On a Security vs Privacy Trade-off in Interconnected Dynamical Systems

Vaibhav Katewa, Rajasekhar Anguluri, and Fabio Pasqualetti

Department of Mechanical Engineering, University of California, Riverside, CA, USA

Abstract

We study a security problem for interconnected systems, where each subsystem aims to detect local attacks using local measurements and information exchanged with neighboring subsystems. The subsystems also wish to maintain the privacy of their states and, therefore, use privacy mechanisms that share limited or noisy information with other subsystems. We quantify the privacy level based on the estimation error of a subsystem's state and propose a novel framework to compare different mechanisms based on their privacy guarantees. We develop a local attack detection scheme without assuming the knowledge of the global dynamics, which uses local and shared information to detect attacks with provable guarantees. Additionally, we quantify a trade-off between security and privacy of the local subsystems. Interestingly, we show that, for some instances of the attack, the subsystems can achieve a better detection performance by being more private. We provide an explanation for this counter-intuitive behavior and illustrate our results through numerical examples.

Keywords: Privacy, Attack-detection, Interconnected Systems, Chi-squared test

1. Introduction

Dynamical systems are becoming increasingly more distributed, diverse, complex, and integrated with cyber components. Usually, these systems are composed of multiple subsystems, which are interconnected among each other via physical, cyber and other types of couplings [1]. An example of such system is the smart city, which consists of subsystems such as the power grid, the transportation network, the water distribution network, and others. Although these subsystems are interconnected, it is usually difficult to directly measure the couplings and dependencies between them [1]. As a result, they are often operated independently without the knowledge of the other subsystems' models and dynamics.

Modern dynamical systems are also increasingly more vulnerable to cyber/physical attacks that can degrade their performance or may even render them inoperable [2]. There have been many recent studies on analyzing the effect of different types of attacks on dynamical systems and possible remedial strategies (see [3] and the references therein). A key component of these strategies is detection of attacks using the measurements generated by the system. Due to the autonomous nature of the subsystems, each subsystem is primarily concerned with detection of local attacks which affect its operation directly. However, local attack detection capability of each subsystem is limited due to the absence of knowledge of the dynamics and couplings with external subsystems. One way to mutually improve the detection performance is to share information and measurements among the subsystems. However, these measurements may contain some confidential information about the subsystem and, typically, subsystem operators may be willing to share only limited information due to privacy concerns. In this paper, we propose a privacy mechanism that limits the shared information and characterize its privacy guarantees. Further, we develop a local attack detection strategy using the local measurements and the limited shared measurements from other subsystems. We also characterize the trade-off between the detection performance and the amount/quality of shared measurements, which reveals a counter-intuitive behavior of the involved chi-squared (χ^2) detection scheme.

Related Work: Centralized attack detection and estimation schemes in dynamical systems have been studied in both deterministic [4, 5, 6] and stochastic [7, 8] settings. Recently, there has also been studies on distributed attack detection including information exchange among the components of a dynamical system. Distributed strategies for attacks in power systems are presented in [9, 10, 11]. In [5, 12], centralized and decentralized monitor design was presented for deterministic attack detection and identification. In [13, 14], distributed strategies for joint attacks detection and state estimation are presented. Residual based tests [15] and unknown-input observer-based approaches [16] have also been proposed for attack detection. A comparison between centralized and decentralized attack detection schemes was presented in [17].The local detectors

⋆This material is based upon work supported in part by ARO award 71603NSYIP and in part by UCOP award LFR-18-548175.

Email address: {vkatewa,rangu003,fabiopas}@engr.ucr.edu (Vaibhav Katewa, Rajasekhar Anguluri, and Fabio Pasqualetti)

in [17] use only local measurements, whereas we allow the local detectors to use measurements from other subsystems as well.

Distributed fault detection techniques requiring information sharing among the subsystems have also been widely studied. In [18, 19, 20, 21, 22], fault detection for nonlinear interconnected systems is presented. These works typically use observers to estimate the state/output, compute the residuals and compare them with appropriate thresholds to detect faults. For linear systems, distributed fault detection is studied using consensus-based techniques in [23, 24] and unknown-input observer-based techniques in [25].

There have also been recent studies related to privacy in dynamical systems. Differential privacy based mechanisms in the context of consensus, filtering and distributed optimization have been proposed (see [26] and the references therein). These works develop additive noise-based privacy mechanisms, and characterize the trade-offs between the privacy level and the control performance. Other privacy measures based on information theoretic metrics like conditional entropy [27], mutual information [28, 29] and Fisher information [30] have also been proposed. In [31], a privacy vs. cooperation trade-off for multi-agent systems was presented. In [32], a privacy mechanism for consensus was presented, where privacy is measured in terms of estimation error covariance of the initial state. The authors in [33] showed that the privacy mechanism can be used by an attacker to execute stealthy attacks in a centralized setting.

In contrast to these works, we identify a novel and counter-intuitive trade-off between security and privacy in interconnected dynamical systems. In a preliminary version of this work [34], we compared the detection performance between the cases when the subsystems share full measurements (no privacy mechanism) and when they do not share any measurements. In this paper, we introduce a privacy framework and present an analytic characterization of privacy-performance trade-offs.

Contributions: The main contributions of this paper are as follows. First, we propose a privacy mechanism to keep the states of a subsystem private from other subsystems in an interconnected system. The mechanism limits both the amount and quality of shared measurements by projecting them onto an appropriate subspace and adding suitable noise to the measurements. This is in contrast to prior works which use only additive noise for privacy. We define a privacy ordering and use it to quantify and compare the privacy of different mechanisms. Second, we propose and characterize the performance of a chi-squared (χ^2) attack detection scheme to detect local attacks in absence of the knowledge of the global system model. The detection scheme uses local and received measurements from neighboring subsystems. Third, we characterize the trade-off between the privacy level and the local detection performance. Interestingly, our analysis shows that in some cases both privacy and detection performance can be improved by sharing less information. This reveals a counter-intuitive behavior of the widely used χ^2 test for attack detection [7, 8, 35], which we illustrate and explain.

Mathematical notation: $\text{Tr}(\cdot)$, $\text{Im}(\cdot)$, $\text{Null}(\cdot)$ and $\text{Rank}(\cdot)$ denote the trace, image, null space, and rank of a matrix, respectively. $(\cdot)^\mathsf{T}$ and $(\cdot)^+$ denote the transpose and Moore-Penrose pseudo-inverse of a matrix. A positive (semi)definite matrix A is denoted by $A > 0$ ($A \geq 0$). $\text{diag}(A_1, A_2, \cdots, A_n)$ denotes a block diagonal matrix whose block diagonal elements are $A_1, A_2, \cdots, A_n$. The identity matrix is denoted by I (or I_n to denote its dimension explicitly). A scalar $\lambda \in \mathbb{C}$ is called a generalized eigenvalue of (A, B) if $(A - \lambda B)$ is singular. $\otimes$ denotes the Kronecker product. A zero mean Gaussian random variable y is denoted by $y \sim \mathcal{N}(0, \Sigma_y)$, where Σ_y denotes the covariance of y. The (central) chi-square distribution with q degrees of freedom is denoted by χ^2_q and the noncentral chi-square distribution with noncentrality parameter λ is denoted by $\chi^2_q(\lambda)$. For $x \geq 0$, let $\mathcal{Q}_q(x)$ and $\mathcal{Q}_q(x; \lambda)$ denote the right tail probabilities of a chi-square and noncentral chi-square distributions, respectively.

2. Problem Formulation

We consider an interconnected discrete-time LTI dynamical system composed of N subsystems. Let $\mathcal{S} \triangleq \{1, 2, \cdots, N\}$ denote the set of all subsystems and let $\mathcal{S}_{-i} \triangleq \mathcal{S} \setminus \{i\}$, where $\setminus$ denotes the exclusion operator. The dynamics of the subsystems are given by:

$$x_i(k+1) = A_i x_i(k) + A_{-i} x_{-i}(k) + w_i(k), \tag{1}$$

$$y_i(k) = C_i x_i(k) + v_i(k) \qquad i \in \mathcal{S}, \tag{2}$$

where $x_i \in \mathbb{R}^{n_i}$ and $y_i \in \mathbb{R}^{p_i}$ are the state and output/measurements of subsystem i, respectively. Let $n \triangleq \sum_{i=1}^{N} n_i$. Subsystem i is coupled with other subsystems through the interconnection term $A_{-i} x_{-i}(k)$, where $x_{-i} \triangleq [x_1^\mathsf{T}, \cdots, x_{i-1}^\mathsf{T}, x_{i+1}^\mathsf{T}, \cdots, x_N^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{n-n_i}$ denotes the states of all other subsystems. We refer to x_{-i} as the interconnection signal. Further, $w_i \in \mathbb{R}^{n_i}$ and $v_i \in \mathbb{R}^{p_i}$ are the process and measurement noise, respectively. We assume that $w_i(k) \sim \mathcal{N}(0, \Sigma_{w_i})$ and $v_i(k) \sim \mathcal{N}(0, \Sigma_{v_i})$ for all $k \geq 0$, with $\Sigma_{w_i} > 0$ and $\Sigma_{v_i} > 0$. The process and measurement noise are assumed to be white and independent for different subsystems. Finally, we assume that the initial state $x_i(0) \sim \mathcal{N}(0, \Sigma_{x_i(0)})$ is independent of $w_i(k)$ and $v_i(k)$ for all $k \geq 0$. We make the following assumption regarding the interconnected system:

Assumption 1: Subsystem i has perfect knowledge of its dynamics, i.e., it knows (A_i, A_{-i}, C_i), the statistical properties of w_i, v_i and $x_i(0)$. However, it does not have knowledge of the dynamics, states, and the statistical properties of the noise of the other subsystems. □

Remark 1. ***(Control input)*** *The dynamics in* (1) *typically includes a control input. However, since each subsystem has the knowledge of its control input, its effect*

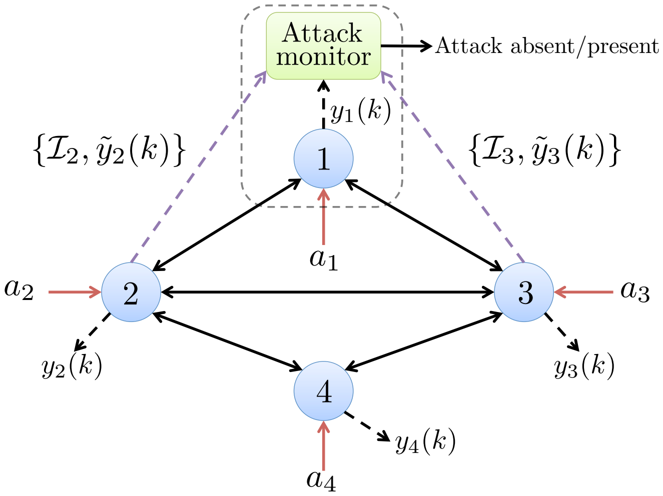


Figure 1: An interconnected system consisting of $N = 4$ subsystems. The solid lines represent state coupling among the subsystems. For attack detection by Subsystem 1, its neighboring agents 2 and 3 communicate their output information to 1 (denoted by dashed lines). The attack monitor associated with Subsystem 1 uses the received information and the local measurements to detect attacks.

can be easily included in the attack detection procedure. Therefore, for the ease of presentation, we omit the control input. □

We consider the scenario where each subsystem can be under an attack. We model the attacks as external linear additive inputs to the subsystems. The dynamics of the subsystems under attack are given by

$$x_i(k+1) = A_i x_i(k) + A_{-i} x_{-i}(k) + \underbrace{B_i^a \tilde{a}_i(k)}_{\triangleq\, a_i(k)} + w_i(k), \quad (3)$$

$$y_i(k) = C_i x_i(k) + v_i(k) \qquad i \in \mathcal{S}, \quad (4)$$

where $\tilde{a}_i \in \mathbb{R}^{r_i}$ is the local attack input for Subsystem i, which is assumed to be a deterministic but unknown signal for all $i \in \mathcal{S}$. The matrix B_i^a dictates how the attack $\tilde{a}_i$ affects the state of Subsystem i, which we assume to be unknown to Subsystem i.

Each subsystem is equipped with an attack monitor whose goal is to detect the local attack using the local measurements. Since Subsystem i does not know B_i^a, it can only detect $a_i = B_i^a \tilde{a}_i$. The detection procedure requires the knowledge of the statistical properties of y_i which depend on the interconnection signal x_{-i}. Since the subsystems do not have knowledge of the interconnection signals (c.f. *Assumption 1*), they share their measurements among each other to aid the local detection of attacks (see Fig. 1). The details of how these shared measurements are used for attack detection are presented in Section 4.

While the shared measurements help in detecting local attacks, they can reveal sensitive information of the subsystems. For instance, some of the states/outputs of a subsystem may be confidential, which it may not be willing to share with other subsystems. To protect the privacy of such states/outputs, we propose a privacy mechanism $\mathcal{M}_i$ through which a subsystem limits the amount and quality of its shared measurements. Thus, instead of sharing the complete measurements in (4), Subsystem i shares limited measurements (denoted as $\tilde{y}_i$) given by:

$$\begin{aligned}\mathcal{M}_i: \qquad \tilde{y}_i(k) &= S_i y_i(k) + \tilde{r}_i(k)\\ &= S_i C_i x_i(k) + S_i v_i(k) + \tilde{r}_i(k), \quad (5)\end{aligned}$$

where $S_i \in \mathbb{R}^{m_i \times p_i}$ is a selection matrix suitably chosen to select a subspace of the outputs, and $\tilde{r}_i(k) \sim \mathcal{N}(0, \Sigma_{\tilde{r}_i})$ is an artificial white noise (independent of w_i and v_i) added to introduce additional inaccuracy in the shared measurements. Without loss of generality, we assume S_i to be full row rank for all $i \in \mathcal{S}$. Thus, a subsystem can limit its shared measurement via a combination of the following two mechanisms (i) by sharing fewer (or a subspace of) measurements, and (ii) by sharing more noisy measurements. Intuitively, when Subsystem i limits its shared measurements, the estimates of its states/outputs computed by the other subsystems become more inaccurate. This prevents other subsystems from accurately determining the confidential states/outputs of Subsystem i, thereby protecting its privacy. We will explain this phenomenon in detail in the next section.

Let the parameters corresponding to the limited measurements of subsystem i be denoted by $\mathcal{I}_i \triangleq \{C_i, S_i, \Sigma_{v_i}, \Sigma_{\tilde{r}_i}\}$. We make the following assumption:
Assumption 2: Each subsystem $i \in \mathcal{S}$ shares its limited measurements $\tilde{y}_i$ in (5) and the parameters $\mathcal{I}_i$ with all subsystems $j \in \mathcal{S}_{-i}$.[1] □

Under *Assumptions 1* and *2*, the goal of each subsystem i is to detect the local attack a_i using its local measurements y_i and the limited measurements $\{\tilde{y}_j\}_{j \in \mathcal{S}_{-i}}$ received from the other subsystems (see Fig. 1).

3. Privacy Quantification

In this section, we quantify the privacy of the mechanism $\mathcal{M}_i$ in terms of the estimation error covariance of the state x_i. The estimation can be performed by any subsystem $j \in \mathcal{S}_{-i}$ that receives the limited measurements from Subsystem i. Then, we use this quantification to compare and rank different privacy mechanisms.

We use a batch estimation scheme in which the estimate is computed based on the collective measurements obtained for $k = 1, 2, \cdots, T$, with $T > 0$. Let $\tilde{y}_i = [\tilde{y}_i^\mathsf{T}(1), \cdots, \tilde{y}_i^\mathsf{T}(T)]^\mathsf{T}$, and let x_i, v_i, $\tilde{r}_i$ be similar time-aggregated vectors of $x_i(k), v_i(k), \tilde{r}_i(k)$, respectively. Then, using (5), we have

$$\tilde{y}_i = \underbrace{(I_T \otimes S_i C_i)}_{\triangleq\, H_i} x_i + \underbrace{(I_T \otimes S_i) v_i + \tilde{r}_i}_{\triangleq\, r_i}, \quad (6)$$

[1] To be precise, this information sharing is required only between *neighboring* subsystems, i.e., between subsystems that are directly coupled with each other in (1).

Preprint submitted to Automatica
Received July 3, 2019 07:18:35 PST

where $r_i \sim \mathcal{N}(0, \Sigma_{r_i})$ with $\Sigma_{r_i} = I_T \otimes (S_i \Sigma_{v_i} S_i^{\mathsf{T}} + \Sigma_{\tilde{r}_i})$. Note that any Subsystem $j \in \mathcal{S}_{-i}$ that receives measurements (6) from Subsystem i knows $\{H_i, \Sigma_{r_i}\}$ (c.f. *Assumption* 2). However, it is oblivious to the statistics of the confidential stochastic signal x_i. Thus, Subsystem j computes an estimate of x_i assuming that it is a deterministic but unknown quantity.

The Maximum Likelihood (ML) estimate of x_i based on $\tilde{y}_i$ is given by (using Lemma A.1):

$$\begin{aligned} &\hat{x}_i = \tilde{H}_i^+ H_i^{\mathsf{T}} \Sigma_{r_i}^{-1} \tilde{y}_i + (I - \tilde{H}_i^+ \tilde{H}_i) d_i, \quad \text{where} \\ &\tilde{H}_i \triangleq H_i^{\mathsf{T}} \Sigma_{r_i}^{-1} H_i \geq 0, \end{aligned} \tag{7}$$

and d_i is any real vector of appropriate dimension. If $\tilde{H}_i$ (or equivalently H_i) is not full column rank, then the estimate can lie anywhere in $\text{Null}(\tilde{H}_i) = \text{Null}(H_i)$ (shifted by $\tilde{H}_i^+ H_i^{\mathsf{T}} \Sigma_{r_i}^{-1} \tilde{y}_i$). Thus, the component of x_i that lies in $\text{Null}(H_i)$ cannot be estimated and only the component of x_i that lies in $\text{Im}(\tilde{H}_i) = \text{Im}(H_i^{\mathsf{T}})$ can be estimated. Let $\mathcal{P}_i \triangleq \tilde{H}_i^+ \tilde{H}_i$ denote the projection operator on $\text{Im}(\tilde{H}_i)$. The estimation error in this subspace is given by:

$$\begin{aligned} e_i = \mathcal{P}_i x_i - \mathcal{P}_i \hat{x}_i &= \tilde{H}_i^+ \tilde{H}_i x_i - \tilde{H}_i^+ H_i^{\mathsf{T}} \Sigma_{r_i}^{-1} \tilde{y}_i \\ &= -\tilde{H}_i^+ H_i^{\mathsf{T}} \Sigma_{r_i}^{-1} r_i, \end{aligned} \tag{8}$$

and the estimation error covariance is given by:

$$\Sigma_{e_i} = \mathbb{E}[\tilde{H}_i^+ H_i^{\mathsf{T}} \Sigma_{r_i}^{-1} r_i r_i^{\mathsf{T}} \Sigma_{r_i}^{-1} H_i \tilde{H}_i^+] = \tilde{H}_i^+. \tag{9}$$

Note that since the model in (6) is linear with Gaussian noise, $\mathcal{P}_i \hat{x}_i$ is the minimum-variance unbiased (MVU) estimate of x_i projected on $\text{Im}(H_i^{\mathsf{T}})$. Thus, the covariance Σ_{e_i} captures the fundamental limit on how accurately $\mathcal{P}_i x_i$ can be estimated and, therefore, it is a suitable metric to quantify privacy.

The privacy level of mechanism $\mathcal{M}_i$ in (5) is determined by two quantities: (i) $\text{rank}(S_i)$, and (ii) Σ_{e_i}. Intuitively, if $\text{rank}(S_i)$ is small, then Subsystem i shares fewer measurements and, as a result, the component of x_i that cannot be estimated ($(I - \tilde{H}_i^+ \tilde{H}_i) x_i$) becomes large. Further, if Σ_{e_i} is large (in a positive semi-definite sense), this implies that the estimation accuracy of the component of x_i that can be estimated ($\tilde{H}_i^+ \tilde{H}_i x_i$) is worse. Thus, a lower value of $\text{rank}(S_i)$ and a larger value of Σ_{e_i} implies a larger level of privacy. Based on this discussion, we next define an ordering between two privacy mechanisms.

Consider two privacy mechanisms $\mathcal{M}_i^{(1)}$ and $\mathcal{M}_i^{(2)}$, and let $\tilde{y}_i^{(k)}, \hat{x}_i^{(k)}, k = 1, 2$ denote the limited measurements and estimates corresponding to the two mechanisms, respectively. Further, let $S_i^{(k)}, H_i^{(k)}, \tilde{H}_i^{(k)}, \mathcal{P}_i^{(k)}, \Sigma_{e_i}^{(k)}, k = 1, 2$ denote the quantities defined above corresponding to $\mathcal{M}_i^{(1)}$ and $\mathcal{M}_i^{(2)}$.

Definition 1. ***(Privacy ordering)*** *Mechanism $\mathcal{M}_i^{(2)}$ is more private than $\mathcal{M}_i^{(1)}$, denoted by $\mathcal{M}_i^{(2)} \geq \mathcal{M}_i^{(1)}$, if*

$$\begin{aligned} &(i) \; \text{Im}\left((S_i^{(2)})^{\mathsf{T}}\right) \subseteq \text{Im}\left((S_i^{(1)})^{\mathsf{T}}\right) \quad \textit{and,} \\ &(ii) \; \Sigma_{e_i}^{(2)} \geq \mathcal{P}_i^{(2)} \Sigma_{e_i}^{(1)} \mathcal{P}_i^{(2)}. \end{aligned} \tag{10}$$

□

The first condition implies that $\tilde{y}_i^{(2)}$ is a limited version of $\tilde{y}_i^{(1)}$ and is required for the ordering to be well defined. Under this condition, it is easy to see that $\text{Im}(H_i^{(2)}) = \text{Im}(\mathcal{P}_i^{(2)}) \subseteq \text{Im}(H_i^{(1)}) = \text{Im}(\mathcal{P}_i^{(1)})$. Thus, the estimated component $\mathcal{P}_i^{(2)} \hat{x}_i^{(2)}$ lies in a subspace that is contained in the subspace of the estimated component $\mathcal{P}_i^{(1)} \hat{x}_i^{(1)}$. For a fair comparison between the two mechanisms, we consider the projection of $\mathcal{P}_i^{(1)} \hat{x}_i^{(1)}$ on $\text{Im}(\mathcal{P}_i^{(2)})$, given by $\mathcal{P}_i^{(2)} \mathcal{P}_i^{(1)} \hat{x}_i^{(1)} = \mathcal{P}_i^{(2)} \hat{x}_i^{(1)}$. Then, we compare its estimation error (given by $\mathcal{P}_i^{(2)} \Sigma_{e_i}^{(1)} \mathcal{P}_i^{(2)}$) with the estimation error of $\mathcal{P}_i^{(2)} \hat{x}_i^{(2)}$ (given by $\Sigma_{e_i}^{(2)}$) to obtain the second condition in (10). Next, we present an example to illustrate Definition 1.

Example 1. *Let $x_i \in \mathbb{R}^2$, $C_i = I_2$, $T = 1$, and consider two privacy mechanisms given by:*

$$\begin{aligned} \mathcal{M}_i^{(1)}: \qquad & \tilde{y}_i^{(1)} = (x_i + v_i) + \tilde{r}_i^{(1)}, \\ \mathcal{M}_i^{(2)}: \qquad & \tilde{y}_i^{(2)} = \begin{bmatrix} 1 & 0 \end{bmatrix} (x_i + v_i) + \tilde{r}_i^{(2)}, \end{aligned}$$

with $\Sigma_{v_i} = \Sigma_{\tilde{r}_i}^{(1)} = I_2$ and $\Sigma_{\tilde{r}_i}^{(2)} = \alpha \geq 0$. Mechanism $\mathcal{M}_i^{(1)}$ shares both components of the measurement vector y_i ($S_i^{(1)} = I_2$) whereas $\mathcal{M}_i^{(2)}$ shares only the first component ($S_i^{(2)} = [1 \;\, 0]$), and both add some artificial noise. The state estimates under the two mechanisms (using (7)) are given by

$$\hat{x}_i^{(1)} = \tilde{y}_i^{(1)} \quad \textit{and} \quad \hat{x}_i^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tilde{y}_i^{(2)} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} d_i.$$

Thus, under $\mathcal{M}_i^{(1)}$ both components of x_i can be estimated while under $\mathcal{M}_i^{(2)}$, only the first component can be estimated. Further, we have $\Sigma_{e_i}^{(1)} = 2I_2, \Sigma_{e_i}^{(2)} = \left[\begin{smallmatrix} 1+\alpha & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $\mathcal{P}_i^{(2)} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$. Thus, the estimation error covariance of the first component of x_i under $\mathcal{M}_i^{(1)}$ and $\mathcal{M}_i^{(2)}$ are 2 and $1 + \alpha$, respectively, and $\mathcal{M}_i^{(2)}$ is more private than $\mathcal{M}_i^{(1)}$ if $\alpha \geq 1$.

On the other hand, if $\alpha < 1$, then an ordering between the mechanisms cannot be established. In this case, under $\mathcal{M}_i^{(1)}$, both the state components can be estimated but the estimation error in first component is large. In contrast, under $\mathcal{M}_i^{(2)}$, only the first component can be estimated but its estimation error is small. □

Next, we state a sufficient condition on the noise added by two privacy mechanisms that guarantee the ordering of the mechanisms. This condition implies that, if one privacy mechanism shares a subspace of the measurements of the other mechanism and injects a sufficiently large amount of noise, then it is more private.

Lemma 3.1. ***(Sufficient condition for privacy ordering)*** *Consider two privacy mechanisms $\mathcal{M}_i^{(1)}$ and $\mathcal{M}_i^{(2)}$ in (5) with parameters $(S_i^{(k)}, \Sigma_{\tilde{r}_i}^{(k)})$, $k = 1, 2$ that satisfy* *condition* (i) *of* (10). *Let* P *be a full row rank matrix that satisfies* $S_i^{(2)} = PS_i^{(1)}$. *If*

$$\Sigma_{\tilde{r}_i}^{(2)} \geq P\Sigma_{\tilde{r}_i}^{(1)} P^\mathsf{T}, \tag{11}$$

then $\mathcal{M}_i^{(2)}$ *is more private than* $\mathcal{M}_i^{(1)}$.

Proof: From (6) and (7), we have

$$\tilde{H}_i^{(k)} = I_T \otimes \underbrace{(S_i^{(k)}C_i)^\mathsf{T} \left[S_i^{(k)}\Sigma_{v_i}(S_i^{(k)})^\mathsf{T} + \Sigma_{\tilde{r}_i}^{(k)}\right]^{-1} S_i^{(k)}C_i}_{\triangleq Y^{(k)}}.$$

Since $(S_i^{(1)}, S_i^{(2)})$ satisfy (10) (i), there always exist a full row rank matrix P satisfying $S_i^{(2)} = PS_i^{(1)}$. Next we have,

$$\begin{aligned} Y^{(2)} &= (S_i^{(1)}C_i)^\mathsf{T} P^\mathsf{T} \left[PS_i^{(1)}\Sigma_{v_i}(S_i^{(1)})^\mathsf{T} P^\mathsf{T} + \Sigma_{\tilde{r}_i}^{(2)}\right]^{-1} PS_i^{(1)}C_i \\ &= (S_i^{(1)}C_i)^\mathsf{T} P^\mathsf{T} \left[P(S_i^{(1)}\Sigma_{v_i}(S_i^{(1)})^\mathsf{T} + \Sigma_{\tilde{r}_i}^{(1)})P^\mathsf{T} + E\right]^{-1} PS_i^{(1)}C_i \\ &\overset{(a)}{\leq} (S_i^{(1)}C_i)^\mathsf{T}\left[S_i^{(1)}\Sigma_{v_i}(S_i^{(1)})^\mathsf{T} + \Sigma_{\tilde{r}_i}^{(1)}\right]^{-1} S_i^{(1)}C_i = Y^{(1)} \end{aligned} \tag{12}$$

where $E \triangleq \Sigma_{\tilde{r}_i}^{(2)} - P\Sigma_{\tilde{r}_i}^{(1)}P^\mathsf{T}$ and (a) follows from $E \geq 0$ (using (11)) and Lemma A.3. From (12), it follows that

$$\begin{aligned} &\tilde{H}_i^{(2)} \leq \tilde{H}_i^{(1)} \overset{(b)}{\Rightarrow} \tilde{H}_i^{(2)} \geq \tilde{H}_i^{(2)}(\tilde{H}_i^{(1)})^+\tilde{H}_i^{(2)} \\ &\overset{(c)}{\Rightarrow} (\tilde{H}_i^{(2)})^+\tilde{H}_i^{(2)}(\tilde{H}_i^{(2)})^+ \geq (\tilde{H}_i^{(2)})^+\tilde{H}_i^{(2)}(\tilde{H}_i^{(1)})^+\tilde{H}_i^{(2)}(\tilde{H}_i^{(2)})^+ \\ &\overset{(d)}{=} \text{Condition } (ii) \text{ in } (10), \end{aligned}$$

where (b) follows from [36, Lemma 1], and (c), (d) follow from facts that $(\tilde{H}_i^{(k)})^+$ is symmetric and $(\tilde{H}_i^{(k)})^+\tilde{H}_i^{(k)} = \tilde{H}_i^{(k)}(\tilde{H}_i^{(k)})^+$. Thus, both conditions in (10) are satisfied and $\mathcal{M}_i^{(2)} \geq \mathcal{M}_i^{(1)}$. ■

We conclude this section by showing that the privacy mechanism in (5) exhibits an intuitive post-processing property. It implies that if we further limit the measurements produced by a privacy mechanism, then this operation cannot decrease the privacy of the measurements. This post-processing property also holds in the differential privacy framework [26].

Lemma 3.2. ***(Post-processing increases privacy)*** *Consider two privacy mechanisms* $\mathcal{M}_i^{(1)}$ *and* $\mathcal{M}_i^{(2)}$, *where* $\mathcal{M}_i^{(2)}$ *further limits the measurements of* $\mathcal{M}_i^{(1)}$ *as:*

$$\begin{aligned} \mathcal{M}_i^{(1)}: &\qquad \tilde{y}_i^{(1)}(k) = S_i^{(1)}y_i(k) + \tilde{r}_i^{(1)}(k) \\ \mathcal{M}_i^{(2)}: &\qquad \tilde{y}_i^{(2)}(k) = S\tilde{y}_i^{(1)}(k) + n_i(k), \end{aligned}$$

where S *is full row rank and* $n_i(k) \sim \mathcal{N}(0, \Sigma_{n_i})$. *Then,* $\mathcal{M}_i^{(2)}$ *is more private than* $\mathcal{M}_i^{(1)}$.

Proof: It is easy to observe that $S_i^{(2)} = SS_i^{(1)}$ and $\tilde{r}_i^{(2)}(k) = S\tilde{r}_i^{(1)}(k) + n_i(k)$. Thus,

$$\Sigma_{\tilde{r}_i}^{(2)} = S\Sigma_{\tilde{r}_i}^{(1)}S^\mathsf{T} + \Sigma_{n_i} \geq S\Sigma_{\tilde{r}_i}^{(1)}S^\mathsf{T},$$

and the result follows from Lemma 3.1. ■

4. Local Attack detection

In this section we present the local attack detection procedure of the subsystems and characterize their detection performance. For the ease or presentation, we describe the analysis for Subsystem 1 and remark that the procedure is analogous for the other subsystems.

4.1. Measurement collection

We employ a batch detection scheme in which each subsystem collects the measurements for $k = 1, 2, \cdots, T$, with $T > 0$, and performs detection based on the collective measurements. In this subsection, we model the collected local and shared measurements for Subsystem 1.

Local measurements: Let the time-aggregated local measurements, interconnection signals, attacks, process noise and measurement noise corresponding to Subsystem 1 be respectively denoted by

$$\begin{aligned} y_L &\triangleq [y_1^\mathsf{T}(1), y_1^\mathsf{T}(2), \cdots, y_1^\mathsf{T}(T)]^\mathsf{T}, \\ x &\triangleq [x_{-1}^\mathsf{T}(0), x_{-1}^\mathsf{T}(1), \cdots, x_{-1}^\mathsf{T}(T-1)]^\mathsf{T}, \\ \tilde{a} &\triangleq [\tilde{a}_1^\mathsf{T}(0), \tilde{a}_1^\mathsf{T}(1), \cdots, \tilde{a}_1^\mathsf{T}(T-1)]^\mathsf{T}, \\ w &\triangleq [w_1^\mathsf{T}(0), w_1^\mathsf{T}(1), \cdots, w_1^\mathsf{T}(T-1)]^\mathsf{T}, \\ v &\triangleq [v_1^\mathsf{T}(1), v_1^\mathsf{T}(2), \cdots, v_1^\mathsf{T}(T)]^\mathsf{T}, \quad \text{and let} \\ F(Z) &\triangleq \begin{bmatrix} C_1Z & 0 & \cdots & 0 \\ C_1A_1Z & C_1Z & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_1A_1^{T-1}Z & C_1A_1^{T-2}Z & \cdots & C_1Z \end{bmatrix} \\ &= F(I)(I_T \otimes Z). \end{aligned}$$

By using (3) recursively and (4), the local measurements can be written as

$$y_L = Ox_1(0) + F_x x + F_{\tilde{a}}\tilde{a} + F_w w + v, \tag{13}$$

$$\text{where} \quad F_x = F(A_{-1}),\ F_{\tilde{a}} = F(B_1^a),\ F_w = F(I), \quad \text{and}$$

$$O \triangleq \begin{bmatrix}(C_1A_1)^\mathsf{T} & (C_1A_1^2)^\mathsf{T} & \cdots & (C_1A_1^T)^\mathsf{T}\end{bmatrix}^\mathsf{T}.$$

Note that $w \sim \mathcal{N}(0, \Sigma_w)$ and $v \sim \mathcal{N}(0, \Sigma_v)$ with

$$\Sigma_w = I_T \otimes \Sigma_{w_1} > 0 \quad \text{and} \quad \Sigma_v = I_T \otimes \Sigma_{v_1} > 0.$$

Let $v_L \triangleq Ox_1(0) + F_w w + v$ denote the effective local noise in the measurement equation (13). Using the fact that $(x_1(0), w, v)$ are independent, the overall local measurements of the subsystem are given by

$$\begin{aligned} &y_L = F_x x + F_{\tilde{a}}\tilde{a} + v_L, \quad \text{where} \\ &v_L \sim \mathcal{N}(0, \Sigma_{v_L}),\ \Sigma_{v_L} = O\Sigma_{x_1(0)}O^\mathsf{T} + F_w\Sigma_w F_w^\mathsf{T} + \Sigma_v > 0. \end{aligned} \tag{14}$$

Shared measurements: Let $\tilde{y}_{-1}(k) \triangleq [\tilde{y}_2^\mathsf{T}(k), \tilde{y}_3^\mathsf{T}(k), \cdots, \tilde{y}_N^\mathsf{T}(k)]^\mathsf{T}$ denote the limited measurements received by Subsystem 1 from all the other subsystems at time k. Further, let $v_{-1}(k)$ and $\tilde{r}_{-1}(k)$ denote similar aggregated vectors of $\{v_j(k)\}_{j\in\mathcal{S}_{-1}}$ and $\{\tilde{r}_j(k)\}_{j\in\mathcal{S}_{-1}}$, respectively. Then, from (5) we have

$$
\begin{aligned}
&\tilde{y}_{-1}(k) = S_{-1}C_{-1}x_{-1}(k) + S_{-1}v_{-1}(k) + \tilde{r}_{-1}(k), \qquad (15)\\
&\text{where } \; S_{-1} \triangleq \operatorname{diag}(S_2, \cdots, S_N), C_{-1} \triangleq \operatorname{diag}(C_2, \cdots, C_N),\\
&v_{-1}(k) \sim \mathcal{N}(0, \Sigma_{v_{-1}}),\ \Sigma_{v_{-1}} = \operatorname{diag}(\Sigma_{v_2}, \cdots, \Sigma_{v_N}) > 0,\\
&\tilde{r}_{-1}(k) \sim \mathcal{N}(0, \Sigma_{\tilde{r}_{-1}}),\ \Sigma_{\tilde{r}_{-1}} = \operatorname{diag}(\Sigma_{\tilde{r}_2}, \cdots, \Sigma_{\tilde{r}_N}) \geq 0.
\end{aligned}
$$

Further, let the time-aggregated limited measurements received by Subsystem 1 be denoted by $y_R \triangleq [\tilde{y}_{-1}^\mathsf{T}(0), \tilde{y}_{-1}^\mathsf{T}(1), \cdots, \tilde{y}_{-1}^\mathsf{T}(T-1)]^\mathsf{T}$, and let v_R denote similar time-aggregated vector of $\{S_{-1}v_{-1}(k) + \tilde{r}_{-1}(k)\}_{k=0,\cdots,T-1}$. Then, from (15), the overall limited measurements received by Subsystem 1 read as

$$
\begin{aligned}
y_R &= Hx + v_R, \quad \text{where} \qquad (16)\\
H &\triangleq I_T \otimes S_{-1}C_{-1}, \quad \text{and} \quad v_R \sim \mathcal{N}(0, \Sigma_{v_R})\\
\text{with} \quad \Sigma_{v_R} &= I_T \otimes (S_{-1}\Sigma_{v_{-1}}S_{-1}^\mathsf{T} + \Sigma_{\tilde{r}_{-1}}) > 0.
\end{aligned}
$$

The goal of Subsystem 1 is to detect the local attack using the local and received measurements given by (14) and (16), respectively.

4.2. Measurement processing

Since Subsystem 1 does not have access to the interconnection signal x, it uses the received measurements to obtain an estimate of x. Similar to the previous section, the estimate is computed assuming x to be a deterministic but unknown quantity.

The maximum likelihood (ML) estimate of x using the received measurements in (16) is (using Lemma A.1)

$$
\begin{aligned}
\hat{x} &= \tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}y_R + (I - \tilde{H}^+\tilde{H})d, \quad \text{where}\\
\tilde{H} &\triangleq H^\mathsf{T}\Sigma_{v_R}^{-1}H \geq 0,
\end{aligned} \qquad (17)
$$

and d is any real vector of appropriate dimension. The component of x that lies in the null space of $\tilde{H}$ cannot be estimated. We decompose x as

$$
\begin{aligned}
x &= (I - \tilde{H}^+\tilde{H})x + \tilde{H}^+\tilde{H}x\\
&= (I - \tilde{H}^+\tilde{H})x + \tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}Hx\\
&\overset{(16)}{=} (I - \tilde{H}^+\tilde{H})x + \tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}(y_R - v_R). \qquad (18)
\end{aligned}
$$

Substituting x from (18) in (14), we get

$$
\begin{aligned}
y_L = {} & F_x(I - \tilde{H}^+\tilde{H})x + F_x\tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}(y_R - v_R)\\
& + F_{\tilde{a}}\tilde{a} + v_L. \qquad (19)
\end{aligned}
$$

Next, we process the local measurements in two steps. First, we subtract the known term $F_x\tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}y_R$. Second, we eliminate the component $(I - \tilde{H}^+\tilde{H})x$ (which cannot be estimated) by premultiplying (19) with a matrix M^T, where

$$
\begin{aligned}
M &= \text{Basis of Null}\left([F_x(I - \tilde{H}^+\tilde{H})]^\mathsf{T}\right),\\
&\Rightarrow M^\mathsf{T}F_x(I - \tilde{H}^+\tilde{H}) = 0. \qquad (20)
\end{aligned}
$$

Since the columns of M are basis vectors, M is full column rank. The processed measurements are given by

$$
\begin{aligned}
z &= M^\mathsf{T}(y_L - F_x\tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}y_R)\\
&\overset{(19),(20)}{=} M^\mathsf{T}F_{\tilde{a}}\tilde{a} + \underbrace{M^\mathsf{T}(v_L - F_x\tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}v_R)}_{\triangleq v_P}, \qquad (21)
\end{aligned}
$$

where $v_P \sim \mathcal{N}(0, \Sigma_{v_P})$. The random variables v_L and v_R are independent because they depend exclusively on the local and external subsystems' noise, respectively. Using this fact

$$
\begin{aligned}
\Sigma_{v_P} &= M^\mathsf{T}\left[\Sigma_{v_L} + F_x\tilde{H}^+H^\mathsf{T}\Sigma_{v_R}^{-1}\Sigma_{v_R}\Sigma_{v_R}^{-\mathsf{T}}H(\tilde{H}^+)^\mathsf{T}F_x^\mathsf{T}\right]M\\
&\overset{\tilde{H}^\mathsf{T}=\tilde{H}}{=} M^\mathsf{T}\Sigma_{v_L}M + M^\mathsf{T}F_x\tilde{H}^+F_x^\mathsf{T}M \overset{(a)}{>} 0, \qquad (22)
\end{aligned}
$$

where (a) follows from the facts that M is full column rank and $\Sigma_{v_L} > 0$. The processed measurements z in (21) depend only on the local attack $\tilde{a}$, and the Gaussian noise v_P whose statistics is known to Subsystem 1 (c.f. *Assumptions 1 and 2*), i.e. $z \sim \mathcal{N}(M^\mathsf{T}F_{\tilde{a}}\tilde{a}, \Sigma_{v_P})$. Thus, Subsystem 1 uses z to perform attack detection. Note that the attack vectors that belong to $\text{Null}(M^\mathsf{T}F_{\tilde{a}})$ cannot be detected.

The operation of elimination of the unknown component $(I - \tilde{H}^+\tilde{H})x$ from y_L also eliminates a component of the attack $\tilde{a}$. As a result, this operation increases the space of undetectable attack vectors from $\text{Null}(F_{\tilde{a}})$ to $\text{Null}(M^\mathsf{T}F_{\tilde{a}})$. In some cases, this operation could also result in complete elimination of attacks as shown in the next result.

Lemma 4.1. *Consider equation* (3) *and the limited measurements in* (5)*, and let* S_{-1}, C_{-1}, M *be defined in* (15) *and* (20)*. If*

$$
\operatorname{Im}(B_1^a) \subseteq \operatorname{Im}\left(A_{-1}\left[I - (S_{-1}C_{-1})^+(S_{-1}C_{-1})\right]\right), \qquad (23)
$$

then $M^\mathsf{T}F_{\tilde{a}} = 0$.

Proof: Since $\text{Null}(\tilde{H}) = \text{Null}(H)$, we have

$$
\tilde{H}^+\tilde{H} = H^+H = I_T \otimes (S_{-1}C_{-1})^+(S_{-1}C_{-1}).
$$

Let $Z \triangleq (S_{-1}C_{-1})^+(S_{-1}C_{-1})$. Then, substituting F_x from (13) in (20), we get

$$
\begin{aligned}
&M^\mathsf{T}F(I)(I_T \otimes A_{-1})[I - I_T \otimes Z] = 0\\
\Rightarrow &M^\mathsf{T}F(I)(I_T \otimes A_{-1})[I_T \otimes (I - Z)] = 0\\
\Rightarrow &M^\mathsf{T}F(I)(I_T \otimes A_{-1}[I - Z]) = 0. \qquad (24)
\end{aligned}
$$

If (23) holds, then there exists a matrix P such that $B_1^a = A_{-1}[I - Z]P$. Thus, from (13), we have

$$
\begin{aligned}
M^\mathsf{T}F_{\tilde{a}} &= M^\mathsf{T}F(I)(I_T \otimes A_{-1}[I - Z]P)\\
&= M^\mathsf{T}F(I)(I_T \otimes A_{-1}[I - Z])(I_T \otimes P) \overset{(24)}{=} 0.
\end{aligned}
$$

■

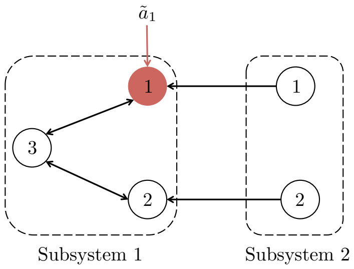


Figure 2: An interconnected system consisting of two subsystems. The nodes denote the states of the subsystems and solid edges denote the couplings and interconnections of Subsystem 1 (self edges are omitted). The attacked node is shaded in red.

The above result has the following intuitive interpretation: if the attacks lie in the subspace of the interconnections that cannot be estimated, then eliminating these interconnections also eliminates the attacks. In this case, the processed measurements do not have any signature of the attacks, which, therefore, cannot be detected. This result highlights the limitation of our measurement processing procedure. Next, we illustrate the result using an example.

Example 2. *Consider an interconnected subsystem consisting of two subsystems with the following parameters (see Fig. 2):*

$$A_1 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}, \; A_{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \; B_1^a = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$C_1 = I_3, C_2 = I_2$ *and* $T = 1$. *We have* $F_x = A_{-1}$ *and* $F_{\tilde{a}} = B_1^a$. *Consider the following two cases:*
Case (i)*: Subsystem 2 shares its 2nd state, i.e.,* $S_2 = S_{-1} = \begin{bmatrix} 0 & 1 \end{bmatrix}$. *In this case, Subsystem* 1 *does not get information about the interconnection affecting its 1st state and the elimination of this interconnection also eliminates the attack. It can be verified that* $M = \left[\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]^\mathsf{T}$ *and* $M^\mathsf{T} B_1^a = 0$.
Case (ii)*: Subsystem 2 shares its 1st state, i.e.,* $S_2 = S_{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix}$. *In this case, Subsystem* 1 *gets information about the interconnection affecting its 1st state. Thus, its elimination is not required and this preserves the attack. It can be verified that* $M = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]^\mathsf{T}$ *and* $M^\mathsf{T} B_1^a \neq 0$.

4.3. Statistical hypothesis testing

The goal of Subsystem 1 is to determine whether it is under attack or not using the processed measurements z in (21). Recall that, since Subsystem 1 does not know B_1^a, it can only detect $a_1 = B_1^a \tilde{a}_1$. Let $a \triangleq [(B_1^a \tilde{a}_1(0))^\mathsf{T}, \cdots, (B_1^a \tilde{a}_1(T-1))^\mathsf{T}]^\mathsf{T}$. Then, from (13), we have $F_{\tilde{a}} \tilde{a} = F_a a$, where $F_a = F(I)$. Thus, processed measurements are distributed according to $z \sim \mathcal{N}(M^\mathsf{T} F_a a, \Sigma_{v_P})$. We cast the attack detection problem as a binary hypothesis testing problem. Since Subsystem 1 does not know the attack a, we consider the following *composite* (simple vs. composite) testing problem

$$\begin{aligned} &H_0: \quad a = 0 \quad \text{(Attack absent)} \qquad \text{vs} \\ &H_1: \quad a \neq 0 \quad \text{(Attack present)} \end{aligned}$$

We use the Generalized Likelihood Ratio Test (GLRT) criterion [37] for the above testing problem, which is given by

$$\frac{f(z|H_0)}{\sup\limits_a f(z|H_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \tau' \quad \text{where,} \tag{25}$$

$$\begin{aligned} f(z|H_0) &= \frac{1}{\sqrt{2\pi|\Sigma_{v_P}|}} e^{-\frac{1}{2} z^\mathsf{T} \Sigma_{v_P}^{-1} z} \quad \text{and,} \\ f(z|H_1) &= \frac{1}{\sqrt{2\pi|\Sigma_{v_P}|}} e^{-\frac{1}{2}(z - M^\mathsf{T} F_a a)^\mathsf{T} \Sigma_{v_P}^{-1} (z - M^\mathsf{T} F_a a)}, \end{aligned}$$

are the probability density functions of the multivariate Gaussian distribution of z under hypothesis H_0 and H_1, respectively, and τ' is a suitable threshold. Using the result in Lemma A.1 to compute the denominator in (25) and taking the logarithm, the test (25) can be equivalently written as

$$t(z) \triangleq z^\mathsf{T} \Sigma_{v_P}^{-1} M^\mathsf{T} F_a \tilde{M}^+ F_a^\mathsf{T} M \Sigma_{v_P}^{-1} z \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{26}$$
$$\text{where} \quad \tilde{M} = F_a^\mathsf{T} M \Sigma_{v_P}^{-1} M^\mathsf{T} F_a,$$

and $\tau \geq 0$ is the threshold. The above test is a χ^2 test since the test statistics $t(z)$ follows a chi-squared distribution (see Lemma 4.3). The next result simplifies the test statistics $t(z)$ and provides an interpretation of the test.

Lemma 4.2. ***(Simplification of test statistics)*** *Let* $\Sigma_{v_P}^{-1} = R^\mathsf{T} R$ *denote the Cholesky decomposition of* $\Sigma_{v_P}^{-1}$. *Then,*

$$\Sigma_{v_P}^{-1} M^\mathsf{T} F_a \tilde{M}^+ F_a^\mathsf{T} M \Sigma_{v_P}^{-1} = R^\mathsf{T} U U^\mathsf{T} R, \tag{27}$$

where U *is a matrix whose columns are the orthonormal basis vectors of* $\text{Im}(R M^\mathsf{T} F_a)$.

Proof: Let $M_1 \triangleq M^\mathsf{T} F_a$. Then

$$\tilde{M}^+ = (M_1^\mathsf{T} R^\mathsf{T} R M_1)^+ = (R M_1)^+ ((R M_1)^+)^\mathsf{T}.$$

Thus, we have

$$\begin{aligned} &\Sigma_{v_P}^{-1} M^\mathsf{T} F_a \tilde{M}^+ F_a^\mathsf{T} M \Sigma_{v_P}^{-1} \\ &\quad = (R^\mathsf{T} R) M_1 (R M_1)^+ ((R M_1)^+)^\mathsf{T} M_1^\mathsf{T} (R^\mathsf{T} R) \\ &\quad = R^\mathsf{T} (R M_1)(R M_1)^+ (R M_1)(R M_1)^+ R \\ &\quad = R^\mathsf{T} (R M_1)(R M_1)^+ R. \end{aligned}$$

Since $R M_1 (R M_1)^+$ is the orthogonal projection operator on $\text{Im}(R M_1)$, $R M_1 (R M_1)^+ = U U^\mathsf{T}$, and the proof is complete. ■

Using Lemma 4.2, the test (26) can be written as

$$t(z) = z^\mathsf{T} R^\mathsf{T} U U^\mathsf{T} R z \underset{H_0}{\overset{H_1}{\gtrless}} \tau. \tag{28}$$

Thus, the test compares the energy of the signal $U^\mathsf{T} Rz$ with a given threshold to detect the attacks. Next, we derive the distribution of the test statistics under both hypothesis.

Lemma 4.3. ***(Distribution of test statistics)*** *The distribution of test statistics $t(z)$ in* (28) *is given by*

$$t(z) \sim \chi^2_q \quad \textit{under } H_0, \tag{29}$$

$$t(z) \sim \chi^2_q(\lambda \triangleq a^\mathsf{T} \Lambda a) \quad \textit{under } H_1, \tag{30}$$

where $q = \mathrm{Rank}(M^\mathsf{T} F_a)$ *and* $\Lambda = F_a^\mathsf{T} M \Sigma_{v_P}^{-1} M^\mathsf{T} F_a$.

Proof: By the definition of U in (27), and recalling $\Sigma_{v_P}^{-1} = R^\mathsf{T} R$ with R being non-singular, we have

$$\mathrm{Rank}(U^\mathsf{T} U) = \mathrm{Rank}(U) = \mathrm{Rank}(RM^\mathsf{T} F_a) = \mathrm{Rank}(M^\mathsf{T} F_a).$$

Let $z' = U^\mathsf{T} Rz$. Under H_0, $z \sim \mathcal{N}(0, \Sigma_{v_P})$. Thus,

$$z' \sim \mathcal{N}(0, U^\mathsf{T} R \Sigma_{v_P} R^\mathsf{T} U) \overset{(a)}{=} \mathcal{N}(0, I_q),$$

where (a) follows from $R\Sigma_{v_P} R^\mathsf{T} = I$ and $U^\mathsf{T} U = I_q$. Therefore, $t(z) = (z')^\mathsf{T} z' \sim \chi^2_q$.

Let $M_1 = M^\mathsf{T} F_a$. Under H_1, $z \sim \mathcal{N}(M_1 a, \Sigma_{v_P})$. Thus,

$$z' \sim \mathcal{N}(U^\mathsf{T} R M_1 a, I_q)$$
$$\Rightarrow t(z) = (z')^\mathsf{T} z' \sim \chi^2_q(a^\mathsf{T} M_1^\mathsf{T} R^T U U^\mathsf{T} R M_1 a).$$

Using $UU^\mathsf{T} = RM_1(RM_1)^+$ from the proof of Lemma 4.2, we have

$$a^\mathsf{T} M_1^\mathsf{T} R^T U U^\mathsf{T} R M_1 a = a^\mathsf{T} (RM_1)^\mathsf{T} (RM_1)(RM_1)^+ (RM_1) a$$
$$= a^\mathsf{T} (RM_1)^\mathsf{T} (RM_1) a = a^\mathsf{T} M_1^\mathsf{T} \Sigma_{v_P}^{-1} M_1 a = \lambda,$$

and the proof is complete. ■

Remark 2. ***(Interpretation of detection parameters*** (q, λ)***)*** *The parameter q denotes the number of independent observations of the attack vector a in the processed measurements* (21)*. The parameter λ can be interpreted as the signal to noise ratio (SNR) of the processed measurements in* (21)*, where the signal of interest is the attack.* □

Next, we characterize the performance of the test (26). Let the probability of false alarm and probability of detection for the test be respectively denoted by

$$P_F = \mathrm{Prob}(t(z) > \tau | H_0) \overset{(a)}{=} \mathcal{Q}_q(\tau) \quad \text{and,}$$

$$P_D = \mathrm{Prob}(t(z) > \tau | H_1) \overset{(b)}{=} \mathcal{Q}_q(\tau; \lambda),$$

where (a) and (b) follow from (29) and (30), respectively. Recall that $\mathcal{Q}_q(x)$ and $\mathcal{Q}_q(x;\lambda)$ denote the right tail probabilities of chi-square and noncentral chi-square distributions, respectively. Inspired by the Neyman-Pearson test framework, we select the size (P_F) of the test and determine the threshold τ which provides the desired size. Then, we use the threshold to perform the test and compute the detection probability. Thus, we have

$$\tau(q, P_F) = \mathcal{Q}_q^{-1}(P_F), \tag{31}$$

$$P_D(q, \lambda, P_F) = \mathcal{Q}_q(\tau(q, P_F); \lambda). \tag{32}$$

The arguments in $\tau(q, P_F)$ and $P_D(q, \lambda, P_F)$ explicitly denote the dependence of these quantities on the detection parameters (q, λ) and the probability of false alarm (P_F). Note that the detection performance of Subsystem 1 is characterized by the pair (P_F, P_D), where a lower value of P_F and a higher value of P_D is desirable. Later, in order to compare the performance of two different tests, we select a common value of P_F for both of them, and then compare the detection probability P_D.

The next result states the dependence of the detection probability on the detection parameters (q, λ).

Lemma 4.4. ***(Dependence of detection performance on detection parameters*** (q, λ)***)*** *For any given false alarm probability P_F, the detection probability $P_D(q, \lambda, P_F)$ is decreasing in q and increasing in λ.*

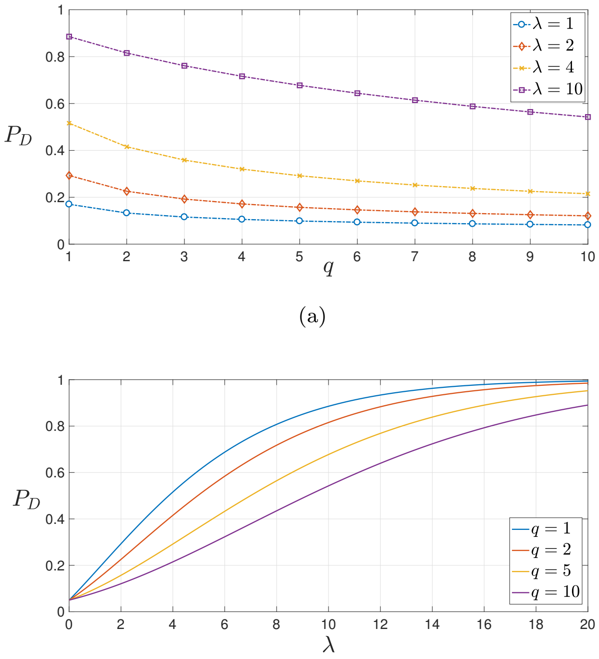
Proof: Since P_F is fixed, we omit it in the notation. It is a standard result that for a fixed q (and $\tau(q)$), the CDF $(= 1 - \mathcal{Q}_q(\tau(q); \lambda) = 1 - P_D(q, \lambda))$ of a noncentral chi-square random variable is decreasing in λ [38]. Thus, $P_D(q, \lambda)$ is increasing in λ.

Next, we have [38]

$$P_D(q, \lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} \mathcal{Q}_{q+2j}(\tau(q)).$$

From [39, Corollary 3.1], it follows that $\mathcal{Q}_{q+2j}(\tau(q)) = \mathcal{Q}_{q+2j}(\mathcal{Q}_q^{-1}(P_F))$ is decreasing in q for all $j > 0$. Thus, $P_D(q, \lambda)$ is decreasing in q. ■

Figure 3 illustrates the dependence of the detection probability on the parameters (q, λ). Lemma 4.4 implies that for a fixed q, a higher SNR (λ) leads to a better detection performance, which is intuitive. However, for a fixed λ, an increase in the number of independent observations (q) results in degradation of the detection performance. This counter-intuitive behavior is due to the fact that the GLRT in (25) is not an uniformly most powerful (UMP) test for all values of the attack a. In fact, a UMP test does not exist in this case [40]. Thus, the test can perform better for some particular attack values while it may not perform as good for other attack values. This suboptimality is an inherent property of the GLRT in (25). It arises due to the composite nature of the test and the fact that the value of the attack vector a is not known to the attack monitor.



(a)

(b)

Figure 3: Dependence of the detection probability P_D on the detection parameters (q, λ) for a fixed $P_F = 0.05$. P_D decreases monotonically with q (subfigure (a)), whereas it increases monotonically with λ (subfigure (b)).

Remark 3. ***(Composite vs. simple test)*** *If the value of the attack vector (say a_1) is known, we can cast a* simple *(simple vs. simple) binary hypothesis testing problem as $H_0 : a = 0$ vs. $H_1 : a = a_1$ and use the standard Likelihood Ratio Test criterion for detection. In this case the detection probability depends only on P_F and SNR (λ), and for any given P_F, the detection performance improves as the SNR increases.* □

5. Detection performance vs privacy trade-off

In this section, we present a trade-off between the attack detection performance and privacy of the subsystems. As before, we focus on detection for Subsystem 1 and consider two measurement sharing privacy mechanisms $\mathcal{M}_j^{(1)}$ and $\mathcal{M}_j^{(2)}$ for all other subsystems $j \in \mathcal{S}_{-1}$. The trade-off is between the *detection performance of Subsystem 1* and the *privacy level of all other subsystems.* The next result states the relation between the detection parameters corresponding to these two sets of privacy mechanisms.

Theorem 5.1. ***(Relation among the detection parameters of privacy mechanisms)*** *Let $\mathcal{M}_j^{(2)}$ be more private than $\mathcal{M}_j^{(1)}$ for all $j \in \mathcal{S}_{-1}$. Given any attack vector a, let $q^{(k)}$ and $\lambda^{(k)} = a^\mathsf{T}\Lambda^{(k)}a$ denote the detection parameters under the privacy mechanisms $\left\{\mathcal{M}_j^{(k)}\right\}_{j\in\mathcal{S}_{-1}}$, for $k = 1, 2$. Then, we have*

$$\begin{aligned} &(i)\ q^{(1)} \geq q^{(2)} \quad \text{and}, \\ &(ii)\ \lambda^{(2)}\mu_{max} \geq \lambda^{(1)} \geq \lambda^{(2)}\mu_{min} \geq \lambda^{(2)}, \end{aligned} \tag{33}$$

where μ_{max} and μ_{min} are the largest and smallest generalized eigenvalues of $(\Lambda^{(1)}, \Lambda^{(2)})$, respectively.

Proof: From (5), (15) and (16), for $k = 1, 2$, we have

$$\begin{aligned} H^{(k)} &= I_T \otimes \text{diag}\left(S_2^{(k)}C_2, \cdots, S_N^{(k)}C_N\right) = S_{-1}^{(k)}H, \\ \Sigma_{v_R}^{(k)} &= S_{-1}^{(k)}\Sigma_{v_R}(S_{-1}^{(k)})^\mathsf{T} + \Sigma_{\tilde{r}_{-1}}^{(k)} > 0 \quad \text{where}, \\ S_{-1}^{(k)} &= I_T \otimes \text{diag}\left(S_2^{(k)}, \cdots, S_N^{(k)}\right), \\ \Sigma_{\tilde{r}_{-1}}^{(k)} &= I_T \otimes \text{diag}\left(\Sigma_{\tilde{r}_2}^{(k)}, \cdots, \Sigma_{\tilde{r}_N}^{(k)}\right) \geq 0. \end{aligned}$$

Since $\mathcal{M}_j^{(2)} \geq \mathcal{M}_j^{(1)}$ for all $j \in \mathcal{S}_{-1}$, the first condition in (10) results in

$$\text{Im}\left((S_{-1}^{(1)})^\mathsf{T}\right) \supseteq \text{Im}\left((S_{-1}^{(2)})^\mathsf{T}\right) \Rightarrow \text{Im}\left((H^{(1)})^\mathsf{T}\right) \supseteq \text{Im}\left((H^{(2)})^\mathsf{T}\right).$$

From (17), we have $\tilde{H}^{(k)} = (H^{(k)})^\mathsf{T}(\Sigma_{v_R}^{(k)})^{-1}H^{(k)}$. Since $\text{Null}(\tilde{H}^{(k)}) = \text{Null}(H^{(k)})$, from (20), it follows that $\text{Im}(M^{(1)}) \supseteq \text{Im}(M^{(2)})$. Recalling from (30) that $q^{(k)} = \text{Rank}((M^{(k)})^\mathsf{T}F_a)$, it follows that $q^{(1)} \geq q^{(2)}$.

Since $\text{Im}(M^{(1)}) \supseteq \text{Im}(M^{(2)})$, we have $M^{(2)} = M^{(1)}P$ for some full column rank matrix P. Let $Z \triangleq F_x^\mathsf{T}M^{(1)}P$. From (22), we have

$$\begin{aligned} \Sigma_{v_P}^{(2)} &= (M^{(2)})^\mathsf{T}\Sigma_{v_L}M^{(2)} + (M^{(2)})^\mathsf{T}F_x(\tilde{H}^{(2)})^+F_x^\mathsf{T}M^{(2)}, \\ &= P^\mathsf{T}\Sigma_{v_P}^{(1)}P + \underbrace{Z^\mathsf{T}[(\tilde{H}^{(2)})^+ - (\tilde{H}^{(1)})^+]Z}_{\triangleq E}. \end{aligned} \tag{34}$$

Next, we show that $E \geq 0$. . Using $M^{(2)} = M^{(1)}P$, and using (20) for both $\{M^{(k)}, \tilde{H}^{(k)}\}$, $k = 1, 2$, we have

$$Z^\mathsf{T}(\tilde{H}^{(1)})^+\tilde{H}^{(1)} = Z^\mathsf{T}(\tilde{H}^{(2)})^+\tilde{H}^{(2)}. \tag{35}$$

Thus, we get

$$\begin{aligned} E &= Z^\mathsf{T}[(\tilde{H}^{(2)})^+ - (\tilde{H}^{(1)})^+\tilde{H}^{(1)}(\tilde{H}^{(1)})^+\tilde{H}^{(1)}(\tilde{H}^{(1)})^+]Z \\ &= Z^\mathsf{T}[(\tilde{H}^{(2)})^+ - (\tilde{H}^{(2)})^+\tilde{H}^{(2)}(\tilde{H}^{(1)})^+(\tilde{H}^{(2)})^+\tilde{H}^{(2)}]Z \end{aligned} \tag{36}$$

where the last inequality follows from (35) and the fact that $\tilde{H}^{(k)}(\tilde{H}^{(k)})^+ = (\tilde{H}^{(k)})^+\tilde{H}^{(k)}$. Next, we have,

$$\begin{aligned} \tilde{H}^{(k)} &= I_T \otimes \text{diag}\Big[(S_2^{(k)}C_2)^\mathsf{T}(S_2^{(k)}\Sigma_{v_2}(S_2^{(k)})^\mathsf{T} + \Sigma_{\tilde{r}_2}^{(k)})^{-1}S_2^{(k)}C_2, \\ &\cdots, (S_N^{(k)}C_N)^\mathsf{T}\left(S_N^{(k)}\Sigma_{v_N}(S_N^{(k)})^\mathsf{T} + \Sigma_{\tilde{r}_N}^{(k)}\right)^{-1}S_N^{(k)}C_N\Big] \\ &= \Pi^\mathsf{T}\text{diag}\Big[I_T \otimes (S_2^{(k)}C_2)^\mathsf{T}(S_2^{(k)}\Sigma_{v_2}(S_2^{(k)})^\mathsf{T} + \Sigma_{\tilde{r}_2}^{(k)})^{-1}S_2^{(k)}C_2, \\ &\cdots, I_T \otimes (S_N^{(k)}C_N)^\mathsf{T}\left(S_N^{(k)}\Sigma_{v_N}(S_N^{(k)})^\mathsf{T} + \Sigma_{\tilde{r}_N}^{(k)}\right)^{-1}S_N^{(k)}C_N\Big]\Pi \\ &= \Pi^\mathsf{T}\text{diag}\left[\tilde{H}_2^{(k)}, \cdots, \tilde{H}_N^{(k)}\right]\Pi \quad \text{and}, \end{aligned} \tag{37a}$$

$$(\tilde{H}^{(k)})^+ = \Pi^\mathsf{T}\text{diag}\left[(\tilde{H}_2^{(k)})^+, \cdots, (\tilde{H}_N^{(k)})^+\right]\Pi, \tag{37b}$$

where Π is a permutation matrix with $\Pi^{-1} = \Pi^{\mathsf{T}}$. Substituting (37a) and (37b) in (36), we have

$$E = Z^{\mathsf{T}}\Pi^{\mathsf{T}}\mathrm{diag}\Big[(\tilde{H}_2^{(2)})^{+} - \mathcal{P}_2^{(2)}(\tilde{H}_2^{(1)})^{+}\mathcal{P}_2^{(2)}, \cdots$$
$$(\tilde{H}_N^{(2)})^{+} - \mathcal{P}_N^{(2)}(\tilde{H}_N^{(1)})^{+}\mathcal{P}_N^{(2)}\Big]\Pi Z \overset{(a)}{\geq} 0,$$

where (a) follows from the second condition in (10) for all $j \in \mathcal{S}_{-1}$. Next, from (30), we have,

$$\begin{aligned}\Lambda^{(2)} &= F_a^{\mathsf{T}} M^{(2)}(\Sigma_{v_P}^{(2)})^{-1}(M^{(2)})^{\mathsf{T}}F_a \\ &\overset{(34)}{=} F_a^{\mathsf{T}} M^{(1)} \underbrace{P(P^{\mathsf{T}}\Sigma_{v_P}^{(1)}P + E)^{-1}P^{\mathsf{T}}}_{\triangleq Y}(M^{(1)})^{\mathsf{T}}F_a \\ &\overset{(b)}{\leq} F_a^{\mathsf{T}} M^{(1)}(\Sigma_{v_P}^{(1)})^{-1}(M^{(1)})^{\mathsf{T}}F_a = \Lambda^{(1)}, \\ \Rightarrow \lambda^{(1)} &= a^{\mathsf{T}}\Lambda^{(1)}a \geq a^{\mathsf{T}}\Lambda^{(2)}a = \lambda^{(2)},\end{aligned}$$

where (b) follows from Lemma A.3, and the facts that $E \geq 0$ and P is full column rank. Finally, the second condition in (33) follows from Lemma A.4 and the proof is complete. ■

Theorem 5.1 shows that when the subsystems $j \in \mathcal{S}_{-1}$ share measurements with Subsystem 1 using more private mechanisms, both the number of processed measurements and the SNR reduce. This has implications on the detection performance of Subsystem 1, as explained next. To compare the performance corresponding to the two sets of privacy mechanisms, we select the same false alarm probability P_F for both the cases and compare the detection probability. Theorem 5.1 and Lemma 4.4 imply that $P_D(q^{(2)}, \lambda^{(2)}, P_F)$ can be greater or smaller than $P_D(q^{(1)}, \lambda^{(1)}, P_F)$ depending on the actual values of the detection parameters. In fact, ignoring the dependency on P_F since it is same for both cases, we have

$$P_D(q^{(2)}, \lambda^{(2)}) - P_D(q^{(1)}, \lambda^{(1)}) =$$
$$\underbrace{P_D(q^{(2)}, \lambda^{(2)}) - P_D(q^{(2)}, \lambda^{(1)})}_{\leq 0} + \underbrace{P_D(q^{(2)}, \lambda^{(1)}) - P_D(q^{(1)}, \lambda^{(1)})}_{\geq 0}.$$

Intuitively, if the decrease in P_D due to the decrease in the SNR[2] ($\lambda^{(1)} \to \lambda^{(2)}$) is larger than the increase in P_D due to the decrease in the number of measurements ($q^{(1)} \to q^{(2)}$), then the the detection performance decreases, and vice-versa.

This is an interesting and counter-intuitive trade-off between the detection performance and privacy/information sharing, and it implies that, in certain cases, sharing less information can lead to a better detection performance. This phenomenon occurs because the GLRT for the considered hypothesis testing problem is a sub-optimal test, as discussed before.

Next, we compare the detection performance corresponding to two privacy mechanisms that share the same subspace of measurements.

Corollary 5.2. ***(Strict security-privacy trade-off)*** *Consider two privacy mechanisms $\mathcal{M}_j^{(2)} \geq \mathcal{M}_j^{(1)}$ such that $\mathrm{Im}\left((S_j^{(2)})^{\mathsf{T}}\right) = \mathrm{Im}\left((S_j^{(1)})^{\mathsf{T}}\right)$ for $j \in \mathcal{S}_{-1}$. Let $(q^{(k)}, \lambda^{(k)})$ denote the detection parameters of Subsystem 1 under the privacy mechanisms $\left\{\mathcal{M}_j^{(k)}\right\}_{j\in\mathcal{S}_{-1}}$, for $k = 1, 2$. Then, for any given P_F, we have*

$$P_D(q^{(2)}, \lambda^{(2)}, P_F) \leq P_D(q^{(1)}, \lambda^{(1)}, P_F).$$

Proof: Since the mechanisms share the same subspace of measurements, from the proof of Theorem 5.1, we have

$$\mathrm{Im}\Big((S_{-1}^{(1)})^{\mathsf{T}}\Big) = \mathrm{Im}\Big((S_{-1}^{(2)})^{\mathsf{T}}\Big) \Rightarrow \mathrm{Im}\Big((H^{(1)})^{\mathsf{T}}\Big) = \mathrm{Im}\Big((H^{(2)})^{\mathsf{T}}\Big)$$
$$\Rightarrow \mathrm{Im}\Big(M^{(1)}\Big) = \mathrm{Im}\Big(M^{(2)}\Big) \Rightarrow q^{(1)} = q^{(2)}.$$

The fact that $\lambda^{(1)} \geq \lambda^{(2)}$ follows from Theorem 5.1, and the result then follows from Lemma 4.4. ■

The above result implies that there is strict trade-off between privacy and detection performance when the subspace of the shared measurements is fixed and the privacy level is varied by changing the noise level. In this case, more private mechanisms result in a poorer detection performance, and vice-versa.

6. Simulation Example

Consider an interconnected system with $N = 3$ subsystems with the following parameters:

$$A_1 = \frac{1}{3}\begin{bmatrix} -1 & -16 & 2 & -4 \\ 0 & -6 & 1 & -1 \\ 0 & 2 & 1 & 1 \\ 1 & 28 & -3 & 6 \end{bmatrix}, A_{12} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix},$$

$$A_{13} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, C_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$A_{-1} = \begin{bmatrix} A_{12} & A_{13} \end{bmatrix}$, $\Sigma_{x_1(0)} = 0.2I_4$, $\Sigma_{w_1} = 0.1I_4$, $C_2 = I_3$ $C_3 = I_2$, $\Sigma_{v_1} = \Sigma_{v_2} = I_3$, $\Sigma_{v_3} = I_2$, $T = 2$.

We focus on the attack detection for Subsystem 1, where Subsystems 2 and 3 use privacy mechanisms to share their measurements with Subsystem 1. We consider the following three cases of privacy mechanisms for Subsystems 2 and 3:

- $\mathcal{M}^{(0)} = \{\mathcal{M}_2^{(0)}, \mathcal{M}_3^{(0)}\}$: Subsystems 2 and 3 do not use any privacy mechanisms and share actual measurements, i.e., $S_2 = I_3, S_3 = I_2, \Sigma_{\tilde{r}_2} = 0$, and $\Sigma_{\tilde{r}_3} = 0$.

[2] Note that the SNR depends upon the attack vector a (via (30)), which we do not know a-priori. Thus, depending on the actual attack value, the SNR can take any positive value.

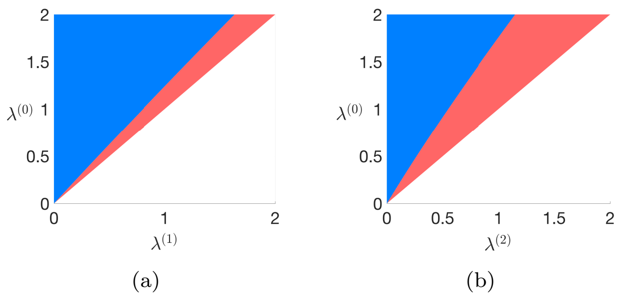


Figure 4: Comparison between detection performance of case 0 with: (a) case 1, and (b) case 2. In the blue region, case 0 performs better than case0/case 1, and vice versa in red square region. Since $\lambda^{(0)} \geq \lambda^{(k)}$ for $k = 1, 2$ (c.f. Lemma 5.1), the white region is inadmissible.

- $\mathcal{M}^{(1)}$: $S_2 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right], S_3 = I_2, \Sigma_{\tilde{r}_2} = 0$, and $\Sigma_{\tilde{r}_3} = I_2$.
- $\mathcal{M}^{(2)}$: $S_2 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right], S_3 = \left[\begin{smallmatrix} 0 & 1 \end{smallmatrix}\right], \Sigma_{\tilde{r}_2} = 0$, and $\Sigma_{\tilde{r}_3} = 1.8$.

Using Lemma 3.1, it can be easily verified that the following privacy ordering holds: $\mathcal{M}^{(2)} > \mathcal{M}^{(1)} > \mathcal{M}^{(0)}$. Recall that the detection performance is completely characterized by P_F and the detection parameters (q, λ). We choose $P_F = 0.05$ for all the cases. Let $(q^{(k)}, \lambda^{(k)})$, $k = 0, 1, 2$ denote the detection parameters for the above three cases. Recall that the parameter q depends only the system parameters, whereas the parameter λ depends on the system parameters as well as the attack values. For the above cases, we have $q^{(0)} = 6, q^{(1)} = 4$ and $q^{(3)} = 2$. Recalling (30), the value of $\lambda^{(k)} = a^{\mathsf{T}}\Lambda^{(k)}a$ can lie anywhere between $[0, \infty)$ depending on the attack value a. Thus, for simplicity, we present the results in this section in terms of $\lambda^{(k)}$.

We aim to compare the detection performance of case 0 with cases 1 and 2, respectively. We are interested in identifying the ranges of the detection parameters for which one case performs better than the other. As mentioned previously, the parameters $q^{(k)}$ are fixed for the three cases, so we compare the performance for different values of the parameter $\lambda^{(k)}$. Fig. 4 presents the performance comparison of case 0 with case 1 (Fig. 4(a)) and case 2 (Fig. 4(a)). Any point (x, y) in the colored regions are achievable by an attack, i.e., there exists an attack a such that $a^{\mathsf{T}}\Lambda^{(k)}a = x$ and $a^{\mathsf{T}}\Lambda^{(0)}a = y$, whereas the white region is inadmissible (see (33)). The blue region corresponds to the pairs $(\lambda^{(k)}, \lambda^{(0)})$ for which case 0 performs better than case k, i.e., $P_D(q^{(0)}, \lambda^{(0)}, P_F) \geq P_D(q^{(k)}, \lambda^{(k)}, P_F)$ for $k = 1, 2$. In the red region, case k performs better that case 0, $k = 1, 2$.

We observe that case 0 performs better than case k if $\frac{\lambda^{(0)}}{\lambda^{(k)}}$ is large, and vice versa. This shows that if the attack vector a is such that $\frac{\lambda^{(0)}}{\lambda^{(k)}}$ is small, then the detection performance corresponding to a more private mechanism ($\mathcal{M}^{(k)} > \mathcal{M}^{(0)}$) is better. This implies that there is non-strict trade-off between privacy and detection performance. This counter-intuitive result is due to the suboptimality of the GLRT used to perform detection, as explained before (c.f. discussion above Remark 3). Further, we observe that the red region of Fig. 4(b) is larger than (and contains) the red region of Fig. 4(b). This is because $\mathcal{M}^{(2)}$ is more private than $\mathcal{M}^{(1)}$.

Finally, we consider the case where Subsystems 2 and 3 implement their privacy mechanisms by only adding artificial noise in (5). Thus, $S_2 = I_3, S_3 = I_2$, and the artificial noise covariances are given by $\Sigma_{\tilde{r}_2} = \sigma^2 I_3$ and $\Sigma_{\tilde{r}_2} = \sigma^2 I_2$. The attack value is $\tilde{a}(k) = [1, 1]^{\mathsf{T}}$ for $k = 1, 2$. Clearly, as the noise level σ increases, the privacy level also increases. Fig. 5 shows the detection performance of Subsystem 1 for varying noise level σ. We observe that the detection performance is a decreasing function of the noise level (c.f. Corollary 5.2), implying a strict trade-off between detection performance and privacy in this case.

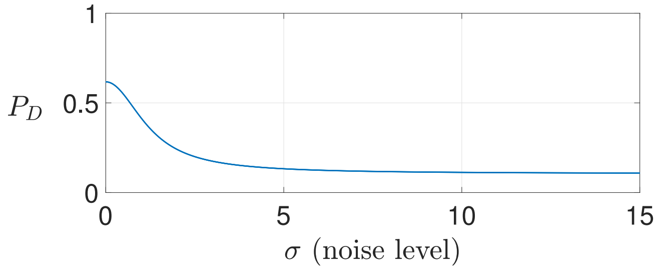


Figure 5: Detection performance for varying level of noise parameter σ.

7. Conclusion

We study an attack detection problem in interconnected dynamical systems where each subsystem is tasked with detection of local attacks without any knowledge of the dynamics of other subsystems and their interconnection signals. The subsystems share measurements among themselves to aid attack detection, but they also limit the amount and quality of the shared measurements due to privacy concerns. We show that there exists a non-strict trade-off between privacy and detection performance, and in some cases, sharing less measurements can improve the detection performance. We reason that this counter-intuitive result is due the suboptimality of the considered χ^2 test.

Future work includes exploring if this counter-intuitive trade-off exist for alternative detection schemes (for instance, unknown-input observers) and for other types of statistical tests. Also, recursive schemes to compute the state estimates, eliminate interconnections and compute the detection probability should also be explored.

APPENDIX

Lemma A.1. *The optimal solutions of the following weighted least squares problem:*

$$\min_{x} \quad J(x) = (y - Hx)^{\mathsf{T}}\Sigma^{-1}(y - Hx), \tag{38}$$

with $\Sigma > 0$ are given by

$$x^* = \tilde{H}^+ H^\mathsf{T} \Sigma^{-1} y + (I - \tilde{H}^+ \tilde{H}) d, \tag{39}$$

where $\tilde{H} = H^\mathsf{T} \Sigma^{-1} H$, and d is any real vector of appropriate dimension. Further, the optimal value of the cost is

$$J(x^*) = y^\mathsf{T} (\Sigma^{-1} - \Sigma^{-1} H \tilde{H}^+ H^\mathsf{T} \Sigma^{-1}) y. \tag{40}$$

Lemma A.2. *Let $\left[\begin{smallmatrix} A & B \\ B^\mathsf{T} & D \end{smallmatrix}\right]$ be a positive definite matrix with $A > 0$, $D \geq 0$. Further, let $M \geq 0$. Then,*

$$\left[\begin{smallmatrix} A & B \\ B^\mathsf{T} & D \end{smallmatrix}\right]^{-1} \geq \left[\begin{smallmatrix} (A+M)^{-1} & 0 \\ 0 & 0 \end{smallmatrix}\right],$$

Proof: Using the Schur complement, we have

$$\left[\begin{smallmatrix} A & B \\ B^\mathsf{T} & D \end{smallmatrix}\right]^{-1} = \left[\begin{smallmatrix} I & -A^{-1}B \\ 0 & I \end{smallmatrix}\right] \left[\begin{smallmatrix} A^{-1} & 0 \\ 0 & (D-B^\mathsf{T}A^{-1}B)^{-1} \end{smallmatrix}\right] \left[\begin{smallmatrix} I & 0 \\ -B^\mathsf{T}A^{-1} & I \end{smallmatrix}\right],$$

where the Schur complement $D - B^\mathsf{T} A^{-1} B > 0$. Further,

$$\left[\begin{smallmatrix} (A+M)^{-1} & 0 \\ 0 & 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} I & -A^{-1}B \\ 0 & I \end{smallmatrix}\right] \left[\begin{smallmatrix} (A+M)^{-1} & 0 \\ 0 & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} I & 0 \\ -B^\mathsf{T}A^{-1} & I \end{smallmatrix}\right]$$

Since $A + M \geq A$, $A^{-1} \geq (A + M)^{-1}$. Thus,

$$\left[\begin{smallmatrix} A^{-1}-(A+M)^{-1} & 0 \\ 0 & (D-B^\mathsf{T}A^{-1}B)^{-1} \end{smallmatrix}\right] \geq 0,$$

and the result follows. ■

Lemma A.3. *Let $\Sigma > 0 \in \mathbb{R}^{n \times n}$ and $\Sigma_a \geq 0 \in \mathbb{R}^{m \times m}$, with $m \leq n$, and let $S \in \mathbb{R}^{n \times m}$ be full (column) rank. Then,*

$$\Sigma^{-1} \geq S(S^\mathsf{T} \Sigma S + \Sigma_a)^{-1} S^\mathsf{T}. \tag{41}$$

Proof: Since S is full column rank, $S^\mathsf{T}\Sigma S > 0$, $S^\mathsf{T}\Sigma S + \Sigma_a$ is invertible and $S^+ S = I_m = S^\mathsf{T}(S^\mathsf{T})^+$. Thus, $I_n = \left[\begin{smallmatrix} S^\mathsf{T}(S^\mathsf{T})^+ & 0 \\ 0 & I_{n-m} \end{smallmatrix}\right]$. Let $N \in \mathbb{R}^{n \times (n-m)}$ denote a matrix whose columns are the basis of $\text{Null}(S^\mathsf{T})$. Then, $\left[\begin{smallmatrix} S^\mathsf{T}(S^\mathsf{T})^+ & 0 \end{smallmatrix}\right] = S^\mathsf{T} \left[\begin{smallmatrix} (S^\mathsf{T})^+ & N \end{smallmatrix}\right] \triangleq S^\mathsf{T} R$. Since, $\text{Im}((S^\mathsf{T})^+) = \text{Im}(S) \perp \text{Null}(S^\mathsf{T})$, R is non-singular. Let $T \triangleq \left[\begin{smallmatrix} 0 & I_{n-m} \end{smallmatrix}\right] R^{-1}$. Then, we have $I_n = \left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right] R = R \left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right]$. Thus,

$$\begin{aligned}
\Sigma^{-1} &= I_n^\mathsf{T} (I_n \Sigma I_n^\mathsf{T})^{-1} I_n \\
&= \left[\begin{smallmatrix} S & T^\mathsf{T} \end{smallmatrix}\right] R^\mathsf{T} \left(R \left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right] \Sigma \left[\begin{smallmatrix} S & T^\mathsf{T} \end{smallmatrix}\right] R^\mathsf{T}\right)^{-1} R \left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right] \\
&= \left[\begin{smallmatrix} S & T^\mathsf{T} \end{smallmatrix}\right] \left(\left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right] \Sigma \left[\begin{smallmatrix} S & T^\mathsf{T} \end{smallmatrix}\right]\right)^{-1} \left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right] \\
&= \left[\begin{smallmatrix} S & T^\mathsf{T} \end{smallmatrix}\right] \left[\begin{smallmatrix} S^\mathsf{T}\Sigma S & S^\mathsf{T}\Sigma T^\mathsf{T} \\ T\Sigma S & T\Sigma T^\mathsf{T} \end{smallmatrix}\right]^{-1} \left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right], \quad \text{and}
\end{aligned}$$

$$S(S^\mathsf{T}\Sigma S + \Sigma_a)^{-1} S^\mathsf{T} = \left[\begin{smallmatrix} S & T^\mathsf{T} \end{smallmatrix}\right] \left[\begin{smallmatrix} (S^\mathsf{T}\Sigma S+\Sigma_a)^{-1} & 0 \\ 0 & 0 \end{smallmatrix}\right]^{-1} \left[\begin{smallmatrix} S^\mathsf{T} \\ T \end{smallmatrix}\right].$$

The result follows from Lemma A.2. ■

Lemma A.4. *Let $M_1 \geq M_2 \geq 0$, $\lambda \geq 0$ and let $J(x) = x^\mathsf{T} M_1 x$. Then, the maximum and minimum values of $J(x)$ subject to $x^\mathsf{T} M_2 x = \lambda$ are given by $\lambda\mu_{max}$ and $\lambda\mu_{min}$ respectively, where μ_{max} and μ_{min} are the largest and smallest generalized eigenvalues of (M_1, M_2), respectively.*

Proof: Consider the following optimization problem

$$\min_x / \max \quad J(x) = x^\mathsf{T} M_1 x, \quad \text{subject to} \quad x^\mathsf{T} M_2 x = \lambda.$$

The Lagrangian of this problem is given by $l = x^\mathsf{T} M_1 x - \mu(x^\mathsf{T} M_2 x - \lambda)$, where $\mu \in \mathbb{R}$ is the Lagrange multiplier. By differentiating l, the first order optimality condition is given by $(M_1 - \mu M_2)x = 0$. Thus, μ is a generalized eigenvalue of (M_1, M_2). Further, using $M_1 x = \mu M_2 x$, the cost at the optimum is given by $\lambda\mu$ and the maximum and minimum values of the cost given in the lemma follow. ■

References

[1] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine*, 21(6):11–25, 2001.

[2] A. Cardenas, S. Amin, and S. Sastry. Secure control: Towards survivable cyber-physical systems. In *International Conference on Distributed Computing Systems Workshops*, page 495—500, Beijing, China, 2008.

[3] J. Giraldo, E. Sarkar, A. Cardenas, M. Maniatakos, and M. Kantarcioglu. Security and privacy in cyber-physical systems: A survey of surveys. *IEEE Design & Test*, 34(4):7–17, 2017.

[4] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[5] F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[6] Y. Chen, S. Kar, and J. M. F. Moura. Dynamic attack detection in cyber-physical systems with side initial state information. *IEEE Transactions on Automatic Control*, 62(9):4618–4624, 2017.

[7] Y. Mo and B. Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61(9):2618–2624, 2016.

[8] Y. Chen, S. Kar, and J. M. F. Moura. Optimal attack strategies subject to detection constraints against cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 5(3):1157–1168, 2018.

[9] H. Nishino and H. Ishii. Distributed detection of cyber attacks and faults for power systems. In *IFAC World Congress*, pages 11932–11937, Cape Town, South Africa, August 2014.

[10] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer. Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions. *IEEE Signal Processing Magazine*, 29(5):106–115, 2012.

[11] F. Dörfler, F. Pasqualetti, and F. Bullo. Distributed detection of cyber-physical attacks in power networks: A waveform relaxation approach. In *Allerton Conf. on Communications, Control and Computing*, September 2011.

[12] F. Pasqualetti, F. Dörfler, and F. Bullo. A divide-and-conquer approach to distributed attack identification. In *IEEE Conf. on Decision and Control*, pages 5801–5807, Osaka, Japan, December 2015.

[13] N. Forti, G. Battistelli, L. Chisci, S. Li, B. Wang, and B. Sinopoli. Distributed joint attack detection and secure state estimation. *IEEE Transactions on Signal and Information Processing over Networks*, 4(1):96–110, 2018.

[14] Y. Guan and X. Ge. Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks. *IEEE Transactions on Signal and Information Processing over Networks*, 4(1):48–59, 2018.

[15] F. Boem, A. J. Gallo, G. Ferrari-Trecate, and T. Parisini. A distributed attack detection method for multi-agent systems governed by consensus-based control. In *IEEE Conf. on Decision and Control*, pages 5961–5966, Melbourne, Australia, 2017.

[16] A. Teixeira, H. Sandberg, and K. H. Johansson. Networked control systems under cyber attacks with applications to power networks. In *American Control Conference*, pages 3690–3696, 2010.

[17] R. Anguluri, V. Katewa, and F. Pasqualetti. Centralized versus decentralized detection of attacks in stochastic interconnected systems. *arXiv:1903.10109*, 2019.

[18] R. M. G. Ferrari, T. Parisian, and M. M. Polycarpou. Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach. *IEEE Transactions on Automatic Control*, 57(2):275–290, 2012.

[19] C. Kiliris, M. M. Polycarpou, and T. Parisian. A robust nonlinear observer-based approach for distributed fault detection of input–output interconnected systems. *Automatica*, 53:408–415, 2015.

[20] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou. Distributed sensor fault diagnosis for a network of interconnected cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 2(1):11–23, 2015.

[21] X. Zhang and Q. Zhang. Distributed fault diagnosis in a class of interconnected nonlinear uncertain systems. *International Journal of Control*, 85(11):1644–1662, 2012.

[22] X. G. Yan and C. Edwards. Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer. *International Journal of Control*, 81(4):591–606, 2008.

[23] E. Franco, R. Olfati-Saber, T. Parisini, and M. M. Polycarpou. Distributed fault diagnosis using sensor networks and consensus-based filters. In *IEEE Conf. on Decision and Control*, pages 386–391, San Diego, CA, USA, December 2006.

[24] S. Stankovic, N. Ilic, Z. Djurovic, M. Stankovic, and K. H. Johansson. Consensus based overlapping decentralized fault detection and isolation. In *Conference on Control and Fault Tolerant Systems*, Nice, France, 2010.

[25] I. Shames, A. M. H. Teixeira, H. Sandberg, and K. H. Johansson. Distributed fault detection for interconnected second-order systems. *Automatica*, 47:2757–2764, 2011.

[26] J. Cortes, G. E. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. J. Pappas. Differential privacy in control and network systems. In *IEEE Conf. on Decision and Control*, pages 4252–4272, Las Vegas, USA, 2016.

[27] E. Akyol, C. Langbort, and T. Basar. Privacy constrained information processing. In *IEEE Conf. on Decision and Control*, Osaka, Japan, 2015.

[28] F. Farokhi and G. Nair. Privacy-constrained communication. In *IFAC Workshop on Distributed Estimation and Control in Networked Systems*, pages 43–48, Tokyo, Japan, September 2016.

[29] T. Tanaka, M. Skoglund, H. Sandberg, and K. H. Johansson. Directed information and privacy loss in cloud-based control. In *American Control Conference*, Seattle, USA, 2017.

[30] F. Farokhi and H. Sandberg. Ensuring privacy with constrained additive noise by minimizing fisher information. *Automatica*, 99:275–288, 2019.

[31] V. Katewa, F. Pasqualetti, and V. Gupta. On privacy vs cooperation in multi-agent systems. *International Journal of Control*, 91(7):1693–1707, 2018.

[32] Y. Mo and R. M. Murray. Privacy-preserving average consensus. *IEEE Transactions on Automatic Control*, 62(2):753–765, 2017.

[33] J. Giraldo, A. Cardenas, and M. Kantarcioglu. Security and privacy trade-offs in cps by leveraging inherent differential privacy. In *IEEE Conference on Control Technology and Applications*, pages 1313–1318, Hawaii, USA, 2017.

[34] R. Anguluri, V. Katewa, and F. Pasqualetti. On the role of information sharing in the security of interconnected systems. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Honolulu, Hi, 2018.

[35] A. S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12:601–611, 1976.

[36] R. E. Hartwig. A note on the partial ordering of positive semidefinite matrices. *Linear and Multilinear Algebra*, 6(3):223–226, 1978.

[37] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

[38] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Volume 2*. Wiley-Interscience, 1995.

[39] E. Furman and R. Zitikis. A monotonicity property of the composition of regularized and inverted-regularized gamma functions with applications. *Journal of Mathematical Analysis and Applications*, 348(2):971–976, 2008.

[40] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer-Verlag New York, 2005.