Privacy-Preserving Anomaly Detection in Stochastic Dynamical Systems: Synthesis of Optimal Gaussian Mechanisms

Haleh Hayati, Nathan van de Wouw, Carlos Murguia

Abstract—We present a framework for designing distorting mechanisms that allow remotely operating anomaly detectors while preserving privacy. We consider the problem setting in which a remote station seeks to identify anomalies using system input-output signals transmitted over communication networks. However, disclosing true data of the system operation is not desired as it can be used to infer private information – modeled here as a system private output. To prevent accurate estimation of private outputs by adversaries, we pass original signals through distorting (privacy-preserving) mechanisms and send the distorted data to the remote station (which inevitably leads to degraded monitoring performance). We formulate the design of these mechanisms as a privacy-utility trade-off problem. We cast the synthesis of dependent Gaussian mechanisms as the solution of a convex program where we seek to maximize privacy quantified using information-theoretic metrics (mutual information and differential entropy) over a finite window of realizations while guaranteeing a bound on monitoring performance degradation.

I. INTRODUCTION

In today’s hyperconnected world, the rapid advancement of science and technology has resulted in massive user data being collected and processed by numerous companies over public networks. Companies use this data to provide personalized services and targeted advertising. However, these technologies have come at the cost of a significant loss of privacy. Adversaries with sufficient resources can infer sensitive (private) information about users/systems from public data available on the internet and/or unsecured networks/servers. This is why researchers from different fields have been drawn to privacy and security of Cyber-Physical Systems (CPSs), [1]–[5].

In most engineering applications, information about the state of systems is obtained through sensor measurements and sent to a remote station through communication networks for signal processing and decision-making. One of these applications is remote anomaly detection, where a remote station seeks to identify anomalies based on system data transmitted over communication networks. However, if the communication network is public/unsecured and/or the remote station is not trustworthy, adversaries might access and estimate systems’ private data. To avoid this, we propose a distorting mechanism to randomize sensor and input data before transmission using additive dependent Gaussian random vectors designed to hide the private part of the state. The aim of this work is to devise synthesis tools to design these mechanisms such that private information is protected and, at the same time, the implementation of remote anomaly detectors is still possible using distorted data. Therefore, the synthesis tool must consider the trade-off between privacy and monitoring performance. We remark that we protect against both eavesdropping adversaries and the (potentially untrustworthy) remote station. We aim to develop a privacy scheme that considers no trusted party and seeks to destroy information in the direction of private outputs as much as possible while keeping monitoring performance degradation bounded.

In this manuscript, we follow an information-theoretic approach to privacy. As privacy metric, we propose a combination of mutual information and entropy [6] between disclosed and private data. To characterize anomaly detection performance, we consider the false alarm rate of standard chi-squared detectors as performance metric. Hence, the synthesis algorithm seeks to optimize the privacy metric by selecting the stochastic properties of the injected dependent noise (their joint covariance matrices over a finite window of realizations) while not degrading the monitoring false alarm rate by more than a desired amount. We prove that this problem can be posed as a constrained convex optimization problem – log-determinant cost with linear matrix inequality constraints.

Using additive random noise is common practice to enforce privacy of sensitive data. Differential Privacy (DP) is a popular method in the context of privacy of dynamical systems [3], [7]. However, in most cases, privacy is defined at every time-step, i.e., they do not consider information disclosed over a history of observations. While this gives some privacy guarantees, it is well known that for dynamical systems, the more data is collected, the more accurate the input/state estimation can be. For instance, in the deterministic case, if the system is observable and of dimension \( n \), after \( n \) time-steps, the state can be completely reconstructed without error. Therefore, if the privacy scheme is to avoid input/state estimation, giving guarantees at every time-step is certainly not sufficient. Moreover, injecting i.i.d. noise at every time-step [8], [9] is also conservative as there is an obvious correlation among measurements and inputs in dynamical systems.

There are existing techniques to address privacy in dynamical systems from an information-theoretic perspective [2], [10], [11]. In these methods, privacy is characterized using information-theoretic metrics – e.g., mutual information, entropy, and Fisher information. However, independently of the metric being used, 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 [1]. This issue has been addressed by assuming the data to be kept private is deterministic [1]. But, in a CPSs context, the inherent system dynamics and unavoidable system and sensor noise lead to stochastic non-stationary data, and thus, existing tools do not fit. In [12], [13], we have started exploring these ideas assuming Gaussianity in system disturbances and
without considering any particular application at the remote station – we mainly focused on formulating the privacy-utility problem over a finite horizon and characterized utility degradation by the expected squared distortion between distorted and original data. Motivated by these results, in this manuscript, we propose a general class of Gaussian mechanisms and seek the optimal parametrization of the additive dependent noise distribution (its joint covariance matrix over a window of realizations) given the system dynamics and the anomaly detector at the remote station.

To the best of the authors’ knowledge, there is only one related result in the literature. For remote anomaly detection in stochastic dynamical systems, the authors in [8] consider differential privacy and provide tools to select the variance of i.i.d. additive noise so that disclosed data guarantees a certain level of differential privacy. Their results consider privacy of input signals (so no private state-dependent outputs), and at every time-step, i.e., they do not consider information disclosed over a history of observations. Although these results provide a DP privacy guarantee on input data, the use of i.i.d. noise at every time-step leads, in general, to conservative results without considering any particular application at the remote station. Moreover, results in [8] do not provide a mathematical characterization of privacy-detection tradeoffs nor provide synthesis methods to design optimal distorting mechanisms (in the sense of optimizing these tradeoffs). Taking into account these points, our contributions are:

- The characterization of proper privacy and detection metrics tailored to the specific application domain (private remote anomaly detection in stochastic dynamical systems) in terms of system matrices. This allows quantifying (enforcing) given (desired) trade-off between privacy and anomaly detection performance.

- A synthesis framework (in terms of a series of convex programs) to design additive-dependent Gaussian mechanisms over finite horizons that minimize information leakage while guaranteeing a desired anomaly detection performance. To accomplish this, we have reformulated the original design optimization problem (which is non-convex in cost and constraints) as a log-determinant cost with linear matrix inequality constraints for the lifted system dynamics.

- We prove that our framework can handle prior distributions in the system dynamics that are not necessarily Gaussian, as long as they are log-concave [14]. The latter further generalizes the class of stochastic dynamical systems that our framework can handle. Due to space limitations, the latter is included in an extended pre-print of this paper [15].

II. SYSTEM DESCRIPTION AND ANOMALY DETECTION

A. System Description

We consider discrete-time stochastic systems of the form:

\[
\begin{align*}
    x_{k+1} &= Ax_k + Bu_k + t_k + G\delta_k, \\
    y_k &= Cx_k + w_k + H\delta_k, \\
    s_k &= Dx_k,
\end{align*}
\]  

(1)

with time-index \( k \in \mathbb{N} \), state \( x_k \in \mathbb{R}^{n_s} \), measurable output \( y_k \in \mathbb{R}^{n_y} \), known input \( u_k \in \mathbb{R}^{n_u} \), private performance output \( s_k \in \mathbb{R}^{n_s} \), and matrices \( A, B, C, D, G, H \) of appropriate dimensions. Vector \( \delta_k \in \mathbb{R}^{n_s} \) models changes in the system dynamics due to faults or anomalies (thus, in the absence of anomalies \( \delta_k = 0 \)). Matrix \( D \) is full row rank, and the pair \((A, C)\) is detectable. The state disturbance \( t \) and the output disturbance \( w \) are multivariate i.i.d. Gaussian processes with zero mean and positive definite covariance matrices \( \Sigma_t \) and \( \Sigma_w \), respectively. The initial state \( x_1 \) is assumed to be a Gaussian random vector with \( E[x_1] = \mu_x^0 \in \mathbb{R}^{n_s} \) and covariance matrix \( \Sigma_x^0 \in \mathbb{R}^{n_s \times n_s}, \Sigma_x^0 > 0 \). Disturbances \( t_k \) and \( w_k \) and the initial condition \( x_1 \) are mutually independent. We assume that matrices (vectors) \((A, B, C, D, \Sigma_t^0, \mu_t^0, \Sigma_w^0)\) and input \( u_k \) are known for all \( k \).

B. Anomaly Detection

The aim of the anomaly detection algorithm is to identify anomalies (i.e., when \( \delta_k \neq 0 \) in (1)) using system input-output signals. Here, we consider standard Kalman filter-based chi-squared change detection procedures [16]. The main idea is to use an estimator to anticipate the system evolution in the absence of anomalies. This prediction is subsequently compared with actual measurements from sensors. If the difference between what is measured and the estimation (commonly referred to as residuals) is larger than expected, there might be
an anomaly in the system. We use steady-state Kalman filters as the state estimator.

1) Steady-state Kalman Filter (anomaly-free case): Consider the following one step-ahead Kalman filter \[17\] for (1):

\[
\hat{x}_{k+1} = A \hat{x}_k + Bu_k + L(y_k - C \hat{x}_k),
\]

with estimated state \( \hat{x}_k \in \mathbb{R}^{n_x} \), \( \hat{x}_1 = E[x_1] \), and output injection gain matrix \( L \in \mathbb{R}^{n_x \times n_y} \). Define the estimation error \( e_k := y_k - \hat{y}_k \). The optimal filter gain \( L \) minimizing the trace of the asymptotic estimation error covariance matrix \( P := \lim_{k \to \infty} E[e_k r_k^T] \) in the absence of anomalies is given by

\[
L := \left( APCT^T + \Sigma w + CPC^T \right)^{-1},
\]

where \( P \) is the solution of the Riccati equation:

\[
APA^T + P = \Sigma = \left( APCT^T + \Sigma w + CPC^T \right)^{-1} CPA^T.
\]

Equation (4) always has a unique solution \( P \) under the assumption of detectability of the pair \((A,C)\) [17].

2) Residuals and Hypothesis Testing: Consider the process dynamics \[1\] and the steady-state Kalman filter \[2\], and define the residual sequence \( r_k := y_k - C \hat{x}_k \), which can be shown to evolve according to the difference equation:

\[
\begin{align*}
\hat{e}_{k+1} &= (A - LC) \hat{e}_k + t_k + Lw_k + (G - LH) \delta_k, \\
r_k &= C \hat{e}_k + w_k + H \delta_k.
\end{align*}
\]

If there are no anomalies \( (\delta_k = 0) \), the mean and the covariance matrix of the residual are as follows:

\[
\begin{align*}
E[r_k] &= CE[\hat{e}_k] + E[w_k] = 0_{n_y \times 1}, \\
E[r_k r_k^T] &= CPC^T + \Sigma_w =: \Sigma \in \mathbb{R}^{n_y \times n_y}.
\end{align*}
\]

For this residual, we identify two hypotheses to be tested: \( \mathcal{H}_0 \) the normal mode (no anomalies) and \( \mathcal{H}_1 \) the faulty mode (with anomalies). Under the normal mode, \( \mathcal{H}_0 \), the statistics of the residual is given by (6). However, in the presence of anomalies (\( \mathcal{H}_1 \)), we expect that the statistics of the residual are different from those in the normal mode, i.e.,

\[
\mathcal{H}_1 : \begin{cases} E[r_k] \neq 0_{n_y \times 1}, \text{ and/or} \\ E[r_k r_k^T] \neq \Sigma. \end{cases}
\]

3) Distance Measure and Chi-Squared Procedure: The input to the chi-squared procedure is a distance measure \( z_k \in \mathbb{R} \), i.e., a measure of how deviated the estimator is from the sensor measurements. We use the quadratic distance measure

\[
z_k = r_k^T \Sigma^{-1} r_k,
\]

where \( r_k \) and \( \Sigma \) are the residual sequence and its covariance matrix introduced in (6). If there are no anomalies \( (\delta_k = 0) \), i.e. \( E[r_k] = 0 \) and \( E[r_k r_k^T] = \Sigma \), it follows that

\[
\begin{align*}
E[z_k] &= \text{tr}[\Sigma^{-1} \Sigma] + E[r_k^T \Sigma^{-1} r_k] = n_y, \\
\text{var}[z_k] &= 2 \text{tr}[\Sigma^{-1} \Sigma \Sigma^{-1} \Sigma] + 4 E[r_k^T \Sigma^{-1} \Sigma \Sigma^{-1} E[r_k] = 2n_y,
\end{align*}
\]

see, e.g. [20] for details. Because \( r_k \sim \mathcal{N}(0, \Sigma) \), \( z_k \) follows a chi-squared distribution with \( n_y \) degrees of freedom. The chi-squared procedure is characterized by \( z_k \) and its cumulative distribution as follows:

Chi-Squared procedure:
If \( z_k = r_k^T \Sigma^{-1} r_k > \alpha \), \( \hat{k} = k \).

Design parameter: threshold \( \alpha \in \mathbb{R}_+ \).
Output: alarm time(s) \( \hat{k} \).

The idea is that if \( z_k \) exceeds the threshold \( \alpha \), alarms are triggered, and the time instant \( k \) is identified as an alarm time (denoted by \( \hat{k} \)). The parameter \( \alpha \) is selected to satisfy a desired false alarm rate \( A^* \). Assume that there are no anomalies and consider the chi-squared procedure with threshold \( \alpha \in \mathbb{R}_+ \), \( r_k \sim \mathcal{N}(0, \Sigma) \). Let \( \alpha = \alpha^* := 2P^{-1}(\frac{\alpha}{2}, 1 - A^*) \), where \( P^{-1}(\cdot, \cdot) \) denotes the inverse regularised lower incomplete gamma function, then the false alarm rate \( A \), induced by the threshold \( \alpha \), satisfies \( A = A^* \) (see [19] for proof and details).

C. Privacy Mechanism and Adversarial Capabilities

We aim to prevent adversaries from accurately estimating the private output \( s_k \) (see [1]). To this end, we randomize measurements \( y_k \) and input signals \( u_k \) before transmission and send the corrupted data to the remote station instead. The idea is to randomize \( y_k \) and \( u_k \) through a stochastic mapping \( M(k, y_k, u_k) \) of the form:

\[
M(k, y_k, u_k) := \begin{cases} \hat{y}_k = y_k + v_k, \\ \hat{u}_k = u_k + j_k, \end{cases}
\]

for some Gaussian processes, \( v_k \sim \mathcal{N}(0, \Sigma_v^k) \) and \( j_k \sim \mathcal{N}(0, \Sigma_j^k) \) with zero mean and time-varying covariance matrices \( \Sigma_v^k \) and \( \Sigma_j^k \) to be designed for each time instant. We consider \( v_k \) and \( j_k \) as dependent Gaussian processes, implying that privacy noises sequences, \( v^K = \text{col}(v_1, \ldots, v_K) \) and \( j^K = \text{col}(j_1, \ldots, j_K) \) for \( K := \{1, \ldots, K\} \) are dependent sequences. The randomized vectors \( \hat{y}_k \) and \( \hat{u}_k \) are disclosed to the remote station to be used by the remote anomaly detector to detect faults/attacks and trigger alarms, see Figure 1. Since at the remote station, only the received distorted signals \( (\hat{y}_k, \hat{u}_k) \) can be used, the dynamics of the system used at the remote
station in the presence of anomalies and privacy distortions is as follows:

\[
\begin{align*}
    x_{k+1} &= Ax_k + Bu_k + t_k + G\delta_k \\
    \hat{y}_k &= y_k + v_k = Cx_k + w_k + H\delta_k + v_k.
\end{align*}
\] (11)

That is, the original plant driven by \( \hat{u}_k \) and producing \( \hat{y}_k \). 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 distorted sensor measurements \( \hat{y}_k \) and distorted input signals \( \hat{u}_k \), but also have prior knowledge of the dynamics and the stochastic properties of the system, i.e., matrices \( (A, B, C, D, \Sigma^y, \Sigma^u, \Sigma^w) \) are known by the adversary. Moreover, the adversary also knows the means and covariance matrices of \( \hat{y}_k, (\mu^y_k, \Sigma^y_k) \), and \( \hat{u}_k, (\mu^u_k, \Sigma^u_k) \), as these can be estimated from the disclosed data \( (\hat{y}_k, \hat{u}_k) \). We assume that the adversary uses a linear MMSE estimator (see [21]) to reconstruct the private output \( s_k \), which, for jointly Gaussian vectors, produces the best estimation performance among all unbiased estimators [21]. In practice, actual adversaries would typically not have all the capabilities that we assume here. However, if we maximize privacy under such worst-case adversaries, we ensure that adversaries with less capabilities perform even worse (or equal at most).

\[ \Pr[\hat{z}_k > \alpha] < A^* + \epsilon, \] (15)
which is equivalent to

$$\Pr[\hat{z}_k \leq \alpha] > 1 - A^* - \epsilon,$$

(16)

where $\epsilon$ is the false alarm distortion level. Using the definition of Cumulative Distribution Function (CDF), we can write (16) as

$$F_{\hat{z}_k}(\alpha) > 1 - A^* - \epsilon,$$

(17)

where $F_{\hat{z}_k}(\alpha)$ is the CDF of $\hat{z}_k$, i.e., $F_{\hat{z}_k}(\alpha) = \Pr[\hat{z}_k \leq \alpha]$. Therefore, the trade-off optimization problem for maximizing privacy while considering the above constraint on the false alarm rate can be written as follows:

**Problem 1** Given the system dynamics (1), $k \in K$ with $K = \{1, \ldots, K\}$, $K \in \mathbb{N}$, desired false alarm rate $A^*$, maximum distortion level $\epsilon \in \mathbb{R}^+$, the chi-squared procedure threshold $\alpha$, the distorting mechanism (10), and the remote Kalman filter (12), find the sequence of covariance matrices $\Sigma_k^0$ and $\Sigma_k^1$ in (10) as the solution to the following optimization problem:

$$\begin{align*}
\min_{\Sigma_k^0, \Sigma_k^1, k \in K} & \quad I[s^K; \hat{y}^K] - h[j^K], \\
\text{s.t.} & \quad F_{\hat{z}_k}(\alpha) > 1 - A^* - \epsilon, \\
\text{and} & \quad (u^K, j^K, \hat{y}^K) \text{ mutually independent.}
\end{align*}$$

(18)

**IV. Solution to Problem 1**

To solve the optimization problem (18), we first need to write the cost function and constraints in terms of our design variables $\Sigma_k^0$ and $\Sigma_k^1$, $k \in K$.

**A. Cost Function: Formulation and Convexity**

We first formulate the cost function, $I[s^K; \hat{y}^K] - h[j^K]$, in terms of the design variables $\Sigma_k^0$ and $\Sigma_k^1$. The differential entropy $h[j^K]$ is fully characterized by the covariance matrix of $j^K$, $\Sigma_k^1 := E[j^K j^K^T]$. Then, $h[j^K]$ is given by (6)

$$h[j^K] = \frac{1}{2} \log \det (\Sigma_k^1) + \frac{K_{nu}}{2} + \frac{K_{nu}}{2} \log(2\pi).$$

(19)

The mutual information $I[s^K; \hat{y}^K]$ can be written in terms of differential entropies as $I[s^K; \hat{y}^K] = h[s^K] + h[\hat{y}^K] - h[s^K, \hat{y}^K]$, see (6). Moreover, these entropies are fully characterized by the covariance matrices of the corresponding random vectors $\hat{y}^K$. So, to write $I[s^K; \hat{y}^K]$ in terms of $\Sigma_k^0$, we need to write the covariance matrices of $s^K$ and $\hat{y}^K$, and their joint covariance in terms of $\Sigma_k^0$. By lifting the system dynamics (1) over $K$, the stacked vector $((\hat{y}^K)^T, (s^K)^T)^T \in \mathbb{R}^{(n_y+n_z)K}$ can be written as

$$\begin{align*}
\begin{bmatrix}
\hat{y}^K \\
\hat{s}^K
\end{bmatrix} &= \begin{bmatrix}
\hat{C}_K \\
\hat{D}_K
\end{bmatrix} F_K x_1 + \begin{bmatrix}
\hat{C}_K \\
\hat{D}_K
\end{bmatrix} J_K u^{K-1},
\end{align*}$$

(20)

with sequence of inputs $u^{K-1} = (u_1^T, \ldots, u_{K-1}^T)^T$, stacked matrices $N_K := J_K(I_{K-1} \otimes B)$, $\hat{C}_K := I_K \otimes C$, $\hat{D}_K := I_K \otimes D_K$, and

$$\begin{align*}
\begin{bmatrix}
F_K := & I \ A^T \ \ldots \ \ (A^T)^{K-1}^T, \\
J_K := & \begin{bmatrix}
0 & 0 & 0 & \ldots & 0 \\
I & 0 & 0 & \ldots & 0 \\
A & I & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(A)^{K-2} & (A)^{K-3} & (A)^{K-4} & \ldots & I
\end{bmatrix}
\end{align*}$$

(21)

Let $\Sigma_k^0 \in \mathbb{R}^{K_{nu} \times K_{nu}}$ and $\Sigma_k^1 \in \mathbb{R}^{K_{nu} \times K_{nu}}$ denote the non-diagonal covariance matrix of the stacked additive dependent vectors $v^K$ and $j^K$. In the following lemma, we give a closed-form expression of the joint density of $(\hat{y}^K, s^K)$.

**Lemma 1** with mean $\mu_{\hat{y}, s}^K := \begin{bmatrix} \mu_{\hat{y}}^K \\
\mu_{s}^K
\end{bmatrix}$ and covariance matrix $\Sigma_{\hat{y}, s}^K \in \mathbb{R}^{(n_y+n_z)K_{nu} \times (n_y+n_z)K_{nu}}$, $\Sigma_{\hat{y}, s}^K > 0$:

$$\begin{align*}
\mu_{\hat{y}, s}^K & := \begin{bmatrix} \hat{C}_K \\
\hat{D}_K
\end{bmatrix} F_K \mu_1 + \begin{bmatrix} \hat{C}_K \\
\hat{D}_K
\end{bmatrix} N_K u^{K-1}, \\
\Sigma_{\hat{y}, s}^K & := \begin{bmatrix} (I) \Sigma_{s} \Sigma_{\hat{y}} & (I) \Sigma_{\hat{y}} \Sigma_{\hat{y}}^{T} \\
(I) \Sigma_{\hat{y}} \Sigma_{\hat{y}}^{T} & (I) \Sigma_{\hat{y}} \Sigma_{\hat{y}}^{T}
\end{bmatrix} + \begin{bmatrix} \hat{C}_K \\
\hat{D}_K
\end{bmatrix} F_K \Sigma_{u} F_K^{T} \begin{bmatrix} \hat{C}_K \\
\hat{D}_K
\end{bmatrix}^{T} \\
+ \begin{bmatrix} \hat{C}_K \\
\hat{D}_K
\end{bmatrix} J_K (I_{K-1} \otimes \Sigma_{s}) J_K^{T} \begin{bmatrix} \hat{C}_K \\
\hat{D}_K
\end{bmatrix}^{T}.
\end{align*}$$

(22)

(23)

**Proof:** To simplify notation, we introduce the stacked vector $\Theta^K := ((\hat{y}^K)^T, (s^K)^T)^T$. By assumption, the initial condition $x_1$, and the processes, $t_k$, $u_k$, and $v_k$, $k \in \mathbb{N}$, are mutually independent, and $x_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$, $t_k \sim \mathcal{N}(0, \Sigma_t)$, and $u_k \sim \mathcal{N}(0, \Sigma_u)$. Then, see (24) for details, we have $L_1 x (1) \sim \mathcal{N}(L_1 \mu_1, L_1 \Sigma_1 L_1^{T})$, $L_{21} x (K-1) \sim \mathcal{N}(0, L_{21} L_{21}^{T})$, and for any deterministic matrices $L_j$, $j = 1, 2, 3, 4$, of appropriate dimensions. It follows that the stacked vector $\Theta^K$ given in (20) is the sum of a deterministic vector, $\begin{bmatrix} \hat{C}_K^{T}, \hat{D}_K^{T}\end{bmatrix} N_K u^{K-1}$, and four independent normally distributed vectors. Therefore, $\Theta^K$ follows a multivariate normal distribution with $E[\Theta^K] = \mu_{\hat{y}, s}^K$ as in (22). By inspection, using the expression of $\Theta^K$ in (20), mutual independence among $x_1$, $t_k$, $u_k$, and $v_k$, $k \in \mathbb{N}$, and the definition of $\Sigma_1$, $\Sigma_1 = E[(x_1 - \mu_1)(x_1 - \mu_1)^{T}]$, it can be verified that the covariance matrix of $\Theta^K$, $E[(\Theta^K - E(\Theta^K))(\Theta^K - E(\Theta^K))^{T}]$, is given by $\Sigma_{\hat{y}, s}^K$ in (23). It remains to prove that the distribution of $\Theta^K$ is not degenerate, that is, $\Sigma_{\hat{y}, s}^K > 0$. Note that $\Sigma_{\hat{y}, s}^K$ in (23) can be written as

$$\Sigma_{\hat{y}, s}^K = \begin{bmatrix} \Sigma_{\hat{y}}^\Sigma \Sigma_{\hat{y}}^\Sigma & \Sigma_{\hat{y}}^\Sigma \Sigma_{\hat{y}}^\Sigma \\
\Sigma_{\hat{y}}^\Sigma \Sigma_{\hat{y}}^\Sigma & \Sigma_{\hat{y}}^\Sigma \Sigma_{\hat{y}}^\Sigma
\end{bmatrix} + \begin{bmatrix} \hat{C}_K Q \hat{D}_K^{T} \\
\hat{D}_K^{T} Q\hat{D}_K
\end{bmatrix}$$

(24)

with $Q := F_K \Sigma_{u} F_K^{T} + J_K (I_{K-1} \otimes \Sigma_{s}) J_K^{T}$. Necessary and sufficient conditions for $\Sigma_{\hat{y}, s}^K > 0$ are its right-lower block is
positive definite \((\tilde{D}_K \Sigma K \tilde{D}_K^T > 0)\) and that the Schur complement of block \(\tilde{D}_K \Sigma K \tilde{D}_K\) of \(\Sigma^{\tilde{s},s}_K\), denoted as \(\Sigma^{\tilde{s},s}_K/(\tilde{D}_K \Sigma K \tilde{D}_K^T)\), is positive definite (25, Theorem 1.12). 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 (26, Theorem 4.2.15). Note that \(Q\) can be factored as follows:

\[
Q = \begin{bmatrix} F_K & J_K \end{bmatrix} \begin{bmatrix} \Sigma_0^T & 0 \\ 0 & I_{K-1} \otimes \Sigma' \end{bmatrix} \begin{bmatrix} F_K \ I_K \end{bmatrix}.
\]

That is, \(Q\) is a linear transformation of the block diagonal matrix \(Q'\) above. By inspection, it can be verified that matrix \(P = [F_K \ I_K]\), is lower triangular with identity matrices on the diagonal; thus, \(P\) is invertible. It follows that \(Q = PQ'P^T\) is a congruence transformation of \(Q'\) (27). Hence, \(Q\) is positive definite if and only if the block diagonal matrices of \(Q'\) are positive definite (27). Matrices \(\Sigma_0^T\) and \(\Sigma'\) are positive definite by assumption (which implies \(I_{K-1} \otimes \Sigma' > 0\)), and thus we can conclude that \(Q > 0\), which implies \(\tilde{D}_K \Sigma K \tilde{D}_K^T > 0\), because \(\tilde{D}_K\) is full row rank. Then, we need to prove that the Schur complement \(\Sigma^{\tilde{s},s}_K/(\tilde{D}_K \Sigma K \tilde{D}_K^T)\), is positive definite. This Schur complement is given by

\[
\Sigma^{\tilde{s},s}_K/(\tilde{D}_K \Sigma K \tilde{D}_K^T) = I_K \otimes \Sigma^w + \Sigma_K + \tilde{C}_K (Q - Q \tilde{D}_K \Sigma K \tilde{D}_K^T)^{-1} \tilde{D}_K Q \tilde{C}_K^T.
\]

Since \((I_K \otimes \Sigma^w) + \Sigma_K\) is positive definite, a sufficient condition for \(\Sigma^{\tilde{s},s}_K/(\tilde{D}_K \Sigma K \tilde{D}_K^T) > 0\) is

\[
Q'' := (Q - \tilde{D}_K \Sigma K \tilde{D}_K^T)^{-1} \tilde{D}_K Q > 0.
\]

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

\[
Q'' > 0 \iff Q'' = \begin{bmatrix} Q & \tilde{D}_K Q \tilde{D}_K \Sigma K \tilde{D}_K^T \end{bmatrix} \begin{bmatrix} Q & \tilde{D}_K Q \tilde{D}_K \Sigma K \tilde{D}_K^T \end{bmatrix} > 0,
\]

which is trivially true because \(Q^{-1}\) is positive definite since \(Q > 0\). Hence, \(\tilde{D}_K \Sigma K \tilde{D}_K^T\) and \(\Sigma^{\tilde{s},s}_K/(\tilde{D}_K \Sigma K \tilde{D}_K^T)\) are both positive definite, and thus \(\Sigma^{\tilde{s},s}_K > 0\).

To obtain the densities of \(\tilde{y}_K\) and \(s_K\) and their cross-covariance \(\Sigma^{\tilde{s},s}_K\), we just marginalize (see [24]) their joint density in Lemma 1 over \(s_K\) and \(\tilde{y}_K\), see the following corollary.

**Corollary 1** \(\tilde{y}_K \sim N[\mu^{\tilde{s}}_K, \Sigma^{\tilde{s}}_K]\) and \(s_K \sim N[\mu^K, \Sigma^K]\):

- \(\mu^{\tilde{s}}_K = \tilde{C}_K F_K \mu^K + \tilde{C}_K N_K u^K_{-1}\),
- \(\mu^K = \tilde{D}_K F_K \mu^K + \tilde{D}_K N_K u^K_{-1}\),
- \(\Sigma^{\tilde{s}}_K = (I_n)\Sigma^{\tilde{s},s}_K(I_n) = \mathbb{R}^{Kn_y \times Kn_y}\),
- \(\Sigma^K = (I_n)\Sigma^{\tilde{s},s}_K(I_n) = \mathbb{R}^{Kn_y \times Kn_n}\),
- \(\Sigma^{\tilde{s},s}_K = (I_n)\Sigma^{\tilde{s},s}_K(I_n) = \mathbb{R}^{Kn_y \times Kn_n}\).

At this point, we have the three covariance matrices \((\Sigma^{\tilde{s},s}_K, \Sigma^{\tilde{s},s}_K, \Sigma^{\tilde{s},s}_K)\) that we need to compute the mutual information \(I[s_K; \tilde{y}_K]\). Note that \(\Sigma^{\tilde{s},s}_K\) and \(\Sigma^{\tilde{s},s}_K\) are independent of the design variables \(\Sigma^K\) and \(\Sigma^{\tilde{s},s}_K\), and \(\Sigma^{\tilde{s},s}_K\) depends on \(\Sigma^K\):

\[
\Sigma^{\tilde{s},s}_K = \begin{bmatrix} \Sigma^{\tilde{s}}_K \\ \Sigma^{\tilde{s}}_K^T \\ \Sigma^{\tilde{s}}_K \\ \Sigma^{\tilde{s}}_K^T \\ \Sigma^{\tilde{s}}_K \\ \Sigma^{\tilde{s}}_K^T \\ \Sigma^{\tilde{s}}_K \\ \Sigma^{\tilde{s}}_K^T \\ \Sigma^{\tilde{s}}_K \end{bmatrix},
\]

where \(\Sigma^{\tilde{s}}_K\), \(\Sigma^{\tilde{s}}_K\), \(\Sigma^{\tilde{s}}_K\), and \(\Sigma^{\tilde{s}}_K\) in (31)–(34), we can write \(I[s_K; \tilde{y}_K]\) as follows

\[
I[s_K; \tilde{y}_K] = h[s_K] + h[\tilde{y}_K] - h[\tilde{y}_K, s_K]
\]

where (35b) follows from standard determinant and logarithm formulas. In the following lemma, we prove that minimizing the cost function \(I[s_K; \tilde{y}_K] - h[\tilde{y}_K] using \(\Sigma^{\tilde{s},s}_K\) as optimization variables is equivalent to solving a convex program subject to Linear Matrix Inequalities (LMI) constraints.

**Lemma 2** Minimizing \(I[s_K; \tilde{y}_K] - h[\tilde{y}_K]\) is equivalent to solving the following convex program:

\[
\begin{bmatrix} \min_{\Sigma_K, \Pi_K, \Sigma_K} \end{bmatrix} \begin{bmatrix} - \log \det (\Pi_K) - \log \det (\Sigma_K) \\ \Sigma_K - \Pi_K \\ \Sigma_K^\top \\ \Sigma_K^\top \end{bmatrix} \geq 0, \Pi_K > 0.
\]

**Proof:** (a) For any positive definite matrix \(\Sigma\), the function \(-\log \det (\Sigma)\) is convex in \(\Sigma\) (28). It follows that \(-h[\tilde{y}_K]\) in (19) is a convex function of \(\Sigma^{\tilde{s}}_K\). Minimizing \(-h[\tilde{y}_K]\) amounts to minimizing \(-\log \det (\Sigma^{\tilde{s}}_K)\) as all other terms in (19) are constants.

(b) Next, consider the expression for \(I[s_K; \tilde{y}_K]\) in (35b). Due to the monotonicity of the determinant function and the fact that \(\Sigma_K\) is independent of the design variables, minimizing (35b) is equivalent to

\[
\begin{bmatrix} \min_{\Pi_K, \Sigma_K} \end{bmatrix} \begin{bmatrix} - \log \det (\Pi_K) \\ \Pi_K - \Sigma_K \\ \Sigma_K^\top \\ \Sigma_K^\top \end{bmatrix} \geq 0, \Pi_K > 0.
\]

Since \(\Sigma_K\) is constant in the design variables (see (24), and \(\Sigma_K\) is convex in \(\Sigma_K\), we can conclude that minimizing \(I[s_K; \tilde{y}_K] - h[\tilde{y}_K]\) is equivalent to solving the convex program in (36).

By Lemma 2 minimizing the cost in (18) is equivalent to solving the convex program in (36). Then, if the distortion
constraint in (18) is convex in the design variables, we can find optimal distorting mechanisms efficiently using off-the-shelf optimization algorithms.

B. Detection Performance Constraints: Formulation and Convexity

We consider a constraint on the false alarm rate in each time step, see (15), as a metric of performance degradation of the detector. This constraint is fully characterized by the CDF of the distorted distance measure \( \hat{z}_k \). Unfortunately, no closed-form expression is available for this CDF. To address this, we seek a lower bound on the CDF of \( \hat{z}_k \) to use it to enforce constraint (17). Define a standard Gaussian vector \( m_k \sim \mathcal{N}(0, I_{n_k}) \). Since \( \tilde{r}_k \sim \mathcal{N}(0, \Sigma_k) \), it is easy to verify that \( \tilde{r}_k \) can be written in terms of \( m_k \) as follows:

\[
\tilde{r}_k = \left( \Sigma_k \right)^{1/2} m_k,
\]

where \( \left( \Sigma_k \right)^{1/2} \) is the square root of \( \Sigma_k \). Hence, \( \hat{z}_k \) can be written as:

\[
\hat{z}_k = \tilde{z}_k^\top \Sigma_k^{-1} \tilde{z}_k = m_k^\top \left( \Sigma_k^{1/2} \right)^\top \Sigma_k^{-1} \Sigma_k^{1/2} m_k.
\]

(39)

Define the sequence \( q_k := \beta m_k^\top m_k, \beta \in \mathbb{R}^+, k \in \mathcal{K} \), and let \( \Sigma_k' := \left( \left( \Sigma_k \right)^{1/2} \right)^\top \Sigma_k^{-1} \left( \Sigma_k \right)^{1/2} \); then, \( \hat{z}_k \leq q_k \) for all \( k \in \mathcal{K} \) implies \( m_k^\top \Sigma_k m_k \leq \beta m_k^\top m_k \). It follows that \( \hat{z}_k \leq q_k \) if and only if

\[
\Sigma_k' \leq \beta I_{n_q}.
\]

(41)

To be able to find the constraints of Problem 1 in terms of the CDF of the upper bound \( q_k \), we need to find a relation between the CDFs of \( \hat{z}_k \) and \( q_k \). In the next lemma, we show such a relation.

Lemma 3: \( \hat{z}_k \leq q_k, k \in \mathcal{K} \), implies:

\[
F_{\hat{z}_k}(\alpha) \geq F_{q_k}(\alpha),
\]

(42)

where \( F_{\hat{z}_k}(\alpha) \) and \( F_{q_k}(\alpha) \) denote the CDFs of \( \hat{z}_k \) and \( q_k \), respectively.

Proof: The proof is given in Appendix A.

From (41), Lemma 3, and (17), we can conclude that the constraints of the optimization problem (18) hold if \( \Sigma_k' \leq \beta I_{n_q} \) and

\[
F_{\beta m_k^\top m_k}(\alpha) > 1 - \alpha/2 - \epsilon \quad \text{for all} \quad k \in \mathcal{K}.
\]

(43)

Therefore, if we choose \( \beta \) to satisfy (43), the false alarm constraint in (18) translate to \( \Sigma_k' \leq \beta I_{n_q} \). Given a desired distortion level \( \epsilon \) and desired false alarm rate \( A^* \), a lower bound on \( F_{\beta m_k^\top m_k}(\alpha) \) can be computed. Since \( m_k^\top m_k \) follows a chi-squared distribution with \( n_q \) degree of freedom, the CDF \( F_{q_k}(\alpha) \) can be computed as follows (see [29]):

\[
F_{q_k}(\alpha) = P \left( n_q/2, \alpha/2\beta \right),
\]

(44)

where \( P(\cdot, \cdot) \) is the regularized gamma function [29]. Substituting (44) into (43) yields:

\[
P \left( n_q/2, \alpha/2\beta \right) > 1 - A^* - \epsilon.
\]

(45)

Because the regularized gamma function is an increasing function of \( \alpha/2\beta \) [30], the largest \( \beta \) satisfying (45) for given \( \alpha, \epsilon, \) and \( A^* \) can be calculated using the inverse of the lower incomplete gamma function [10] as follows:

\[
\beta < \beta^* := \frac{\alpha}{2} \left( n_q/2, 1 - A^* - \epsilon \right).
\]

(46)

Then, once we have \( \beta^* \), the false alarm constraint in Problem 1 amounts to \( \Sigma_k' \leq \beta I_{n_q} \), which is a characterization in terms of the design variables (as \( \Sigma_k' \) depends on \( \Sigma_k' \) and \( \Sigma_k' \)). We state this explicitly in the following lemma.

Lemma 4: Inequality \( \Sigma_k' = \left( \left( \Sigma_k \right)^{1/2} \right)^\top \Sigma_k^{-1} \left( \Sigma_k \right)^{1/2} < \beta^* I_{n_q} \) can be written in terms of \( \Sigma_k' \) and \( \Sigma_k' \) as:

\[
\hat{z}_k < \beta^* \Sigma_k.
\]

(47)

with \( \beta^* \) defined in (46) and \( \hat{z}_k = \Sigma + \Sigma_k' + CL \Sigma_k' L^\top C^\top + CBS_k' B^\top C^\top \).

Proof: The proof is given in Appendix B.

By Lemma 2 and Lemma 4, the cost \( I[\tilde{y}^K; \tilde{y}^K] - h[j^K] \) and false alarm constraint \( F_{\hat{z}_k}(\alpha) > 1 - A^* - \epsilon \) are written linearly in terms of the optimization variables \( \Sigma_k' \) and \( \Sigma_k' \). In what follows, we pose the complete nonlinear convex program to solve Problem 1.

Theorem 1: Consider the system dynamics (1), distorting mechanism (10), \( k \in \mathcal{K} \) with \( K = \{1, \ldots, K\} \) and \( K \in \mathbb{N} \), desired false alarm rate \( A^* \), maximum distortion level \( \epsilon \in \mathbb{R}^+ \), the chi-squared procedure threshold \( \alpha \), the anomaly detector Kalman filter (12), and covariance matrices \( \Sigma_k', \Sigma_k' \), and \( \Sigma_k' \) (given in (31)–(34)). Then, the covariance matrices \( \Sigma_k' > 0 \) and \( \Sigma_k' > 0 \) that minimize \( I[\tilde{y}^K; \tilde{y}^K] - h[j^K] \) subject to \( F_{\hat{z}_k}(\alpha) > 1 - A^* - \epsilon \) can be found by solving the convex program in (48).

Proof: The expressions for the cost and constraints and their convexity follow from Lemma 2 and Lemma 4.

Solving the convex program in (48) provides an upper bound on the solution for the original optimization problem in (18), considering the relaxation of constraints. This approach allows us to navigate the challenges of solving the

\[
\begin{aligned}
\min_{\Pi_K, \Sigma_k', \Sigma_k'} & - \log \det[\Sigma_k'] - \log \det[\Pi_K], \\
\text{s.t.} & \begin{cases}
\Pi_K > 0, \\
\Sigma_k' - \Pi_K \begin{bmatrix} \Sigma_k' \Sigma_k' \end{bmatrix} \geq 0, \\
\Sigma_k' > 0, \\
\Sigma + \Sigma_k' + CL \Sigma_k' L^\top C^\top + CBS_k' B^\top C^\top < \beta^* \Sigma_k', \\
\beta^* > 2P^{-1}(\frac{n_q/2}{2}, 1 - A^* - \epsilon).
\end{cases}
\end{aligned}
\]

(48)
Remark 1 In the following proposition, we prove that we can employ Theorem 1 to find optimal privacy-preserving mechanisms for prior distributions that are not necessarily Gaussian, as long as they are log-concave [14]. In particular, we prove that if \( s^K \) and \( \tilde{y}^K \) follow non-Gaussian log-concave distributions, their mutual information is upper-bounded by an affine function of the mutual information between Gaussian random vectors with the same mean and covariance matrices as \( s^K \) and \( \tilde{y}^K \), respectively. Examples of log-concave distributions are normal, exponential, uniform, Laplace, chi, beta, and logistic distributions. We refer the reader to [31], [32] for a comprehensive list and properties of log-concave distributions.

**Proposition 1** Let \( s^K \) and \( \tilde{y}^K \) be jointly distributed following log-concave multivariate probability distributions. The mutual information \( I[s^K; \tilde{y}^K] \) can be upper-bounded as follows:

\[
I[s^K; \tilde{y}^K] \leq I[s^K_N; \tilde{y}^K_N] + C_n, \tag{49}
\]

where \( C_n = \frac{1}{2} \log(2\pi e c(n))^n \), \( n = K(n_a + n_y) \), \( c(n) = \frac{c^2 n^2}{4 \sqrt{2(n+2)}} \), and \( s^K_N \) and \( \tilde{y}^K_N \) denote Gaussian random vectors with the same covariance matrices as \( s^K \) and \( \tilde{y}^K \), respectively.

**Proof:** The proof is given in Appendix C.

Remark 2 Proposition 1 implies that by minimizing \( I[s^K_N; \tilde{y}^K_N] - h[j^K] \) (based on the optimization problem in Theorem 1), we can effectively decrease the information leakage \( I[s^K; \tilde{y}^K] - h[j^K] \) for any \((s^K, \tilde{y}^K)\) following log-concave distributions.

A tighter bound for the mutual information of log-concave random vectors can be achieved based on their maximum density. The maximum density of a random vector is calculated by its \( L_\infty \) norm [33]. In the following proposition, we prove that the mutual information between log-concave random vectors \( s^K \) and \( \tilde{y}^K \) can be upper-bounded by an affine function of the mutual information of Gaussian random vectors with maximum density being the same as those of \( s^K \) and \( \tilde{y}^K \).

**Proposition 2** Let \( s^K \) and \( \tilde{y}^K \) be jointly distributed following a log-concave multivariate joint probability distribution with mutual information \( I[s^K_N; \tilde{y}^K_N] \), then:

\[
I[s^K; \tilde{y}^K] \leq I[s^K_N; \tilde{y}^K_N] + n, \tag{50}
\]

where \( n = K(n_a + n_y) \), and \( s^K_N \) and \( \tilde{y}^K_N \) denote multivariate normally distributed random vectors with maximum density being the same as those of \( s^K \) and \( \tilde{y}^K \), respectively.

**Proof:** The proof is given in Appendix D.

Remark 3 In Proposition 2 we prove that the difference between \( I[s^K_N; \tilde{y}^K_N] \) and \( I[s^K_N; \tilde{y}^K_N] \) can be upper-bounded by \( n \). This upper bound is still conservative since it needs to cover all log-concave distributed random vectors. However, the authors in [33] observe that every log-concave random vector has approximately the same entropy per coordinate as a related Gaussian vector (which resulted in approximately the same mutual information), and log-concave distributions resemble Gaussian distributions. Therefore, by minimizing \( I[s^K_N; \tilde{y}^K_N] \) (based on the optimization problem in Theorem 1), we can effectively decrease the information leakage \( I[s^K; \tilde{y}^K] - h[j^K] \) for any \((s^K, \tilde{y}^K)\) following a log-concave distribution.

Remark 4 In this manuscript, we have addressed the problem of designing optimal probabilistic mappings to maximize privacy for a finite time horizon. In [15], it has been proven that the proposed mechanisms can be implemented for an arbitrarily large time horizon by solving finite horizon problems repeatedly and sequentially implementing the obtained mechanisms.

Remark 5 In this work, we consider the problem of solving an optimization problem to hold a trade-off between privacy and the performance of the anomaly detection algorithm in order to design optimal privacy-distorting mechanisms. For designing the privacy mechanisms in the general case, i.e., for utilities of data other than anomaly detection, another optimization problem to hold a trade-off between privacy and the amount of distortion is described in [12], [13]. This generalizes the applicability of the results in the paper to a much wider range of applications (beyond anomaly detection) in which a balance between privacy and performance (challenged by the distortion of data) is needed.

V. ON ANOMALY DETECTION PERFORMANCE METRICS

In order to assess the cost of the proposed privacy-preserving mechanisms in terms of detectability loss, three different metrics are employed.

The first metric for evaluating the performance of the anomaly detector is the false alarm rate. In the absence of privacy mechanisms, the false alarm rate is \( A^* \). After applying the privacy mechanisms, based on distortion constraint \( \Pr[\tilde{z}_k > \alpha] \approx A^* + \epsilon \) in optimization problem [18], the false alarm rate \( \tilde{A} = \Pr[\tilde{z}_k > \alpha] \) can be increased by \( \epsilon \). Therefore, the privacy distorting mechanism degrades the performance of the anomaly detector in terms of the false alarm rate. We calculate the false alarm rate based on the CDF of \( \tilde{z}_k \) as follows:

\[
\tilde{A} = \Pr[\tilde{z}_k > \alpha | \delta_k = 0] = 1 - F_{\tilde{z}_k}(\alpha | \delta_k = 0). \tag{51}
\]

Since \( \tilde{z}_k \) approximately follows a gamma distribution (see Appendix E), its CDF can be determined by the inverse of the lower incomplete gamma function [29].

Another metric is the detection rate of the anomaly detector for deterministic additive fault \( \delta_k \). The detection rate to detect faulty systems with distance measure \( z_k \) is defined as \( B = \Pr[z_k > \alpha | \delta_k \neq 0] \). We start with the case when there is no privacy distortion in the system but \( \delta_k \neq 0 \). In this case, the faulty residual sequence is determined by [13]. According to
the superposition principle, the faulty residual sequence can be written as:

$$r_k = r^\delta_k + r^\pi_k,$$  \hfill (52)

where $r^\delta_k$ is a signal representing the change in the residual due to the fault $\delta_k$, and $r^\pi_k$ represents the change in the residual due to system and measurement noises. In \cite{36}, we proved that if there is no anomaly and no privacy distortion in the system, the residual follows Gaussian distribution $r_k \sim \mathcal{N}(0, \Sigma)$. Besides, since $\delta_k$ is a deterministic additive fault, $r^\delta_k$ will only appear in the mean of the residual. Therefore, the residual follows Gaussian distribution $r_k \sim \mathcal{N}(r^\delta_k, \Sigma)$ when $\delta_k \neq 0$. It can be proved that based on this new residual, the new distance measure is distributed as a non-central chi-squared with $n_y$ degrees of freedom and non-centrality parameter $\lambda_k = \|\Sigma^{-1/2}r^\delta_k\|^2$ \cite{34}. Note that for a fixed $\Sigma$, $\lambda_k$ is an increasing function of the fault. Let $B^*$ be the detection rate without privacy distortion. This detection rate can be calculated based on the CDF of $z_k$ as

$$B^* = 1 - F_{z_k}(\alpha|\delta_k \neq 0).$$  \hfill (53)

Since $z_k$ is a non-central chi-squared variable for $\delta_k \neq 0$, and due to the fact that the CDF of a non-central chi-squared variable is a decreasing function of its non-centrality parameter (see \cite{55}), the CDF of $z_k$, $F_{z_k}(\alpha|\delta_k \neq 0)$, is a decreasing function of $\lambda_k$. As a result, by increasing $\lambda_k$, $B^* = 1 - F_{z_k}(\alpha|\delta_k \neq 0)$ increases. Therefore, given the fact that $B^*$ is an increasing function of $\lambda_k$ and $\lambda_k$ is an increasing function of $\lambda_k$, we can conclude that for larger fault $\delta_k$, detection rate $B^*$ will be increased, which amounts to better detectability.

For the case when there is both privacy distortion and fault in the system, based on the superposition principle, we need to consider another term in the residual for the additive noises for privacy compared with \cite{52}. Therefore, the distorted faulty residual is given by:

$$\tilde{r}_k = r^\delta_k + r^\pi_k + r^p_k,$$  \hfill (54)

where $r^p_k$ is a signal representing the change in the residual due to privacy distortions. In Equation \cite{14}, we showed that the distorted residual follows $\tilde{r}_k \sim \mathcal{N}(0, \Sigma_k)$. Consequently, the distorted faulty residual follows Gaussian distribution as $\tilde{r}_k \sim \mathcal{N}(r^\delta_k, \Sigma_k)$. It can be proved that based on the new residual, the new distance measure $\tilde{z}_k = \tilde{r}_k^\top \Sigma_k^{-1} \tilde{r}_k$ is distributed as a generalized chi-squared \cite{56}. Then, the detection rate $B$ in the presence of privacy mechanisms can be calculated as follows:

$$B = 1 - F_{z_k}(\alpha|\delta_k \neq 0).$$  \hfill (55)

Since $\tilde{z}_k$ follows generalized chi-squared, its PDF and CDF, unfortunately, cannot be expressed in closed form, although approximations are available \cite{56}.

To determine the effect of privacy distortion on the detection rate, the function of the detection rate in terms of $\delta_k$ in with and without privacy distortion cases can be compared. At the start point, when $\delta_k = 0$, the detection rate equals the false alarm rate, and therefore, the detection rate is higher when privacy distortions are added. Besides, as we mentioned before, when there is no privacy distortion, the detection rate is monotonically increasing for an increasing amount of $\delta_k$. In addition, when $\delta_k$ goes to infinity, the detection rate goes to 1 for both with and without privacy distortions cases. Hence, by adding privacy distortions, the function of detection rate in terms of fault starts from a higher point (compared to without privacy distortions case) and reaches the same point for infinitely large fault where $B = 1$. Hence, the privacy distortion flattens the detection rate function in terms of fault magnitude (see Fig. \cite{6} for our implemented numerical example of this occurrence). The detection rate could not be employed as a performance metric in the privacy-utility trade-off optimization problem. Because to be able to use it, we need to know the distribution of the fault, which is not a realistic assumption. Moreover, even when the fault is deterministic, the detection rate in the presence of privacy mechanisms \cite{55} needs to be calculated numerically. Therefore, this metric cannot effectively be employed to define the detection performance constraint as we do with the false alarm rate.

To investigate the overall effect of privacy mechanisms on detection performances, the Receiver Operating Characteristic (ROC) is considered. The ROC curves show how the detection rate changes as the false-alarm rate increases. In \cite{51} and \cite{55}, the false alarm rate and the detection rate are written in terms of $\alpha$, where $\alpha$ is selected to satisfy the desired false alarm rate $A^*$. Therefore, calculating the false alarm rate and detection rate for different $A^*$ will allow characterizing the ROC of the proposed detection scheme. The ROC plot can be employed to explore the net effect of privacy mechanisms on detection performances in terms of false alarm rate and detection rate.

VI. ILLUSTRATIVE CASE STUDY

The authors of \cite{19, 37} studied the anomaly detection problem for a well-stirred chemical reactor with a heat exchanger. We use this case study to demonstrate our results.

The state, input, and output vectors of the considered reactor are:

$$\begin{bmatrix} x_k \end{bmatrix} = \begin{bmatrix} C_0 \\ T_0 \\ T_w \\ T_m \end{bmatrix}, \quad u_k = C_u, \quad y_k = \begin{bmatrix} C_0 \\ T_0 \end{bmatrix},$$

where

$$\begin{align*}
C_0 & : \text{Concentration of the chemical product}, \\
T_0 & : \text{Temperature of the product}, \\
T_w & : \text{Temperature of the jacket water of the heat exchanger}, \\
T_m & : \text{Coolant temperature}, \\
C_u & : \text{Inlet concentration of the reactant}.
\end{align*}$$

We use the discrete-time dynamics of the reactor introduced in \cite{11} for this illustrative simulation study with matrices as given in \cite{56}, normally distributed initial condition $x_1 \sim \mathcal{N}([6.94; 13.76; 1; 1]\top, I_4]$, and reference signal $u_k = 50 \cos[0.5k]^2$. As private output, we use the concentration of the chemical product; then, matrix $D$ in \cite{11} is given by the full row rank matrix $D = (1, 0, 0, 0)$. Therefore, the aim of
First, in Figure 2, we show the effect of false alarm distortion level $\epsilon$ on the amount of optimal information leakage, which is shown by the optimal cost $I[s^K; \tilde{y}^K] - h[j^K]$. This figure depicts the evolution of the optimal cost for increasing $\epsilon$ with time horizon $K = 60$ and desired false alarm rate $\mathcal{A}^* = 0.1$ in the absence of fault. As expected, by increasing the distortion level $\epsilon$, which increases the amount of distortion on the false alarm rate as we discussed in Section VI (i.e., leading to poorer anomaly detection performance), the information leakage decreases monotonically (i.e., leading to improved privacy). This figure illustrates that while designing the privacy distorting mechanisms, we need to balance the trade-off between privacy and anomaly detector performance.

The effect of optimal distorting mechanisms is shown in Figure 3 where we contrast actual and distorted measurement data for different distortion levels $\epsilon \in \{0.1, 0.75, 0.9\}$. We consider $\mathcal{A}^* = 0.1$; therefore, $\epsilon = 0.9$ means that the optimization problem in (48) is solved without considering the distortion constraint (because in this case, the upper bound of the false alarm rate is $\mathcal{A}^* + \epsilon = 1$ which means that there is no constraint on the false alarm rate). As can be seen in these figures, the difference between distorted and actual measurements will increase by increasing the distortion level. The same result is achieved for the comparison of actual and distorted inputs.

Next, in Figure 4, we depict the realization of $s_k$, the estimated private output without distorting $y_k$ and $u_k$ using the MMSE estimator, $\hat{s}_k^{nu}$, and the estimated private output given the distorted vectors $\tilde{y}_k$ and $\tilde{u}_k$, $\hat{s}_k^{nu}$. We consider time horizon $K = 60$, distortion levels $\epsilon \in \{0.1, 0.75, 0.9\}$, and desired false alarm rate $\mathcal{A}^* = 0.1$ in the absence of fault. As can be seen in these figures, by increasing the distortion level $\epsilon$, the accuracy of estimation decreases. This figure demonstrates that by randomizing measurement and input data, we can prevent worst-case adversaries from accurately estimating $s_k$.

To evaluate the effect of privacy mechanisms on anomaly detection performance, we apply three different metrics, as discussed in Section VI. First, in Figure 5 the effect of distortion level $\epsilon$ on the false alarm rate is shown. As discussed in Section VI, we employ the CDF of the approximation of $\tilde{x}_k$, $F_{\tilde{x}_k}(\alpha)$, using the lower incomplete gamma function to determine the false alarm rate. As can be seen, by increasing $\epsilon$, the false alarm rate increases. Therefore as we expected, increasing false alarm distortion level $\epsilon$, which resulted in increasing privacy distortion and decreasing the information leakage, indeed reduces the performance of the anomaly detector in terms of false alarm rate. Besides, this figure shows...
that the upper bound we designed for the false alarm rate in distortion constraint (17), $A^* + \epsilon$, is respected and not so conservative.

Next, as the second detection metric, the effect of privacy distortion on the detection rate of the chi-squared tuning procedure is depicted in Figure 6. We consider deterministic additive constant fault $\delta_k = \delta$ with fault matrices $H = (0, 1)^T$ and $G = 0$. We apply privacy distortions with different distortion levels $\epsilon \in \{0.01, 0.3, 0.5\}$ and desired false alarm rate $A^* = 0.1$, where $\epsilon = 0$ indicates without privacy distortion.

As can be seen in this figure, the detection rate of the anomaly detection algorithm $B$, as defined in Section VI, is changing shape and flattening by increasing the distortion level $\epsilon$. For small $\delta$, the detection rate increases by increasing $\epsilon$, and for large $\delta$, the detection rate slightly decreases by increasing $\epsilon$. It might be mistakenly concluded from this figure that increasing the distortion level can improve the detection performance of the anomaly detector. On the other hand, Figure 5 illustrates that by increasing the distortion level, the false alarm rate will increase. Therefore, to investigate the overall effect of privacy mechanisms on detection performances, the ROC curve is considered for analyzing the anomaly detector performance.

The ROC shows how the detection rate changes as the false alarm rate increases. To characterize the ROC of the anomaly detection scheme, we find the detection rate (true positive) for a given fault $\delta$ as a function of the false alarm rate (false positive). Generally, in ROC analysis, tests are categorized based on the area under the ROC curve. The closer a ROC curve is to the upper left corner, the more efficient the test is. At first, in Figure 7, we plot the ROC when there is no privacy distortion in the system for different amounts of fault $\delta \in \{0.1, 1, 2, 3, 4\}$. As we expected based on (53), by increasing the fault, the detection rate will be increased, and the ROC curve will be closer to the upper left corner, which illustrates the better performance of the anomaly detector. Then in Figure 8 the ROC is depicted for the distorted faulty system with a fixed fault $\delta = 2$ and different distortion levels $\epsilon \in \{0.01, 0.3, 0.5\}$ by increasing the desired false alarm rate.

As can be seen, for the fixed false alarm rate, the detection rate is decreased by increasing $\epsilon$. Also, for the fixed detection rate, the false alarm rate is increased by increasing $\epsilon$. In other words, by increasing $\epsilon$, the ROC curve will get farther from the upper left corner, which illustrates the worse performance of the anomaly detector. To further explore this figure, we marked with circle markers on all plots corresponding to a specific value of the desired false alarm rate $A^* = 0.3$.

It can be seen that by increasing $\epsilon$, the $A^* = 0.3$ point moves toward the top right corner of the ROC, close to point $(1, 1)$ which shows a severely poor detection performance, equivalent to that of a detector that always identifies a residual as faulty, regardless of its real value. Therefore, this figure shows that increasing the privacy distortion will reduce the anomaly detector’s performance.

In Figure 9, we show that, as discussed in Remark 2 and for non-Gaussian random vectors $s^K$ and $y^K$ following joint log-concave distributions, their mutual information can be upper-bounded by affine functions of the mutual information of some Gaussian random vectors with the same mean and covariance matrices. We consider $s^K$ and $y^K$ with joint Laplace distribution and solve the optimization problem in (38) for $I\{s^K; \tilde{y}^K\} - h[\tilde{y}^K]$ to find the optimal distorting parameters for an increasing amount of distortion level $\epsilon$. Since the maximum density (33) of Gaussian random vectors and Laplace random vectors are the same for the same means and covariances, we can conclude that for Laplace variables $s^K = s^K_N, \tilde{y}^K = y^K_N$, and therefore $I\{s^K_N; \tilde{y}^K_N\} - h[\tilde{y}^K_N] = I\{s^K_N; y^K_N\} - h[y^K_N]$. We calculate the optimal $I\{s^K; \tilde{y}^K\} = I\{s^K_N; \tilde{y}^K_N\}$ by solving the optimization problem (38). Then, we calculate $I\{s^K; \tilde{y}^K\}$ (see (38) for the mutual information calculation), where $s^K_N$ and $y^K_N$ follow Laplace distributions with the same mean and covariance as $s^K_N$ and $\tilde{y}^K_N$, and $y^K_N = y^K + z^K$ (with the same distortion variables as the Gaussian case). As can be seen in Figure 9, $I\{s^K; \tilde{y}^K\}$ is even upper-bounded by $I\{s^K_N; \tilde{y}^K_N\}$ without the constant terms $n$ and $C_n$. It implies that by solving the optimization problem for Gaussian variables, the information leakage for Laplace variables decreases for increasing $\epsilon$ and is upper-bounded by the information leakage for Gaussian variables. Furthermore, it is apparent that even if the Gaussian upper bounds are loose for these distributions of $s^K$ and $\tilde{y}^K$, the information leakage is still decreasing for increasing $\epsilon$. It shows that minimizing $I\{s^K_N; \tilde{y}^K_N\} - h[\tilde{y}^K_N]$ or $I\{s^K_N; \tilde{y}^K_N\} - h[y^K_N]$ effectively decreases the information leakage for any log-concave $s^K$ and $\tilde{y}^K$.

VII. CONCLUSIONS

This paper presented an anomaly detection framework to maximize privacy for a class of stochastic dynamical systems. A detailed mathematical framework for synthesizing distorting mechanisms is given to minimize the information leakage.
induced by the use of public/unsecured communication networks. We proposed a class of linear Gaussian distorting mechanisms to randomize sensor measurements and input signals before transmission to prevent adversaries from accurately estimating the private part of the system state (a private performance output). Furthermore, for the class of systems under study, we thoroughly characterized information-theoretic metrics (mutual information and differential entropy) to quantify the information between private outputs and disclosed data for a class of worst-case eavesdropping adversaries.

Anomaly detection is an important problem that has been studied in a variety of applications. We proved that the privacy distorting mechanisms can increase the false alarm rate in the anomaly detector, resulting in reducing the performance of the anomaly detection scheme. Therefore, when designing the privacy distorting mechanism, we considered the tradeoff between privacy and anomaly detection performance. We formulated the synthesis of the distorting mechanisms as a convex program, where we minimize the information leakage quantified using the mutual information between an arbitrarily large sequence of private outputs and the disclosed distorted data and the differential entropy of the sequence of input additive noise while considering a constraint on the amount of false alarm rate in the anomaly detector to limit the effect of privacy mechanisms on the anomaly detection algorithm. We presented simulation results to illustrate the performance of the proposed tools.

REFERENCES

[1] 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.

[2] 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.

[3] J. L. Ny and G. J. Pappas, “Differentially private filtering,” IEEE Transactions on Automatic Control, vol. 59, pp. 341–354, 2014.
8. Appendix

A. Proof of Lemma 3

\( \tilde{z}_k \) and \( q_k \) are defined as functions of variable \( m_k \). For the CDF of a function of a random variable, if \( Y \) and \( X \) are continuous random variables with \( Y = g(X) \), the CDF of \( Y \), \( F_Y(y) \) can be characterized based on Probability Density Function (PDF) of \( X \) as follows

\[
F_Y(y) = P_Y(Y \leq y) = P_Y(g(X) \leq y)
\]

\[
= P_X(x \in \mathcal{X} : g(x) \leq y) = \int_{x \in \mathcal{X} : g(x) \leq y} f_X(x) dx,
\]

where \( f_X(x) \) and \( \mathcal{X} \) are PDF and the sample space of \( X \). Therefore, CDF of \( \tilde{z}_k \) and \( q_k \) can be written as functions of the PDF of \( m_k \) as below:

\[
F_{\tilde{z}_k}(\alpha) = \int_{m_k : \tilde{z}_k m_k \leq \alpha} f_{m_k}(\eta) d\eta,
\]

\[
F_{q_k}(\alpha) = \int_{m_k : q_k m_k \leq \alpha} f_{m_k}(\eta) d\eta.
\]

From \( m_k^\top \Sigma m_k \leq \beta m_k^\top m_k \), we can conclude that

\[
\{ m_k : \beta_k m_k^\top m_k \leq \alpha \} \subseteq \{ m_k : m_k^\top \Sigma m_k \leq \alpha \},
\]

(60) which means that the ellipsoid \( \{ m_k^\top \Sigma m_k = \alpha \} \) is a subset of ellipsoid \( \{ m_k^\top m_k = \alpha \} \) (see [39] for proof and details).

From (60) and the fact that \( f_{m_k}(\eta) \) is always non-negative, we deduce that

\[
\int_{m_k m_k^\top \Sigma m_k \leq \alpha} f_{m_k}(\eta) d\eta \geq \int_{m_k : m_k^\top m_k \leq \alpha} f_{m_k}(\eta) d\eta.
\]
According to (58) and (59), (61) is equivalent to \( F_{z_k}(\alpha) \geq F_{q_k}(\alpha) \).

**B. Proof of Lemma 4**

By multiplying \( \tilde{\Sigma}_k^{-1/2} \) to the right and \( (\tilde{\Sigma}_k^{-1/2})^\top \) to the left side of \( (\tilde{\Sigma}_k^{-1/2})^{-1} - \Sigma^{-1} \Sigma_k \leq \beta^* I_{n_y} \), it will be converted to

\[
\Sigma^{-1} \leq \beta^* (\tilde{\Sigma}_k^{-1/2})^{-1/2} = \beta^* \tilde{\Sigma}_k^{-1}.
\]

(62)

Since \( \Sigma \) and \( \tilde{\Sigma}_k \) are positive definite matrices, (62) can be written as (47) which is a convex function of design variables \( \Sigma_k \) and \( \tilde{\Sigma}_k \).

**C. Proof of Proposition 1**

First, we need to prove that for initial state \( x_1 \) and disturbances \( t_k \) and \( y_k \) following non-Gaussian log-concave distributions, and \( j^K \) and \( v^K \) following Gaussian distributions, \( s^K \) and \( \tilde{y}^K \) follow log-concave distributions. The log-concavity of distribution is preserved under affine transformations (see [40], Lemma 2.1). Besides, based on Hoggar’s theorem, the sum of two independent log-concave random variables is itself log-concave [41], [42]. Therefore, based on the system dynamics (1), the distortion mechanism (10), the indeterminacy between initial state and disturbances, and also indeterminacy between \( v^K, j^K \), and \( y^K \), we can conclude that \( y^K, s^K \), and \( \tilde{y}^K \) follow log-concave distributions.

Mutual information \( I[s^K; \tilde{y}^K] \) can be written in terms of differential entropies as \( I[s^K; \tilde{y}^K] = h[s^K] + h[\tilde{y}^K] - h[s^K, \tilde{y}^K] \). Moreover, because Gaussian distributions maximize entropy for a given covariance matrix [6], we can write:

\[
\begin{cases}
    h[s^K] \leq h[s_N^K], \\
    h[\tilde{y}^K] \leq h[\tilde{y}_{N'}^K].
\end{cases}
\]

(63)

The authors in [43] prove that the entropy of random vectors following log-concave distributions is lower bounded as:

\[
h[s^K, \tilde{y}^K] \geq n \frac{2}{\log} \frac{|\Sigma_{\tilde{y}^K}^n|^{1/n}}{c(n)},
\]

(64)

where \( n = K(n_s + n_y) \) and \( c(n) = \frac{\pi^{n_y}}{n^{n_y} (2n+3)^{n_y}} \). Expanding the lower bound in (64), we obtain:

\[
\frac{n}{2} \log \frac{|\Sigma_{\tilde{y}^K}^n|^{1/n}}{c(n)} = \frac{n}{2} \left( \frac{n}{2} \log \det \Sigma_{\tilde{y}^K}^n - \log c(n) \right)
\]

\[
= \frac{n}{2} \log \det \Sigma_{\tilde{y}^K}^n - \frac{n}{2} \log c(n).
\]

(65)

Let \( (s_N^K, \tilde{y}_{N'}^K) \) be jointly Gaussian with covariance \( \Sigma_{\tilde{y}^K}^n \). Then, we can write its differential entropy as \( h[s_N^K, \tilde{y}_{N'}^K] = \frac{n}{2} \log \det \Sigma_{\tilde{y}^K}^n + \frac{n}{2} \log(2\pi) \). Therefore, using the latter formula for \( h[s_N^K, \tilde{y}_{N'}^K] \) and (65), we can write the following:

\[
\frac{n}{2} \log \frac{|\Sigma_{\tilde{y}^K}^n|^{1/n}}{c(n)} = h[s_N^K] + h[\tilde{y}_{N'}^K] - \frac{n}{2} \log(2\pi) - \frac{n}{2} \log c(n).
\]

(66)

Combining (63), (64), and (66), we can finally conclude that

\[
I[s^K; \tilde{y}^K] \leq h[s_N^K] + h[\tilde{y}_{N'}^K] - h[s_N^K, \tilde{y}_{N'}^K] + \frac{n}{2} \log(2\pi) - \frac{n}{2} \log c(n),
\]

(67)

which is equivalent to (49).

**D. Proof of Proposition 2**

Mutual information \( I[s^K; \tilde{y}^K] \) can be written in terms of differential entropies as \( I[s^K; \tilde{y}^K] = h[s^K] + h[\tilde{y}^K] - h[s^K, \tilde{y}^K] \). Besides, it has been proved in Appendix C that \( y^K, s^K \), and \( \tilde{y}^K \) follow log-concave distributions. Then, from the inequality on the entropy of log-concave distributed random vectors, we can conclude that (see Theorem 1.1 in [33]):

\[
\begin{cases}
    \frac{1}{K_{n_y}} h[K_{n_y}] \leq \frac{1}{K_{n_s}} h[K_{n_s}], \\
    \frac{1}{K_{n_y}} h[y^K] \leq \frac{1}{K_{n_y}} h[\tilde{y}^K], \\
    \frac{1}{n} h[s^K, \tilde{y}^K] \geq \frac{1}{n} h[s_N^K, \tilde{y}_{N'}^K] - \frac{1}{2}.
\end{cases}
\]

(68)

Combining all the inequalities in (68), we have

\[
\frac{1}{n} I[s^K; \tilde{y}^K] \leq \frac{1}{n} (h[s_N^K] + h[\tilde{y}_{N'}^K] - h[s_N^K, \tilde{y}_{N'}^K]) + 1,
\]

(69)

which is equivalent to (50).

**E. Proof of \( \tilde{z}_k \) approximately following a gamma distribution**

At first, we prove that defining \( m_k \) as a standard normal variable \( m_k \sim \mathcal{N}(0, I_n) \) and matrix \( Q \in \mathbb{R}^{n \times n} \) with entries \( n \in \mathbb{R}^+ \), \( Q^i_k = m_k^i Q m_k \) approximately follows a gamma distribution. \( Q^i_k = m_k^i Q m_k \) is a weighted sum of chi-squared distributions with one degree of freedom (\( \chi^2_1 \)) as \( Q^i_k \sim \sum_{i=1}^n \lambda_i \chi_i \). where \( \lambda_i, i \in \{1, \ldots, n\} \), are the eigenvalues of the matrix \( Q \). For each term of the sum \( \lambda_i \chi_i^2 \) we have [29]:

\[
\lambda_i \chi_i^2 \sim \Gamma(K_i = \frac{1}{2}, \theta_i = 2\lambda_i).
\]

(70)

where \( \Gamma(K_i = \frac{1}{2}, \theta_i = 2\lambda_i) \) is a gamma variable with shape parameter \( K_i = \frac{1}{2} \) and scale parameter \( \theta_i = 2\lambda_i \). Therefore, \( Q^i_k \sim \sum_{i=1}^n \Gamma(K_i = \frac{1}{2}, \theta_i = 2\lambda_i) \) is the sum of \( n \) gamma variables. Based on the Welch-Satterthwaite approximation [44], [45], the summation of gamma distributions \( \sum_{i=1}^n \Gamma(K_i, \theta_i) \) approximately follows a gamma distribution with shape \( K_{\text{sum}} = \sum_{i=1}^n \theta_i K_i \) and scale \( \theta_{\text{sum}} = \sum_{i=1}^n \theta_i K_i \). Consequently, \( Q^i_k \) follows a gamma distribution \( Q^i_k \sim \Gamma(K^i, \theta^i) \) where \( K^i = \frac{\sum_{i=1}^n \lambda_i^2}{\sum_{i=1}^n \lambda_i^2} \) and scale \( \theta^i = \frac{2 \sum_{i=1}^n \lambda_i^2}{\sum_{i=1}^n \lambda_i^2} \). The accuracy of this approximation is proved in [34].

Hence, we can conclude that \( \tilde{z}_k = m_k^T \tilde{z}_k m_k \) approximately follows a gamma distribution \( \tilde{z}_k \sim \Gamma(K^i, \theta^i) \) with:

\[
K^i = \frac{\sum_{i=1}^n \lambda_i k^2}{2 \sum_{i=1}^n (\lambda_i k)},
\]

(71)

\[
\theta^i = \frac{2 \sum_{i=1}^n (\lambda_i k)^2}{(\sum_{i=1}^n \lambda_i k)^2},
\]

(72)
where $\lambda_{i,k}, i \in \{1, \ldots, n_y\}$ are the eigenvalues of the matrix $\Sigma'_k = (\tilde{\Sigma}_k^{1/2})^T \Sigma^{-1} \tilde{\Sigma}_k^{1/2}$.

It is shown in [34] that the mean of $\lambda_{i,k}$, which in our case are the mean of eigenvalues of $\Sigma'_k$, does not affect the accuracy of this approximation. Therefore, for added noises with relatively large covariances ($\Sigma'_k$ and $\Sigma'_l$), matrix $\Sigma'_k$ would be larger, which lead to larger eigenvalues of $\Sigma'_k$; but it does not affect the accuracy of this approximation. ■