problems [6], [7]. There are also techniques addressing privacy in dynamical systems from an information-theoretic perspective, see [8], [9]. In this line of work, privacy is characterized using information-theoretic metrics, e.g., mutual information. For information-theoretic methods, if the data to be kept private follows continuous probability distributions, the problem of finding the optimal additive noise to maximize privacy is difficult to solve [8]. This issue has been addressed by assuming the data to be kept private is deterministic [8]. However, for cyber-physical systems, the inherent system dynamics and unavoidable system and sensor noise lead to stochastic non-stationary data, and existing tools do not fit this problem.

Data privacy fundamentally differs between static data, such as databases, and dynamically correlated data, e.g., in feedback control systems. In networked control architectures, information flows bidirectionally between the remote station and the plant. The necessity of privacy masks for information flow directions is demonstrated in [2] by identifying the infinite horizon privacy consequences of bidirectional information flow in feedback control. To the best of our knowledge, no privacy-preserving design tools are offered for MIMO feedback control systems that minimize infinite horizon bidirectional information flow while maintaining a desired closed-loop control performance. There are works addressing information-theoretic infinite-horizon privacy [2], [10] for SISO systems. Also, in [11], [12], the infinite horizon privacy is considered for MIMO feedback control systems, but considering one direction of information flow.

This manuscript presents an optimization-based framework for synthesizing privacy-preserving Gaussian mechanisms that maximize privacy but keep distortion on control performance bounded. The proposed privacy mechanism consists of a coordinate transformation and additive Gaussian vectors that are designed to hide the private state of the plant [9]. We distort disclosed data in both information flow directions: the measurement data in the uplink direction that is transmitted from the plant to the remote station and the control data in the downlink direction from the remote station to the plant. We show that using coordinate transformations in the privacy mechanism (combined with additive Gaussian vectors) can significantly reduce information leakage in comparison with only adding stochastic vectors. When designing the privacy mechanisms, we also consider the trade-off between *privacy* and *control performance degradation*. As

Haleh Hayati, Nathan van de Wouw, and Carlos Murguia are with the Department of Mechanical Engineering, Dynamics and Control Group, Eindhoven University of Technology, The Netherlands. Emails: & h.hayati@tue.nl, & n.v.d.wouw@tue.nl, & c.g.murguia@tue.nl.

performance metric, we use the *LQR control cost* of the system when operating on distorted privacy-preserving data. We follow an information-theoretic approach to privacy. As *privacy metric*, we use the *mutual information* [13] between the system infinite state sequence $x^\infty = (x_1, ..., x_\infty)$ and its optimal estimate $\hat{x}^\infty = (\hat{x}_1, ..., \hat{x}_\infty)$ obtained by Kalman filtering given the infinite sequence of distorted disclosed data. Mutual information $I(x^\infty; \hat{x}^\infty)$ between the two jointly distributed infinite-dimensional vectors, x^∞ and $\hat{x}^\infty$, is a measure of the statistical dependence between them. We design the privacy mechanisms to minimize $I(x^\infty; \hat{x}^\infty)$ for a maximum level of control performance degradation of the closed-loop infinite horizon LQR control cost. We prove that the problem of finding sub-optimal additive random vectors covariance matrices and coordinate transformations can be cast into a constrained convex program (convex cost with LMI constraints). This work provides privacy-preserving design tools for MIMO feedback control systems to minimize infinite horizon bidirectional information flow by optimally distorting disclosed data while maintaining prescribed control performance. Providing infinite-horizon privacy is important for dynamical systems as adversaries can infer private information from disclosed data over time.

II. Problem Formulation

A. System Description

We consider the networked control architecture shown in Fig. 1. The dynamics of the plant is described as follows:

$$\mathcal{P} := \begin{cases} x_{k+1} = Ax_k + Bu_k + w_k, \\ y_k = x_k + h_k, \\ u_k = Ky_k \end{cases} \tag{1}$$

with time-index $k \in \mathbb{N}$, state $x_k \in \mathbb{R}^{n_x}$, measurable output $y_k \in \mathbb{R}^{n_y}$, controller $u_k \in \mathbb{R}^{n_u}$ with control feedback gain K, and matrices (A, B, K) of appropriate dimensions, $n_x, n_y, n_u \in \mathbb{N}$. The state and output disturbances w_k and h_k are multivariate i.i.d. Gaussian processes with zero mean and covariance matrices Σ^w, $\Sigma^h > 0$, respectively. The initial state x_1 is a Gaussian random vector with zero mean and covariance matrix $\Sigma_1^x := E[x_1 x_1^\top] > 0$. w_k, h_k, and x_1 are mutually independent. We assume that matrices $(A, B, \Sigma_1^x, \Sigma^w, \Sigma^h, K)$ are known, and (A, B) is stabilizable.

We consider the setting where the local plant is controlled by a remote station. The user who owns the plant transmits y_k to the remote station through an unsecured/public communication network to compute control actions (a remote LQR controller). Then, the control signal u_k is sent back to the user through the network. To characterize control performance for some given positive definite matrices Q and R, we introduce the associated infinite horizon LQR cost:

$$C_\infty(x, u) := \limsup_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} \mathbb{E}\left(x_k^\top Q x_k + u_k^\top R u_k\right), \tag{2}$$

where $\mathbb{E}(\cdot)$ denotes expectation. For privacy reasons, a full disclosure of the state x_k, $k \in \mathbb{N}$ is not desired. We aim to prevent adversaries from estimating x_k accurately. To this end, the user randomizes y_k before disclosure and requests the remote station to randomize control signals, u_k, before transmission to protect against inference at the network and remote station. The idea is to distort y_k and u_k through random affine transformations of the form:

$$\mathcal{M} := \begin{cases} \tilde{y}_k = Gy_k + v_k, \\ \tilde{u}_k = u_k + z_k, \end{cases} \tag{3}$$

where $G \in \mathbb{R}^{n_y \times n_y}$ is a linear transformation, and v_k and z_k are zero mean i.i.d. Gaussian processes with covariance matrices Σ^v and Σ^z. The distorted vectors $\tilde{y}_k$ and $\tilde{u}_k$ are transmitted over the network, see Fig. 1. Then, the closed-loop dynamics with privacy mechanism (3) is given by

$$\tilde{\mathcal{P}} := \begin{cases} \tilde{x}_{k+1} = A\tilde{x}_k + B\tilde{u}_k + w_k, \\ \tilde{y}_k = G\tilde{x}_k + Gh_k + v_k, \\ \tilde{u}_k = KG\tilde{x}_k + KGh_k + Kv_k + z_k. \end{cases} \tag{4}$$

with distorted state $\tilde{x} \in \mathbb{R}^{n_x}$. Here, we seek to synthesize G, Σ^v, and Σ^z, to make estimating the infinite horizon state trajectory $\tilde{x}_k$, $k \in \mathbb{N}$, as "hard" as possible from the disclosed data, $(\tilde{y}_k, \tilde{u}_k)$, $k \in \mathbb{N}$. We assume the adversary uses a steady-state Kalman filter designed to estimate the state *in the absence of privacy mechanisms*, and the adversary has prior knowledge of the system dynamics $(A, B, \Sigma_1^x, \Sigma^w, \Sigma^h)$ but does not have knowledge about the privacy mechanism (matrices (G, Σ^v, Σ^z). This creates an asymmetry we seek to exploit to increase privacy. The filter has this structure:

$$\begin{cases} \hat{x}_{k|k-1} = A\hat{x}_{k-1} + Bu_{k-1}, \\ \hat{x}_k = \hat{x}_{k|k-1} + L\left(\tilde{y}_k - \hat{x}_{k|k-1}\right), \end{cases} \tag{5}$$

with estimated state $\hat{x}_k \in \mathbb{R}^{n_x}$ and gain $L \in \mathbb{R}^{n_x \times n_y}$. The adversary designs the filter for the distortion-free system (1). Let ρ_k denote the estimation error *in the absence of the privacy distortions:* $\rho_k := x_k - \hat{x}_k$. The observer gain L is designed to minimize the asymptotic covariance matrix $\Sigma^\rho := \lim_{k\to\infty} E\left(\rho_k \rho_k^\top\right)$ [14]. As the system is observable, Σ^ρ always exists. Now let e_k denote the estimation error *in the presence of privacy distortions*, i.e., $e_k := \tilde{x}_k - \hat{x}_k$. Given the distorted dynamics (4), privacy mechanisms (3), and the estimator (5), the estimation error dynamics is governed by:

$$\begin{cases} \tilde{x}_{k+1} = (A + BKG)\tilde{x}_k + BK\tilde{v}_k + Bz_k + w_k, \\ e_{k|k-1} = Ae_{k-1} + Bz_{k-1} + w_{k-1}, \\ e_k = (I - L)e_{k|k-1} - L(G - I)\tilde{x}_k - L\tilde{v}_k, \end{cases} \tag{6}$$

where $\tilde{v}_k := Gh_k + v_k$.

B. Problem Formulation

The aim of our privacy scheme is to make the estimation of the infinite horizon state sequence, $\tilde{x}^\infty := (\tilde{x}_1, \ldots, \tilde{x}_\infty)$, from the disclosed distorted data, $\tilde{y}^\infty := (\tilde{y}_1, \ldots, \tilde{y}_\infty)$ and $\tilde{u}^\infty := (\tilde{u}_1, \ldots, \tilde{u}_\infty)$, as hard as possible without degrading the control performance excessively. Hence, when designing the distorting variables (G, Σ^v, Σ^h), we need to consider the *trade-off between privacy and performance*.

As privacy metric, we use the mutual information rate

$I_\infty(\tilde{x};\hat{x})$ [13] between $\tilde{x}^\infty$ and the infinite sequence of estimates $\hat{x}^\infty := (\hat{x}_1, ..., \hat{x}_\infty)$ obtained by Kalman filtering:

$$I_\infty(\tilde{x};\hat{x}) := \limsup_{N\to\infty} \frac{1}{N+1} I(\tilde{x}^N;\hat{x}^N), \tag{7}$$

where $I(\tilde{x}^N;\hat{x}^N)$ denotes standard mutual information [13].

We use the LQR cost in (2) to quantify control performance *in the absence of attacks*. To quantify the effect of the privacy mechanism (3) on the control performance, we introduced the associated distorted LQR control cost:

$$\tilde{C}_\infty(\tilde{x},\tilde{u}) := \limsup_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} \mathbb{E}\left(\tilde{x}_k^\top Q \tilde{x}_k + \tilde{u}_k^\top R \tilde{u}_k\right). \tag{8}$$

We aim to minimize $I_\infty(\tilde{x};\hat{x})$ subject to a constraint on the LQR cost increase due to the privacy mechanism, $\tilde{C}_\infty(\tilde{x},\tilde{u}) - C_\infty(x,u) \le \epsilon$, for a desired maximum control performance degradation level $\epsilon \in \mathbb{R}^+$, using as synthesis variables the mechanism matrices G, Σ^v, and Σ^z. In what follows, we present the problem we seek to address.

Problem 1 Given the system dynamics (1), distortion-free control performance (2), distorted control performance (8), privacy mechanism (3), distorted dynamics (4), Kalman filter (5), and maximum control degradation level $\epsilon > 0$, find the privacy mechanism variables, G, Σ^v, and Σ^z, as the solution of the following optimization problem:

$$\begin{cases} \min\limits_{G,\Sigma^v,\Sigma^z} I_\infty(\tilde{x};\hat{x}), \\ \text{s.t. } \tilde{C}_\infty(\tilde{x},\tilde{u}) - C_\infty(x,u) \le \epsilon. \end{cases} \tag{9}$$

III. Privacy Mechanism Design

To solve Problem 1, we first need to write the cost function and constraint in (9) in terms of the design variables.

A. Cost Function: Formulation and Convexity

Mutual information $I\left(\tilde{x}^N;\hat{x}^N\right)$ as used in (7) can be written in terms of uplink $I\left(\tilde{x}^N \to \hat{x}^N\right)$ (plant to the remote station) and downlink $I\left(\tilde{x}^N \leftarrow \hat{x}^N\right)$ (remote station to the plant) directed information flows [15]:

$$I\left(\tilde{x}^N;\hat{x}^N\right) = I\left(\tilde{x}^N \to \hat{x}^N\right) + I\left(\tilde{x}^N \leftarrow \hat{x}^N\right). \tag{10}$$

Then, the mutual information rate can be written as

$$I_\infty(\tilde{x};\hat{x}) := \limsup_{N\to\infty} \frac{1}{N+1} \left(I\left(\tilde{x}^N \to \hat{x}^N\right) + I\left(\tilde{x}^N \leftarrow \hat{x}^N\right)\right). \tag{11}$$

The decomposition of $I\left(\tilde{x}^N;\hat{x}^N\right)$ in terms of uplink and downlink directed information is essential in enabling to express mutual information as a stage additive function of covariance matrices. This allows writing $I_\infty(\tilde{x};\hat{x})$ in terms of the solution of Lyapunov equations/inequalities, which enables a convex reformulation of cost. In Lemma 1, we write the resulting expression of $I\left(\tilde{x}^N;\hat{x}^N\right)$ in terms of the design variables. Then, $I_\infty(\tilde{x};\hat{x})$ can be obtained by taking the limit in (11). Please refer to the proof of Lemma 1 for a step-by-step derivation of $I\left(\tilde{x}^N;\hat{x}^N\right)$.

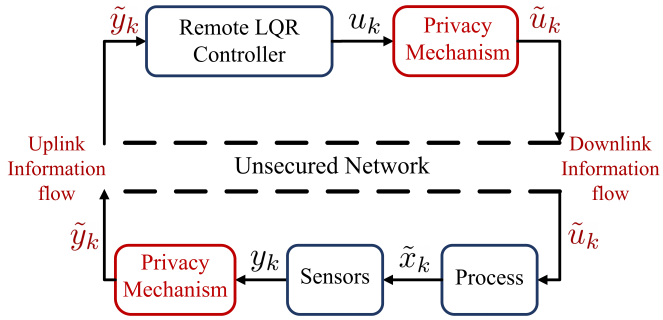

Fig. 1: System configuration.

Lemma 1 *Mutual information $I\left(\tilde{x}^N;\hat{x}^N\right)$ can be written in terms of G, Σ^v, and Σ^z, as follows:*

$$\begin{aligned} I\left(\tilde{x}^N;\hat{x}^N\right) = \sum_{k=1}^{N} \Bigg(& \frac{1}{2}\log\det\left(LG\Sigma^e_{k|k-1}G^\top L^\top + L\Sigma^{\tilde{v}}L^\top\right) \\ & - \frac{1}{2}\log\det\left(L\Sigma^{\tilde{v}}L^\top\right) - \frac{1}{2}\log\det\left(B\Sigma^z B^\top + \Sigma^w\right) \\ & + \frac{1}{2}\log\det\left(BK\Sigma^{\tilde{v}}K^\top B^\top + B\Sigma^z B^\top + \Sigma^w\right)\Bigg), \end{aligned} \tag{12}$$

with covariance matrices $\Sigma^e_{k|k-1} := \mathbb{E}(e_{k|k-1}e_{k|k-1}^\top)$ and $\Sigma^{\tilde{v}} := G\Sigma^h G^\top + \Sigma^v$.

Proof: See Appendix A in [16]. ■

Note that Σ^v only appears in the expression for $\Sigma^{\tilde{v}}$. Given $(G,\Sigma^{\tilde{v}})$, matrix Σ^v is fully determined and vice versa. That is, $(G,\Sigma^{\tilde{v}}) \to (G,\Sigma^v)$ is an invertible transformation. Therefore, we can pose Problem 1 in terms of either $\Sigma^{\tilde{v}}$ or Σ^v. Casting the problem in terms of $\Sigma^{\tilde{v}}$ allows us to formulate a convex cost and a convex constraint. Hereafter, we pose the problem in terms of $(G,\Sigma^{\tilde{v}})$. Once we have found optimal $(G,\Sigma^{\tilde{v}})$, we extract the optimal Σ^v as $\Sigma^v = \Sigma^{\tilde{v}} - G\Sigma^h G^\top$. Note, however, that due to the negative term $-G\Sigma^h G^\top$, the extracted Σ^v might be negative semidefinite, which is wrong as $\Sigma^v > 0$. To avoid this, we enforce that the extracted Σ^v is always positive definite in the synthesis program by adding $\Sigma^{\tilde{v}} - G\Sigma^h G^\top > \mathbf{0}$ as an extra constraint. This constraint can be equivalently written as the following linear inequality in $(G,\Sigma^{\tilde{v}})$ using Schur complement properties [17]:

$$\begin{bmatrix} \Sigma^{\tilde{v}} & G \\ G^\top & (\Sigma^h)^{-1} \end{bmatrix} > \mathbf{0}. \tag{13}$$

We use inequality (13) later when we solve the complete optimization problem to enforce that the optimal $(G,\Sigma^{\tilde{v}})$ leads to a positive definite Σ^v.

In Lemma 1, we have an expression of mutual information in terms of the design variables and the estimation error covariance $\Sigma^e_{k|k-1}$. Consider the closed-loop dynamics (6), and define the extended state $\zeta_k := \text{col}\left[e_{k|k-1}, \tilde{x}_k\right]$, we have

$$\begin{aligned} \zeta_{k+1} = & \begin{bmatrix} A(I-L) & -AL(G-I) \\ \mathbf{0} & A+BKG \end{bmatrix} \zeta_k \\ & + \begin{bmatrix} -AL & B & I \\ BK & B & I \end{bmatrix} \begin{bmatrix} \tilde{v}_k \\ z_k \\ w_k \end{bmatrix}. \end{aligned} \tag{14}$$

Because $(\tilde{v}_k, z_k, w_k)$ are all zero mean i.i.d. processes, the covariance of ζ_k, $\Sigma^\zeta_k := \mathbb{E}(\zeta_k \zeta_k^\top)$, satisfies the following:

$$\Sigma^\zeta_{k+1} = \mathcal{A}\Sigma^\zeta_k \mathcal{A}^T + \mathcal{B}, \tag{15}$$

where

$$\begin{cases} \mathcal{A} := \begin{bmatrix} A(I-L) & -AL(G-I) \\ \mathbf{0} & A+BKG \end{bmatrix}, \\ \mathcal{B} := \begin{bmatrix} -AL & B & I \\ BK & B & I \end{bmatrix} \begin{bmatrix} \Sigma^{\tilde{v}} & & \\ & \Sigma^z & \\ & & \Sigma^w \end{bmatrix} \begin{bmatrix} -AL & B & I \\ BK & B & I \end{bmatrix}^\top. \end{cases} \tag{16}$$

If $\mathcal{A}$ is Schur stable (which is always the case for $G = I$ by construction), the limit $\Sigma^\zeta := \lim_{k\to\infty} \Sigma^\zeta_k$, with Σ^ζ_k solution of (15), exists and coincides with the unique positive definite solution of the Lyapunov equation:

$$\mathcal{A}\Sigma^\zeta \mathcal{A}^T - \Sigma^\zeta + \mathcal{B} = \mathbf{0}. \tag{17}$$

Moreover, because $\zeta_k = \text{col}\left[e_{k|k-1}, \tilde{x}_k\right]$, we have

$$\Sigma^{\tilde{x}} := \lim_{k\to\infty} \Sigma^{\tilde{x}}_k = \begin{bmatrix} \mathbf{0} & I \end{bmatrix} \Sigma^\zeta \begin{bmatrix} \mathbf{0} & I \end{bmatrix}^\top, \tag{18}$$

$$\Sigma^e := \lim_{k\to\infty} \Sigma^e_{k|k-1} = \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \Sigma^\zeta \begin{bmatrix} I & \mathbf{0} \end{bmatrix}^\top, \tag{19}$$

which leads to Corollary 1 by taking the limit in (11).

Corollary 1 *The mutual information rate $I_\infty(\tilde{x};\hat{x})$ in (12) can be written in terms of G, $\Sigma^{\tilde{v}}$, and Σ^z, as follows:*

$$\begin{aligned} I_\infty(\tilde{x};\hat{x}) = &\frac{1}{2}\log\det\left(LG\Sigma^e G^\top L^\top + L\Sigma^{\tilde{v}} L^\top\right) \\ &- \frac{1}{2}\log\det\left(L\Sigma^{\tilde{v}} L^\top\right) - \frac{1}{2}\log\det\left(B\Sigma^z B^\top + \Sigma^w\right) \\ &+ \frac{1}{2}\log\det\left(BK\Sigma^{\tilde{v}} K^\top B^\top + B\Sigma^z B^\top + \Sigma^w\right), \end{aligned} \tag{20}$$

with $\Sigma^e = \lim_{k\to\infty} \Sigma^e_{k|k-1}$ as defined in (19).

The cost $I_\infty(\tilde{x};\hat{x})$ in (20) is non-convex in the design variables. The term $LG\Sigma^e G^\top L^\top$ is quadratic in G and Σ^e depends on the solution of the Lyapunov equation (17), which is quadratic in G. To tackle this, we derive a convex upper bound on the cost (20) and minimize it. We start with an upper bound, Σ, on the solution Σ^ζ of the Lyapunov equation (17). Having Σ and using (19) and monotonicity of $\log\det(\cdot)$ allow us to upper bound the first term in (20). We propose a convex program to find Σ in Lemma 2.

Lemma 2 *An upper bound Σ on the solution Σ^ζ of (17) can be found by solving the following convex program:*

$$\begin{cases} \min\limits_{\Sigma,\Pi_1,\Pi_2} \text{trace}(\Sigma), \\ \text{s.t.} \begin{bmatrix} \Sigma - \mathcal{B} & \mathcal{A}_0\Pi_1 + \mathcal{A}_1\Pi_2 \\ * & \Pi_1 + \Pi_1^\top - \Sigma \end{bmatrix} \geq \mathbf{0}, \\ \Pi_1 = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \mathbf{0} & \Pi_{13} \end{bmatrix}, \quad \Pi_2 = \begin{bmatrix} \mathbf{0} & \Pi_{21} \end{bmatrix}, \end{cases} \tag{21}$$

where

$$\begin{cases} \mathcal{A}_0 := \begin{bmatrix} A(I-L) & AL \\ \mathbf{0} & A \end{bmatrix}, \ \mathcal{A}_1 := \begin{bmatrix} -AL \\ BK \end{bmatrix}. \end{cases} \tag{22}$$

Proof: See Appendix B in [16]. ■

We defined new variables Π_1 and Π_2 to convexify the constraints in (21). Given (Π_1, Π_{21}), matrix G can be extracted as $G = \Pi_{21}\Pi_{13}^{-1}$ (see the proof of Lemma 2 in [16]). We can pose both cost and constraints in terms of either G or Π_{21}. Casting the problem in terms of Π_{21} allows us to linearize some constraints. Hereafter, we pose the problem in terms of (Π_1, Π_{21}). Once we have found optimal (Π_1, Π_{21}), we extract the optimal G using $\Pi_{21} = G\Pi_{13}$.

Lemma 2 allows casting the computation of an upper bound, Σ, on the solution, Σ^ζ, of the Lyapunov equation (17) as the solution of an optimization problem. Matrix Σ, obtained by solving (21), satisfies $\Sigma \geq \Sigma^\zeta = \lim_{k\to\infty} \Sigma^\zeta_k$. Therefore, given Σ, by (18)-(19), we also have the following upper bounds on $\Sigma^{\tilde{x}}$ and Σ^e:

$$\begin{cases} \Sigma^{\tilde{x}} = \lim\limits_{k\to\infty} \Sigma^{\tilde{x}}_k \leq N_{\tilde{x}} \Sigma N_{\tilde{x}}^\top, \\ \Sigma^e = \lim\limits_{k\to\infty} \Sigma^e_{k|k-1} \leq N_e \Sigma N_e^\top, \\ N_{\tilde{x}} := \begin{bmatrix} \mathbf{0} & I \end{bmatrix}, N_e := \begin{bmatrix} I & \mathbf{0} \end{bmatrix}. \end{cases} \tag{23}$$

In Corollary 1, the mutual information rate is written in terms of privacy mechanism variables and Σ^e. Hence, given (23) and monotonicity of the determinant function, an upper bound on $I_\infty(\tilde{x};\hat{x})$ in terms of Σ can be written as follows:

$$\begin{cases} I_\infty(\tilde{x};\hat{x}) \leq \frac{1}{2}\log\det\left(LGN_e\Sigma N_e^\top G^\top L^\top + L\Sigma^{\tilde{v}} L^\top\right) \\ \qquad - \frac{1}{2}\log\det\left(L\Sigma^{\tilde{v}} L^\top\right) - \frac{1}{2}\log\det\left(B\Sigma^z B^\top + \Sigma^w\right) \\ \qquad + \frac{1}{2}\log\det\left(BK_c\Sigma^{\tilde{v}} K_c^\top B^\top + B\Sigma^z B^\top + \Sigma^w\right). \end{cases} \tag{24}$$

So far, we have an upper bound (24) on the cost function in Problem 1 in terms of the solution Σ of program (21) and the mechanism parameters. However, (24) is still non-convex in G and Σ. In Lemma 3, we pose the problem of minimizing the right-hand side of (24) as a convex program. This reformulation is achieved using Schur complement properties, an epigraph reformulation of the minimization problem, and the monotonicity of the logdet$(\cdot)$ function. Moreover, as we will later need to combine the program in Lemma 2 with the convex reformulation of the bound in (24), we write, in Lemma 3, G in terms of Π_2 and Π_1 as we do in Lemma 2 ($G = \Pi_{21}\Pi_{13}^{-1}$, see the discussion below Lemma 2). This is necessary as we have to use the same coordinates in the reformulation of cost and constraints to be able to later solve all together as an optimization problem.

Lemma 3 *Consider the solution of the convex program:*

$$\begin{cases} \min\limits_{\Pi_{13},\Pi_{21},\Pi_3,\Pi_4,\Sigma^{\tilde{v}},\Sigma^z,\Sigma} \left(-\frac{1}{2}\text{logdet}(\Pi_3) - \frac{1}{2}\text{logdet}\left(\Pi_4\right) \right. \\ \left. -\frac{1}{2}\text{logdet}(L\Sigma^{\tilde{v}} L^\top) - \frac{1}{2}\text{logdet}\left(B\Sigma^z B^\top + \Sigma^w\right)\right) \\ \begin{cases} \text{s.t.} \quad 2I - \Pi_4 \geq \left(BK\Sigma^{\tilde{v}} K^\top B^\top + B\Sigma^z B^\top + \Sigma^w\right) \\ \begin{bmatrix} 2I - \Pi_3 - L\Sigma^{\tilde{v}} L^\top & L\Pi_{21} \\ * & \Pi_{13} + \Pi_{13}^\top - N_e\Sigma N_e^\top \end{bmatrix} \geq \mathbf{0}. \end{cases} \end{cases} \tag{25}$$

The resulting Σ, $\Sigma^{\tilde{v}}$, Σ^z, and $G = \Pi_{21}\Pi_{13}^{-1}$ minimize the

upper bound on $I_\infty(\tilde{x};\hat{x})$ in (24).

Proof: See Appendix C in [16]. ■

By Lemma 1, Lemma 2, and Lemma 3, a minimal upper bound on the cost $I_\infty(\tilde{x};\hat{x})$ can be achieved by solving the convex programs in (21) and (25). Then, if the constraints on positive definiteness of Σ^v (13) and control performance, $\tilde{C}_\infty(\tilde{x},\tilde{u})-C_\infty(x,u)\leq\epsilon$, can be written as convex functions of the decision variables, we can find optimal distorting mechanisms efficiently using off-the-shelf optimization algorithms. Regarding (13), it can be verified (see Appendix D in [16]) that (13) can be written in terms of (Π_{13},Π_{21}), the new decision variables, instead of the original G, as follows:

$$\left[\begin{array}{cc}\Sigma^{\tilde{v}} & \Pi_{21}\\ * & \Pi_{13}+\Pi_{13}^\top-\Sigma^h\end{array}\right]\geq\mathbf{0}. \tag{26}$$

We will add this (26) as a new constraint in the synthesis program. It remains to reformulate the control constraint.

B. Control Performance: Formulation and Convexity

Lemma 4 *The constraint on the LQR control cost:*

$$\tilde{C}_\infty(\tilde{x},\tilde{u})-C_\infty(x,u)\leq\epsilon, \tag{27}$$

can be formulated as the following set of LMIs:

$$\begin{cases} tr\left(Q\Sigma^{\tilde{x}}\right)+tr\left(\Pi_5\right)\\ \qquad+\operatorname{tr}\left(K^\top RK\Sigma^{\tilde{v}}+R\Sigma^z\right)\leq C_\infty(x,u)+\epsilon,\\ \left[\begin{array}{cc}\Pi_5 & R^{1/2}K\Pi_{21}\\ * & \Pi_{13}+\Pi_{13}^\top-\Sigma^{\tilde{x}}\end{array}\right]\geq\mathbf{0},\end{cases} \tag{28}$$

with new matrix variable Π_5 to be designed.

Proof: See Appendix E in [16]. ■

In Lemma 1 - Lemma 4, an upper bound on the cost function $I_\infty(\tilde{x};\hat{x})$ and the distortion constraint $\tilde{C}_\infty(x,\tilde{u})-C_\infty(x,u)\leq\epsilon$ are written in terms of convex functions of the design variables. We have, however, two cost functions in Lemma 2 and Lemma 3. The latter leads to a multi-objective optimization problem that can be solved by scalarizing the costs, i.e., introducing a single objective that represents a compromise between both of them. To this aim, we introduce $\alpha\in\mathbb{R}$, $\alpha>0$, as a weighting parameter and define a new cost as the weighted sum of the original ones (see (29)). Since our goal is to achieve a minimal mutual information rate characterizing information leakage, we seek the α that minimizes $I_\infty(\tilde{x};\hat{x})$ by performing a line search over α subject to all constraints in Lemma 1 - Lemma 4. In what follows, we pose the complete nonlinear convex program to find a sub-optimal solution for Problem 1 (sub-optimal as Lemma 3 minimizes an upper bound on the actual cost).

Theorem 1 *Consider the system dynamics* (1)*, distortion-free control performance* (2)*, distorted control performance* (8)*, privacy mechanism* (3)*, distorted dynamics* (4)*, Kalman filter* (5)*, and maximum control degradation level $\epsilon>0$, and matrices in* (16)*,* (22)*, and* (23)*. For a fixed $\alpha>0$, given the solution of the convex program in* (29)*, the mechanism variables G, Σ^v, and Σ^z, that minimize the upper bound on $I_\infty(\tilde{x};\hat{x})$ in* (24) *subject to the control performance degradation constraint, $\tilde{C}_\infty(x,\tilde{u})-C_\infty(x,u)\leq\epsilon$, are given by Σ^z, $G=\Pi_{21}\Pi_{13}^{-1}$, and $\Sigma^v=\Sigma^{\tilde{v}}-\Pi_{21}\Pi_{13}^{-1}\Sigma^h(\Pi_{21}\Pi_{13}^{-1})^\top$.*

$$\begin{cases} \min\limits_{\Pi_1,\Pi_2,\Pi_3,\Pi_4,\Pi_5,\Sigma^{\tilde{v}},\Sigma^z,\Sigma}\ \alpha(-\frac{1}{2}\log\det(\Pi_3)\\ \frac{1}{2}\log\det(L\Sigma^{\tilde{v}}L^\top)-\frac{1}{2}\log\det(\Pi_4)\\ -\frac{1}{2}\log\det(B\Sigma^zB^\top+\Sigma^w))+(1-\alpha)\operatorname{trace}(\Sigma),\\ \left[\begin{array}{cc}2I-\Pi_3-L\Sigma^{\tilde{v}}L^\top & L\Pi_{21}\\ * & \Pi_{13}+\Pi_{13}^\top-N_e\Sigma N_e^\top\end{array}\right]\geq\mathbf{0},\\ 2I-\Pi_4\geq\left(BK\Sigma^{\tilde{v}}K^\top B^\top+B\Sigma^zB^\top+\Sigma^w\right),\\ \left[\begin{array}{cc}\Sigma-\mathcal{B} & \mathcal{A}_0\Pi_1+\mathcal{A}_1\Pi_2\\ * & \Pi_1+\Pi_1^\top-\Sigma\end{array}\right]\geq\mathbf{0},\\ \operatorname{tr}\left(Q\Sigma^{\tilde{x}}\right)+\operatorname{tr}\left(\Pi_5\right)\\ \qquad+\operatorname{tr}\left(K^\top RK\Sigma^{\tilde{v}}+R\Sigma^z\right)\leq C_\infty(x,u)+\epsilon,\\ \left[\begin{array}{cc}\Pi_5 & R^{1/2}K\Pi_{21}\\ * & \Pi_{13}+\Pi_{13}^\top-N_{\tilde{x}}\Sigma N_{\tilde{x}}^\top\end{array}\right]\geq\mathbf{0},\\ \left[\begin{array}{cc}\Sigma^{\tilde{v}} & \Pi_{21}\\ * & \Pi_{13}+\Pi_{13}^\top-\Sigma^h\end{array}\right]>\mathbf{0},\ \ \Sigma^z>\mathbf{0},\ \ \Sigma>\mathbf{0}.\end{cases} \tag{29}$$

Proof: The expressions for the cost and constraints and convexity (linearity) of them follow from Lemma 1, Lemma 2, Lemma 3, Lemma 4, and (26). ■

IV. Illustrative case study

We illustrate the performance of our tools through a case study of a well-stirred chemical reactor with a heat exchanger. We use the discrete-time dynamics of the reactor introduced in [18] for the illustrative simulation study with matrices and more details as given in [16]. We implement the algorithm for two privacy mechanisms: first when the privacy mechanism is as in (3) and the second when the privacy mechanism does not include matrix transformation ($G=I$), to evaluate the effect of G in privacy mechanisms.

First, we show the effect of the control performance degradation level ϵ on the (mutual information-based) privacy cost function. Fig. 2 depicts the evolution of the optimal cost $I_\infty(\tilde{x};\hat{x})$ for increasing ϵ for both with and without matrix transformation in privacy mechanism cases shown by G and $G=I$, respectively. As expected, the objective function decreases monotonically for the increased maximum allowed control performance degradation in both cases. Besides, given that the control cost without privacy distortion is $C_\infty(x,u)=4.3615$, this figure illustrates that in the case of with matrix G, the infinite horizon optimal information leakage, which is shown by optimal $I_\infty(\tilde{x};\hat{x})$, can get very close to zero by a very small control performance degradation level ($\epsilon=0.07$). So, we can minimize the information leakage without degrading the control performance excessively. Hence, the comparison between the information leakage in these cases indicates that adding matrix transformation in the privacy mechanism (3) improves privacy by notably decreasing the information leakage.

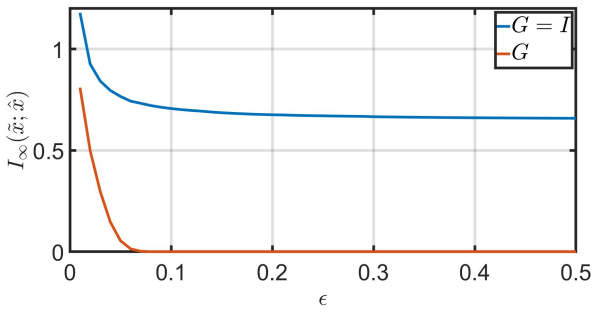

Fig. 2: Evolution of the optimal cost function (information leakage) based on increasing ϵ for with and without matrix transformation in the privacy mechanism.

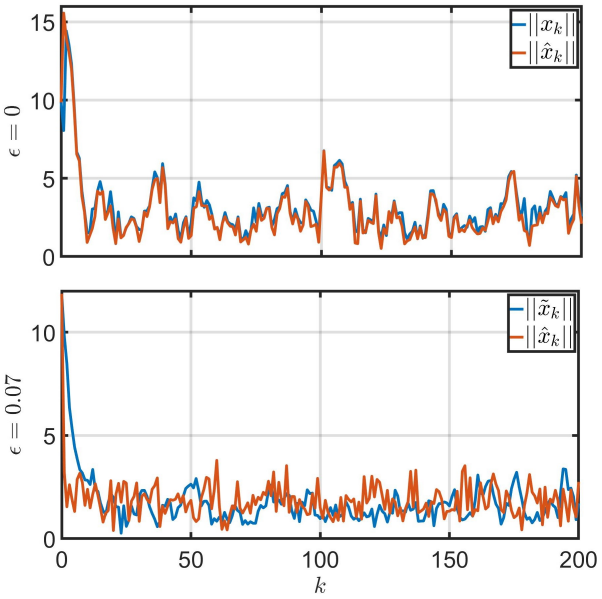

Fig. 3: Comparison between the norm of system state and its Kalman estimate for $\epsilon = 0, 0.07$.

Then, in Fig. 3, we depict the norm of the system state and its Kalman estimate with and without privacy distortion. The accuracy of state estimation based on distorted data $(\tilde{y}_k, \tilde{u}_k)$ with $\epsilon = 0.07$ is less than the estimation accuracy without privacy distortion ($\epsilon = 0$). The mean squared error for state estimation is 4.1304 and 1.5337 with and without the proposed privacy solution. Therefore, the proposed privacy tools can prevent accurate private state estimation.

V. Conclusions

In this paper, for a class of Networked Control Systems (NCSs), we have presented a detailed mathematical framework for synthesizing distorting mechanisms to minimize the infinite horizon information leakage induced by the use of public/unsecured communication networks. We have proposed a class of linear Gaussian distorting mechanisms to randomize sensor and control data before transmission to prevent adversaries from accurately estimating the system state. Furthermore, for the class of systems under study, we have fully characterized an information-theoretic metric to quantify the information between the system state and its optimal estimate given the distorted disclosed data at the remote station for a class of worst-case eavesdropping adversaries. Finally, given the maximum allowed level of control performance degradation (LQR cost), we have provided tools (in terms of convex programs) to design sub-optimal (in terms of maximizing privacy) distorting mechanisms.

VI. Acknowledgment

The research leading to these results has received funding from the European Union's Horizon Europe programme under grant agreement No 101069748 – SELFY project.

References

[1] 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.

[2] A. R. Pedram, T. Tanaka, and M. Hale, "Bidirectional information flow and the roles of privacy masks in cloud-based control," in *2019 IEEE Information Theory Workshop (ITW)*. IEEE, 2019, pp. 1–5.

[3] C. Murguia, F. Farokhi, and I. Shames, "Secure and private implementation of dynamic controllers using semihomomorphic encryption," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3950–3957, 2020.

[4] Y. Kawano and M. Cao, "Design of privacy-preserving dynamic controllers," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3863–3878, 2020.

[5] H. Hayati, C. Murguia, and N. van de Wouw, "Privacy-preserving anomaly detection in stochastic dynamical systems: Synthesis of optimal gaussian mechanisms," *arXiv preprint arXiv:2211.03698*, 2022.

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

[7] J. Cortés, G. E. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 4252–4272.

[8] 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.

[9] H. Hayati, C. Murguia, and N. Van De Wouw, "Finite horizon privacy of stochastic dynamical systems: A synthesis framework for gaussian mechanisms," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 5607–5613.

[10] S. Fang and Q. Zhu, "Fundamental limits of obfuscation for linear gaussian dynamical systems: An information-theoretic approach," in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 4574–4579.

[11] T. Tanaka, P. M. Esfahani, and S. K. Mitter, "Lqg control with minimum directed information: Semidefinite programming approach," *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 37–52, 2017.

[12] K. Yazdani, A. Jones, K. Leahy, and M. Hale, "Differentially private lq control," *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 1061–1068, 2022.

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

[14] K. J. Astrom and B. Wittenmark, *Computer-controlled Systems (3rd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.

[15] J. L. Massey and P. C. Massey, "Conservation of mutual and directed information," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.* IEEE, 2005, pp. 157–158.

[16] H. Hayati, N. van de Wouw, and C. Murguia, "Infinite horizon privacy in networked control systems: Utility/privacy tradeoffs and design tools," *arXiv preprint arXiv:2303.17519*, 2023.

[17] F. Zhang, *The Schur complement and its applications*. Springer Science & Business Media, 2006, vol. 4.

[18] C. Murguia, I. Shames, F. Farokhi, D. Nešić, and H. V. Poor, "On privacy of dynamical systems: An optimal probabilistic mapping approach," *IEEE Transactions on Information Forensics and Security*, 2021.