Automatica 125 (2021) 109426

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

On a security vs privacy trade-off in interconnected dynamical systems[☆]

Vaibhav Katewa [a,*], Rajasekhar Anguluri [b], Fabio Pasqualetti [c]

[a] Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, 560012, India
[b] School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA
[c] Department of Mechanical Engineering, University of California Riverside, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 3 July 2019
Received in revised form 2 July 2020
Accepted 27 October 2020
Available online xxxx

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

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 with examples.

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 (for example, a smart city) (Rinaldi, Peerenboom, & Kelly, 2001). Although these subsystems are interconnected, it is usually difficult to directly measure the couplings and dependencies between them (Rinaldi et al., 2001). 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 (Cardenas, Amin, & Sastry, 2008). There have been many recent studies on attack analysis and possible remedial strategies (see Giraldo, Sarkar, Cardenas, Maniatakos and Kantarcioglu, 2017 and the references therein), where a key component 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. Yet, these measurements may contain confidential information and subsystem operators may be willing to share only limited information due to privacy concerns. In this paper, we study this trade-off between the attack detection performance (security) and the amount/quality of shared measurements (privacy).

Related work: Centralized attack detection and estimation schemes in dynamical systems have been studied in both deterministic (Fawzi, Tabuada, & Diggavi, 2014; Pasqualetti, Dörfler, & Bullo, 2013) and stochastic (Chen, Kar, & Moura, 2018; Mo & Sinopoli, 2016) 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 Cui, Han, Kar, Kim, Poor, and Tajer (2012) and Nishino and Ishii (2014). In Pasqualetti et al. (2013) and Pasqualetti, Dörfler, and Bullo (2015), centralized and decentralized monitor design was

[☆] This material is based upon the work supported in part by ARO, USA award 71603NSYIP and in part by UCOP, USA award LFR-18-548175. The material in this paper was partially presented at the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, November 12–15, 2018, Honolulu, Hawaii, USA. This paper was recommended for publication in revised form by Associate Editor Vijay Gupta under the direction of Editor Christos G. Cassandras.

[*] Corresponding author.
E-mail addresses: vkatewa@iisc.ac.in (V. Katewa), rangu003@engr.ucr.edu (R. Anguluri), fabiopas@engr.ucr.edu (F. Pasqualetti).

https://doi.org/10.1016/j.automatica.2020.109426
0005-1098/© 2020 Elsevier Ltd. All rights reserved.

presented for deterministic attack detection and identification. In Forti, Battistelli, Chisci, Li, Wang, and Sinopoli (2018) and Guan and Ge (2018), distributed strategies for joint attacks detection and state estimation are presented. Residual based tests (Boem, Gallo, Ferrari-Trecate, & Parisini, 2017) and unknown-input observer-based approaches (Teixeira, Sandberg, & Johansson, 2010) have also been proposed for attack detection. A comparison between centralized and decentralized detection schemes was presented in Anguluri, Katewa, and Pasqualetti (2020), where, differently from this work, local detectors use only local measurements.

Distributed fault detection techniques requiring information sharing among the subsystems have also been studied. For instance, in Ferrari, Parisian, and Polycarpou (2012), Yan and Edwards (2008) and Zhang and Zhang (2012) fault detection for non-linear 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 Franco, Olfati-Saber, Parisini, and Polycarpou (2006) and Stankovic, Ilic, Djurovic, Stankovic, and Johansson (2010) and unknown-input observer techniques in Shames, Teixeira, Sandberg, and Johansson (2011).

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 Cortes, Dullerud, Han, Le Ny, Mitra, & Pappas, 2016 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 (Akyol, Langbort, & Basar, 2015) and mutual information (Farokhi & Nair, 2016; Tanaka, Skoglund, Sandberg, & Johansson, 2017) have also been proposed. A privacy vs. cooperation trade-off for multi-agent systems was presented in Katewa, Pasqualetti, and Gupta (2018). In Mo and Murray (2017), a privacy mechanism for consensus was presented, where privacy is measured in terms of estimation error covariance of the initial state. The authors in Giraldo, Cardenas and Kantarcioglu (2017) 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 (Anguluri, Katewa, & Pasqualetti, 2018), 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 an intermediate 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 (Chen et al., 2018; Mo & Sinopoli, 2016), which we illustrate and explain.

Mathematical notation: $\mathrm{Tr}(\cdot)$, $\mathrm{Im}(\cdot)$, $\mathrm{Null}(\cdot)$ and $\mathrm{Rank}(\cdot)$ 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 $A > 0$ ($A \geq 0$). $\mathrm{diag}(A_1, A_2, \ldots, A_n)$ denotes a block diagonal matrix whose block diagonal elements are $A_1, A_2, \ldots, A_n$. The identity matrix is denoted by I. 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 $\mathcal{S} \triangleq \{1, 2, \ldots, N\}$ 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:

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

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

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 $B_i x_{-i}(k)$, where $x_{-i} \triangleq [x_1^\mathsf{T}, \ldots, x_{i-1}^\mathsf{T}, x_{i+1}^\mathsf{T}, \ldots, x_N^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{n-n_i}$ 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:

Assumption 1. Subsystem i has perfect knowledge of its dynamics, i.e., it knows (A_i, B_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 noises of the other subsystems. □

We consider the scenario where each subsystem can be under an attack. We model the attacks as external linear additive inputs to the subsystems, whose dynamics read as

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

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 Subsystem i, and is unknown to Subsystem i.

Each subsystem is equipped with an attack monitor. 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}.

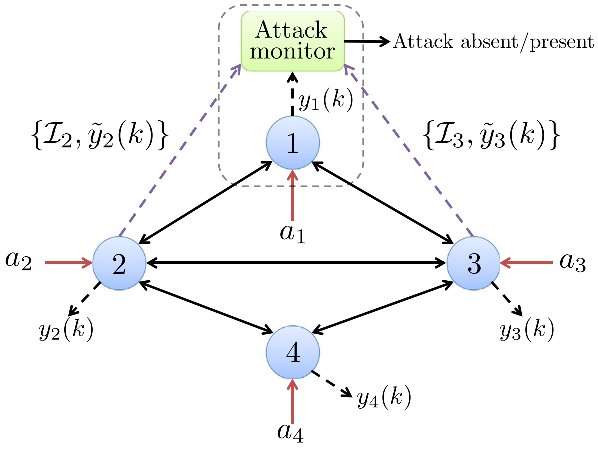


Fig. 1. An interconnected system consisting of $N = 4$ subsystems. The solid lines represent state coupling among the subsystems. For attack detection by Subsystem 1, its neighboring agents 2 and 3 communicate their output information to 1 (denoted by dashed lines). The attack monitor associated with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

Since the subsystems do not have knowledge of the interconnection signals (cf. Assumption 1), they share their measurements among each other to aid the local detection of attacks (see Fig. 1).

While the shared measurements help in detecting local attacks, they can also reveal sensitive information. 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. Instead of sharing the complete measurements in (2), Subsystem i shares limited measurements (denoted as $\tilde{y}_i$) given by:

$$\begin{aligned}\mathcal{M}_i : \qquad \tilde{y}_i(k) &= S_i y_i(k) + \tilde{r}_i(k)\\ &\overset{(2)}{=} S_i C_i x_i(k) + S_i v_i(k) + \tilde{r}_i(k),\end{aligned} \tag{4}$$

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 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, thereby protecting its privacy (a detailed explanation is in Section 4).

Let $\mathcal{I}_i \triangleq \{C_i, S_i, \Sigma_{v_i}, \Sigma_{\tilde{r}_i}\}$ denote the parameters corresponding to the limited measurements of subsystem i.

Assumption 2. Each subsystem $i \in \mathcal{S}$ shares its limited measurements $\tilde{y}_i$ in (4) and the parameters $\mathcal{I}_i$ with all neighboring 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). Further, we are interested in characterizing the trade-off between the privacy level and the detection performance.

3. Local attack detection

In this section we present the local attack detection procedure of the subsystems and characterize their performance. For the ease of presentation and without loss of generality, we describe the analysis for Subsystem 1 only.

3.1. Measurement collection

We employ a batch detection scheme in which each subsystem collects the measurements for $k = 1, 2, \ldots, T$, and performs detection based on the collective measurements.

Local measurements: Let the time-aggregated local measurements be denoted by $y_L \triangleq [y_1^{\mathsf{T}}(1), y_1^{\mathsf{T}}(2), \ldots, y_1^{\mathsf{T}}(T)]^{\mathsf{T}}$. Similarly, denote the time-aggregated ($k = 1$ to T) measurement noise, and time aggregated ($k = 0$ to $T-1$) interconnection signals, attacks, and process noise by $v, x, \tilde{a}$ and w, respectively. Let

$$F(Z) \triangleq \begin{bmatrix} C_1 Z & 0 & \cdots & 0\\ C_1A_1Z & C_1Z & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ C_1A_1^{T-1}Z & C_1A_1^{T-2}Z & \cdots & C_1Z\end{bmatrix} = F(I)(I_T \otimes Z). \tag{5}$$

By using (3) recursively and (2), we have

$$y_L = Ox_1(0) + F_x x + F_{\tilde{a}}\tilde{a} + F_w w + v, \tag{6}$$

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

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

Note that $w \sim \mathcal{N}(0, \Sigma_w)$ and $v \sim \mathcal{N}(0, \Sigma_v)$ with $\Sigma_w = I_T \otimes \Sigma_{w_1} > 0$ and $\Sigma_v = I_T \otimes \Sigma_{v_1} > 0$. Let $v_L \triangleq Ox_1(0) + F_w w + v$ denote the effective local noise in the measurement equation (6). Using the fact that $(x_1(0), w, v)$ are independent, the overall local measurements are given by

$$\begin{aligned}&y_L = F_x x + F_{\tilde{a}}\tilde{a} + v_L, \quad \text{where}\\ &v_L \sim \mathcal{N}(0, \Sigma_{v_L}),\ \Sigma_{v_L} = O\Sigma_{x_1(0)}O^{\mathsf{T}} + F_w\Sigma_w F_w^{\mathsf{T}} + \Sigma_v > 0.\end{aligned} \tag{7}$$

Shared measurements: Let the limited measurements received by Subsystem 1 from all the other subsystems at time k be denoted by $\tilde{y}_{-1}(k) \triangleq [\tilde{y}_2^{\mathsf{T}}(k), \tilde{y}_3^{\mathsf{T}}(k), \ldots, \tilde{y}_N^{\mathsf{T}}(k)]^{\mathsf{T}}$. 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 (4) we have

$$\begin{aligned}&\tilde{y}_{-1}(k) = S_{-1}C_{-1}x_{-1}(k) + S_{-1}v_{-1}(k) + \tilde{r}_{-1}(k),\\ &\text{where}\quad S_{-1} \triangleq \operatorname{diag}(S_2, \ldots, S_N),\ C_{-1} \triangleq \operatorname{diag}(C_2, \ldots, C_N),\\ &v_{-1}(k) \sim \mathcal{N}(0, \Sigma_{v_{-1}}),\ \Sigma_{v_{-1}} = \operatorname{diag}(\Sigma_{v_2}, \ldots, \Sigma_{v_N}) > 0,\\ &\tilde{r}_{-1}(k) \sim \mathcal{N}(0, \Sigma_{\tilde{r}_{-1}}),\ \Sigma_{\tilde{r}_{-1}} = \operatorname{diag}(\Sigma_{\tilde{r}_2}, \ldots, \Sigma_{\tilde{r}_N}) \geq 0.\end{aligned} \tag{8}$$

Further, let the time-aggregated limited measurements received by Subsystem 1 be denoted by $y_R \triangleq [\tilde{y}_{-1}^{\mathsf{T}}(0), \tilde{y}_{-1}^{\mathsf{T}}(1), \ldots, \tilde{y}_{-1}^{\mathsf{T}}(T-1)]^{\mathsf{T}}$, and let v_R denote similar time-aggregated vector of $\{S_{-1}v_{-1}(k) + \tilde{r}_{-1}(k)\}_{k=0,\ldots,T-1}$. Then, from (8), the overall limited measurements received by Subsystem 1 read as

$$\begin{aligned}&y_R = Hx + v_R, \quad \text{where}\\ &H \triangleq I_T \otimes S_{-1}C_{-1}, \quad \text{and} \quad v_R \sim \mathcal{N}(0, \Sigma_{v_R})\\ \text{with}\quad &\Sigma_{v_R} = I_T \otimes (S_{-1}\Sigma_{v_{-1}}S_{-1}^{\mathsf{T}} + \Sigma_{\tilde{r}_{-1}}) > 0.\end{aligned} \tag{9}$$

The goal of Subsystem 1 is to detect local attacks using the local and received measurements given by (7) and (9).

[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).

3.2. 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. Note that Subsystem 1 is oblivious to the statistics of the stochastic signal x. Therefore, it computes an estimate of x assuming that x is a deterministic but unknown quantity.

According to (9), $y_R \sim \mathcal{N}(Hx, \Sigma_{v_R})$, and the Maximum Likelihood (ML) estimate of x based on y_R is computed by maximizing the log-likelihood function of y_R:

$$\begin{aligned}\hat{x} &= \arg\max_{z} \quad -\frac{1}{2}(y_R - Hz)^{\mathsf{T}}\Sigma_{v_R}^{-1}(y_R - Hz)\\ &\overset{(a)}{=} \tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}y_R + (I - \tilde{H}^{+}\tilde{H})d, \quad \text{where}\\ \tilde{H} &\triangleq H^{\mathsf{T}}\Sigma_{v_R}^{-1}H \geq 0,\end{aligned} \tag{10}$$

d is any real vector of appropriate dimension, and equality (a) follows from Lemma A.1 in the Appendix. If $\tilde{H}$ (or equivalently H) is not full column rank, then the estimate can lie anywhere in $\mathrm{Null}(\tilde{H}) = \mathrm{Null}(H)$ (shifted by $\tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}y_R$). Thus, the component of x that lies in $\mathrm{Null}(H)$ cannot be estimated and only the component of x that lies in $\mathrm{Im}(\tilde{H}) = \mathrm{Im}(H^{\mathsf{T}})$ can be estimated. Based on this discussion, we decompose x as

$$\begin{aligned}x &= (I - \tilde{H}^{+}\tilde{H})x + \tilde{H}^{+}\tilde{H}x = (I - \tilde{H}^{+}\tilde{H})x + \tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}Hx\\ &\overset{(9)}{=} (I - \tilde{H}^{+}\tilde{H})x + \tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}(y_R - v_R).\end{aligned} \tag{11}$$

Substituting x from (11) in (7), we get

$$y_L = F_x(I - \tilde{H}^{+}\tilde{H})x + F_x\tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}(y_R - v_R) + F_{\tilde{a}}\tilde{a} + v_L. \tag{12}$$

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 (12) by M^{T}, where

$$\begin{aligned}&M = \text{Basis of Null}\left([F_x(I - \tilde{H}^{+}\tilde{H})]^{\mathsf{T}}\right),\\ &\quad \Rightarrow M^{\mathsf{T}}F_x(I - \tilde{H}^{+}\tilde{H}) = 0.\end{aligned} \tag{13}$$

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

$$\begin{aligned}z &= M^{\mathsf{T}}(y_L - F_x\tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}y_R)\\ &\overset{(12),(13)}{=} M^{\mathsf{T}}F_{\tilde{a}}\tilde{a} + \underbrace{M^{\mathsf{T}}(v_L - F_x\tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}v_R)}_{\triangleq\, v_P},\end{aligned} \tag{14}$$

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. Thus

$$\begin{aligned}\Sigma_{v_P} &= M^{\mathsf{T}}\left[\Sigma_{v_L} + F_x\tilde{H}^{+}H^{\mathsf{T}}\Sigma_{v_R}^{-1}\Sigma_{v_R}\Sigma_{v_R}^{-\mathsf{T}}H(\tilde{H}^{+})^{\mathsf{T}}F_x^{\mathsf{T}}\right]M\\ &\overset{\tilde{H}^{\mathsf{T}}=\tilde{H}}{=} M^{\mathsf{T}}\Sigma_{v_L}M + M^{\mathsf{T}}F_x\tilde{H}^{+}F_x^{\mathsf{T}}M \overset{(a)}{>} 0,\end{aligned} \tag{15}$$

where (a) follows from the facts that M is full column rank and $\Sigma_{v_L} > 0$. The processed measurements z in (14) depend only on the local attack $\tilde{a}$, and the Gaussian noise v_P whose statistics is known to Subsystem 1 (cf. 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.

Remark 1 (*Limitations of Measurement Processing*)**.** The operation of eliminating the unknown component $(I - \tilde{H}^{+}\tilde{H})x$ from y_L also eliminates a component of attack $\tilde{a}$. As a result, the space of undetectable attack vectors increases from $\mathrm{Null}(F_{\tilde{a}})$ to $\mathrm{Null}(M^{\mathsf{T}}F_{\tilde{a}})$. In some cases, this operation can also result in complete elimination of attacks, and attack detection is not possible (Katewa, Anguluri, & Pasqualetti, 2020). □

3.3. Statistical hypothesis testing

The goal of Subsystem 1 is to determine whether it is under attack using the measurements z in (14). Recall that, since Subsystem 1 does not know B_1^a, it can only detect $a_1 = B_1^a\tilde{a}_1$. Let $a \triangleq [(B_1^a\tilde{a}_1(0))^{\mathsf{T}}, \ldots, (B_1^a\tilde{a}_1(T-1))^{\mathsf{T}}]^{\mathsf{T}}$. Then, from (6), 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. Let H_0 and H_1 denote the hypotheses that the attack is absent $(a = 0)$ and present $(a = 1)$, respectively. We use the Generalized Likelihood Ratio Test (GLRT) criterion (Wasserman, 2004) for the above testing problem, which is given by

$$\frac{f(z|H_0)}{\sup\limits_{a} f(z|H_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \tau' \tag{16}$$

where $f(z|H_0)$ and $f(z|H_1)$ are the probability density functions of the multivariate Gaussian distribution of z under hypotheses H_0 and H_1, respectively, and τ' is a suitable threshold. Using the result in Lemma A.1 in the Appendix to compute the denominator in (16) and taking the logarithm, the test (16) can be equivalently written as

$$t(z) \triangleq z^{\mathsf{T}}\Sigma_{v_P}^{-1}M^{\mathsf{T}}F_a\tilde{M}^{+}F_a^{\mathsf{T}}M\Sigma_{v_P}^{-1}z \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{17}$$

$$\text{where} \quad \tilde{M} = F_a^{\mathsf{T}}M\Sigma_{v_P}^{-1}M^{\mathsf{T}}F_a,$$

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.

Lemma 3.1 (*Distribution of Test Statistics*)**.** *The distribution of test statistics $t(z)$ in (17) is given by*

$$t(z) \sim \chi_q^2 \quad \textit{under } H_0, \tag{18}$$

$$t(z) \sim \chi_q^2(\lambda \triangleq a^{\mathsf{T}}\Lambda a) \quad \textit{under } H_1, \tag{19}$$

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. Let $\Sigma_{v_P}^{-1} = R^{\mathsf{T}}R$ (R non-singular) denote the Cholesky decomposition of $\Sigma_{v_P}^{-1}$. Further, let the columns of U be orthonormal basis vectors of $\mathrm{Im}(RM^{\mathsf{T}}F_a)$. Then,

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

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, $t(z) = (z')^{\mathsf{T}}z' \sim \chi_q^2$.

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

$$\begin{aligned}&z' \sim \mathcal{N}(U^{\mathsf{T}}RM_1 a, I_q)\\ &\Rightarrow t(z) = (z')^{\mathsf{T}}z' \sim \chi_q^2(a^{\mathsf{T}}M_1^{\mathsf{T}}R^{\mathsf{T}}UU^{\mathsf{T}}RM_1 a).\end{aligned}$$

Since $RM_1(RM_1)^{+}$ is the orthogonal projection operator on $\mathrm{Im}(RM_1)$, we get $RM_1(RM_1)^{+} = UU^{\mathsf{T}}$. Therefore

$$\begin{aligned}a^{\mathsf{T}}M_1^{\mathsf{T}}R^{\mathsf{T}}UU^{\mathsf{T}}RM_1 a &= a^{\mathsf{T}}(RM_1)^{\mathsf{T}}(RM_1)(RM_1)^{+}(RM_1)a\\ &= a^{\mathsf{T}}(RM_1)^{\mathsf{T}}(RM_1)a = a^{\mathsf{T}}M_1^{\mathsf{T}}\Sigma_{v_P}^{-1}M_1 a = \lambda,\end{aligned}$$

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 (14). The parameter λ can be interpreted as the signal to noise ratio (SNR) of the processed measurements in (14), where the signal of interest is the attack. □

Next, we characterize the performance of the test (17). Let the probability of false alarm and probability of detection for the test be respectively denoted by

$$P_F = \text{Prob}(t(z) > \tau | H_0) \overset{(a)}{=} \mathcal{Q}_q(\tau) \quad \text{and,}$$
$$P_D = \text{Prob}(t(z) > \tau | H_1) \overset{(b)}{=} \mathcal{Q}_q(\tau; \lambda),$$

where (a) and (b) follow from (18) and (19), 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

$$\tau(q, P_F) = \mathcal{Q}_q^{-1}(P_F), \tag{20}$$
$$P_D(q, \lambda, P_F) = \mathcal{Q}_q(\tau(q, P_F); \lambda). \tag{21}$$

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. In order to compare the performance of two different tests, we select a common value of P_F and then compare the values of P_D.

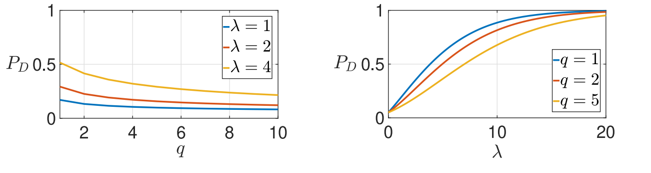
Lemma 3.2 (*Dependence of Detection Performance on Detection Parameters* (q, λ))**.** *For any given false alarm probability P_F, the detection probability $P_D(q, \lambda, P_F)$ is decreasing in q and increasing in λ.*

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 (Johnson, Kotz, & Balakrishnan, 1995). Thus, $P_D(q, \lambda)$ is increasing in λ. Next, we have (Johnson et al., 1995)

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

From Furman and Zitikis (2008, 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. ■

Fig. 2 illustrates the dependence of the detection probability on the parameters (q, λ). Lemma 3.2 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 (16) is not a uniformly most powerful (UMP) test for all values of the attack a. In fact, a UMP test does not exist in this case (Lehmann & Romano, 2005). 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 (16). It arises due to the composite nature of the test and the fact that the value of attack a is not known to the attack monitor.



(a) (b)

Fig. 2. 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 in (a), and it increases monotonically with λ in (b).

4. Privacy quantification

In this section, we quantify the privacy of the mechanism $\mathcal{M}_i$ in (4) in terms of the estimation error covariance of the state x_i. For simplicity, we assume $i \neq 1$, and this estimation is performed by Subsystem 1, which is directly coupled with Subsystem i and receives limited measurements from it. Then, we use the privacy 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 $k = 0, 1, \ldots, T-1$, with $T > 0$. Let $\tilde{y}_i = [\tilde{y}_i^\mathsf{T}(0), \ldots, \tilde{y}_i^\mathsf{T}(T-1)]^\mathsf{T}$, 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,

$$\tilde{y}_i = \underbrace{(I_T \otimes S_i C_i)}_{\triangleq H_i} x_i + \underbrace{(I_T \otimes S_i) v_i + \tilde{r}_i}_{\triangleq r_i}, \tag{22}$$

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 Subsystem 1 that receives measurements (22) from Subsystem i knows $\{H_i, \Sigma_{r_i}\}$ (cf. Assumption 2). However, it is oblivious to the statistics of the confidential stochastic signal x_i. Therefore, it computes an estimate of x_i assuming that x_i is a deterministic but unknown quantity. Further, this estimate is computed by Subsystem 1 using the measurements received only from Subsystem i, and it does not use its local measurements or the measurements received from other subsystems for this purpose.[2]

According to (22), $\tilde{y}_i \sim \mathcal{N}(H_i x_i, \Sigma_{r_i})$, and the ML estimate of x_i based on $\tilde{y}_i$ is:

$$\begin{aligned} \hat{x}_i &\overset{(a)}{=} \tilde{H}_i^+ H_i^\mathsf{T} \Sigma_{r_i}^{-1} \tilde{y}_i + (I - \tilde{H}_i^+ \tilde{H}_i) d_i, \quad \text{where} \\ \tilde{H}_i &\triangleq H_i^\mathsf{T} \Sigma_{r_i}^{-1} H_i \geq 0, \end{aligned} \tag{23}$$

d_i is any real vector of appropriate dimension, and equality (a) follows from Lemma A.1. 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 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 and its covariance in this subspace is:

$$\begin{aligned} e_i &= \mathcal{P}_i x_i - \mathcal{P}_i \hat{x}_i = \tilde{H}_i^+ \tilde{H}_i x_i - \tilde{H}_i^+ H_i^\mathsf{T} \Sigma_{r_i}^{-1} \tilde{y}_i \\ &= -\tilde{H}_i^+ H_i^\mathsf{T} \Sigma_{r_i}^{-1} r_i, \quad \text{and} \end{aligned} \tag{24}$$

$$\begin{aligned} \Sigma_{e_i} &= \mathbb{E}[\tilde{H}_i^+ H_i^\mathsf{T} \Sigma_{r_i}^{-1} r_i r_i^\mathsf{T} \Sigma_{r_i}^{-1} H_i \tilde{H}_i^+] \\ &= \tilde{H}_i^+ H_i^\mathsf{T} \Sigma_{r_i}^{-1} H_i \tilde{H}_i^+ = \tilde{H}_i^+. \end{aligned} \tag{25}$$

Note that since the model in (22) is linear with Gaussian noise, $\mathcal{P}_i \hat{x}_i$ is the minimum-variance unbiased (MVU) estimate of x_i

[2] There are two reasons: (i) unknown attacks a which cannot be eliminated without eliminating x (and thus, x_i), and (ii) unknown dynamics and attacks of other subsystems.

projected on $\mathrm{Im}(H_i^\mathsf{T})$. Thus, the covariance Σ_{e_i} captures the fundamental limit on how accurately $\mathcal{P}_i x_i$ can be estimated and is a suitable metric to quantify privacy.

The privacy level of mechanism $\mathcal{M}_i$ in (4) is characterized by two quantities: (i) $\mathrm{rank}(S_i)$, and (ii) Σ_{e_i}. Intuitively, if $\mathrm{rank}(S_i)$ 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 $\mathrm{rank}(S_i)$ 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. 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

$$\begin{aligned}&(i)\ \mathrm{Im}\left((S_i^{(2)})^\mathsf{T}\right) \subseteq \mathrm{Im}\left((S_i^{(1)})^\mathsf{T}\right) \quad \text{and,}\\ &(ii)\ \Sigma_{e_i}^{(2)} \geq \mathcal{P}_i^{(2)}\Sigma_{e_i}^{(1)}\mathcal{P}_i^{(2)}. \quad \square\end{aligned} \tag{26}$$

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, $\mathrm{Im}(H_i^{(2)}) = \mathrm{Im}(\mathcal{P}_i^{(2)}) \subseteq \mathrm{Im}(H_i^{(1)}) = \mathrm{Im}(\mathcal{P}_i^{(1)})$. 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 $\mathrm{Im}(\mathcal{P}_i^{(2)})$, 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 (26).

Example 1. Let $x_i \in \mathbb{R}^2$, $C_i = I_2$, $T = 1$, and let

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

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 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 (23)) are given by $\hat{x}_i^{(1)} = \tilde{y}_i^{(1)}$ and $\hat{x}_i^{(2)} = \begin{bmatrix}1\\0\end{bmatrix}\tilde{y}_i^{(2)} + \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}d_i$. 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 error covariances 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 $\alpha \geq 1$.

If $\alpha < 1$, then an ordering between the mechanisms cannot be established. Under $\mathcal{M}_i^{(1)}$, both the state components can be estimated but the estimation error in first component is large. Under $\mathcal{M}_i^{(2)}$, only the first component can be estimated but its estimation error is small. □

An important property of the privacy mechanism in (4) is that it exhibits an intuitive post-processing property. It implies that further limiting the measurements produced by a privacy mechanism cannot decrease the privacy of the measurements (see also Katewa et al., 2020). This post-processing property also holds in the differential privacy framework (Cortes et al., 2016).

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 $j \in \mathcal{S}_{-1}$. Note that the trade-off is between the *detection performance of Subsystem 1* and the *privacy level of all other subsystems.*

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*

$$\begin{aligned}&(i)\ q^{(1)} \geq q^{(2)} \quad \text{and,}\\ &(ii)\ \lambda^{(2)}\mu_{max} \geq \lambda^{(1)} \geq \lambda^{(2)}\mu_{min} \geq \lambda^{(2)},\end{aligned} \tag{27}$$

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

Proof. From (4), (8) and (9), for $k = 1, 2$, we have

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

Since $\mathcal{M}_j^{(2)} \geq \mathcal{M}_j^{(1)}$ for all $j \in \mathcal{S}_{-1}$, conditions (26) imply

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

From (10), we have $\tilde{H}^{(k)} = (H^{(k)})^\mathsf{T}(\Sigma_{v_R}^{(k)})^{-1}H^{(k)}$. Since $\mathrm{Null}(\tilde{H}^{(k)}) = \mathrm{Null}(H^{(k)})$, from (13), it follows that
$\mathrm{Im}(M^{(1)}) \supseteq \mathrm{Im}(M^{(2)})$. Recalling from (19) that $q^{(k)} = \mathrm{Rank}((M^{(k)})^\mathsf{T}F_a)$, it follows that $q^{(1)} \geq q^{(2)}$.

Since $\mathrm{Im}(M^{(1)}) \supseteq \mathrm{Im}(M^{(2)})$, $M^{(2)} = M^{(1)}P$ for some full column rank matrix P. Let $Z \triangleq F_x^\mathsf{T}M^{(1)}P$. From (15):

$$\begin{aligned}\Sigma_{v_P}^{(2)} &= (M^{(2)})^\mathsf{T}\Sigma_{v_L}M^{(2)} + (M^{(2)})^\mathsf{T}F_x(\tilde{H}^{(2)})^+F_x^\mathsf{T}M^{(2)},\\ &= P^\mathsf{T}\Sigma_{v_P}^{(1)}P + \underbrace{Z^\mathsf{T}[(\tilde{H}^{(2)})^+ - (\tilde{H}^{(1)})^+]Z}_{\triangleq E}.\end{aligned} \tag{28}$$

Next, we show that $E \geq 0$. Using $M^{(2)} = M^{(1)}P$, and using (13) for both $\{M^{(k)}, \tilde{H}^{(k)}\}$, $k = 1, 2$, we have

$$Z^\mathsf{T}(\tilde{H}^{(1)})^+\tilde{H}^{(1)} = Z^\mathsf{T}(\tilde{H}^{(2)})^+\tilde{H}^{(2)}, \quad \text{and thus} \tag{29}$$

$$\begin{aligned}E &= Z^\mathsf{T}[(\tilde{H}^{(2)})^+ - (\tilde{H}^{(1)})^+\tilde{H}^{(1)}(\tilde{H}^{(1)})^+\tilde{H}^{(1)}(\tilde{H}^{(1)})^+]Z\\ &= Z^\mathsf{T}[(\tilde{H}^{(2)})^+ - (\tilde{H}^{(2)})^+\tilde{H}^{(2)}(\tilde{H}^{(1)})^+(\tilde{H}^{(2)})^+\tilde{H}^{(2)}]Z\end{aligned} \tag{30}$$

where the last inequality follows from (29) and the fact that $\tilde{H}^{(k)}(\tilde{H}^{(k)})^+ = (\tilde{H}^{(k)})^+\tilde{H}^{(k)}$. Next, we have

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

$$\dots, I_T \otimes (S_N^{(k)} C_N)^\mathsf{T} \left(S_N^{(k)} \Sigma_{v_N} (S_N^{(k)})^\mathsf{T} + \Sigma_{\tilde{r}_N}^{(k)} \right)^{-1} S_N^{(k)} C_N \Big] \Pi$$

$$= \Pi^\mathsf{T} \mathrm{diag} \left[\tilde{H}_2^{(k)}, \dots, \tilde{H}_N^{(k)} \right] \Pi \quad \text{and,} \tag{31a}$$

$$(\tilde{H}^{(k)})^+ = \Pi^\mathsf{T} \mathrm{diag} \left[(\tilde{H}_2^{(k)})^+, \dots, (\tilde{H}_N^{(k)})^+ \right] \Pi, \tag{31b}$$

where Π is a permutation matrix with $\Pi^{-1} = \Pi^\mathsf{T}$. Substituting (31a) and (31b) in (30), we have

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

where (a) follows from the second condition in (26) for all $j \in \mathcal{S}_{-1}$. Next, from (19), we have,

$$\begin{aligned} \Lambda^{(2)} &= F_a^\mathsf{T} M^{(2)} (\Sigma_{vp}^{(2)})^{-1} (M^{(2)})^\mathsf{T} F_a \\ &\overset{(28)}{=} F_a^\mathsf{T} M^{(1)} P (P^\mathsf{T} \Sigma_{vp}^{(1)} P + E)^{-1} P^\mathsf{T} (M^{(1)})^\mathsf{T} F_a \\ &\overset{(b)}{\leq} F_a^\mathsf{T} M^{(1)} (\Sigma_{vp}^{(1)})^{-1} (M^{(1)})^\mathsf{T} F_a = \Lambda^{(1)}, \end{aligned}$$
$$\Rightarrow \lambda^{(1)} = a^\mathsf{T} \Lambda^{(1)} a \geq a^\mathsf{T} \Lambda^{(2)} a = \lambda^{(2)},$$

where (b) follows from: (i) Lemma A.2, (ii) $E \geq 0$, and (iii) P is full column rank. Finally, the second condition in (27) follows from Lemma A.3, and 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 3.2 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[3] ($\lambda^{(1)} \to \lambda^{(2)}$) is larger than the increase in P_D due to the decrease in the number of measurements ($q^{(1)} \to q^{(2)}$), then the detection performance decreases.

Theorem 5.2 (*Less Privacy Does Not Always Guarantee More Security*)**.** *Consider the setup in Theorem 5.1 with $\mathcal{M}_j^{(1)}$ less private than $\mathcal{M}_j^{(2)}$ for all $j \in \mathcal{S}_{-1}$. Let the detection probability of Subsystem 1 be as in (21). Then, given P_F, $P_D(q^{(1)}, \lambda^{(1)}, P_F) \geq P_D(q^{(2)}, \lambda^{(2)}, P_F)$ may not hold.*

This is an interesting and counter-intuitive trade-off between the detection performance and privacy/information sharing. It implies that 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 show that this counter-intuitive phenomenon does not occur if the subspace of the measurements shared by the privacy mechanisms is fixed, and the privacy level is varied with the noise level. In this case, a strict trade-off between privacy and detection performance exists.

Corollary 5.3 (*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, $q^{(1)} = q^{(2)}$ and $\lambda^{(1)} \geq \lambda^{(2)}$ follow from the proof of Theorem 5.1. The result then follows from Lemma 3.2. ■

6. Simulation example

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

$$A_1 = \frac{1}{3} \begin{bmatrix} -1 & -16 & 2 & -4 \\ 0 & -6 & 1 & -1 \\ 0 & 2 & 1 & 1 \\ 1 & 28 & -3 & 6 \end{bmatrix}, A_{12} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix},$$
$$A_{13} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, C_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$
$$A_{-1} = \begin{bmatrix} A_{12} & A_{13} \end{bmatrix},\ \Sigma_{x_1(0)} = 0.2 I_4,\ \Sigma_{w_1} = 0.1 I_4,\ C_2 = I_3,$$
$$C_3 = I_2,\ \Sigma_{v_1} = \Sigma_{v_2} = I_3,\ \Sigma_{v_3} = I_2,\ T = 2.$$

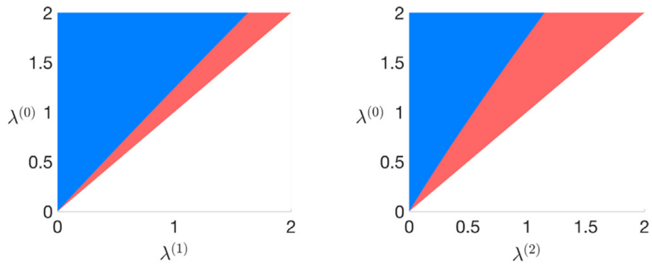
We focus on attack detection for Subsystem 1, where Subsystems 2 and 3 use privacy mechanisms to share their measurements with Subsystem 1. We consider the following three cases for Subsystems 2 and 3:

- $\mathcal{M}^{(0)} = \{\mathcal{M}_2^{(0)}, \mathcal{M}_3^{(0)}\}$: Subsystems 2 and 3 do not use any privacy mechanisms and share actual measurements, i.e., $S_2 = I_3, S_3 = I_2, \Sigma_{\tilde{r}_2} = 0, \Sigma_{\tilde{r}_3} = 0$.
- $\mathcal{M}^{(1)}$: $S_2 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right], S_3 = I_2, \Sigma_{\tilde{r}_2} = 0, \Sigma_{\tilde{r}_3} = I_2$.
- $\mathcal{M}^{(2)}$: $S_2 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right], S_3 = \left[\begin{smallmatrix} 0 & 1 \end{smallmatrix}\right], \Sigma_{\tilde{r}_2} = 0, \Sigma_{\tilde{r}_3} = 1.8$.

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 (19), 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. 3 presents the performance comparison of case 0 with case 1 (Fig. 3(a)) and case 2 (Fig. 3(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 (27)). 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, $k = 1, 2$.

[3] The SNR depends upon the attack vector a (via (19)), which we do not know a-priori. Thus, the SNR can take any positive value.



(a) (b)

Fig. 3. 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 $k = 1, 2$ (cf. Theorem 5.1), the white region is inadmissible. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

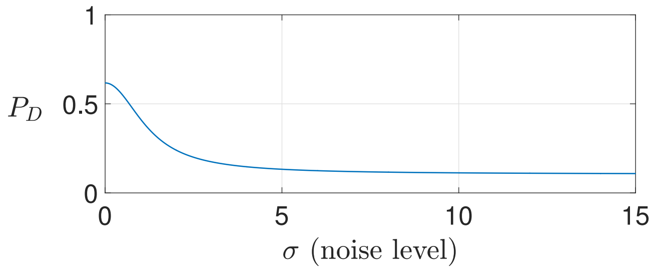


Fig. 4. Detection performance for varying level of noise.

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 a 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. Further, we observe that the red region of Fig. 3(b) is larger than (and contains) the red region of Fig. 3(b) 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 (4). 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 $\Sigma_{\tilde{r}_3} = \sigma^2 I_2$. 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. 4 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 (cf. Corollary 5.3), implying a strict trade-off between detection performance and privacy in this case.

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 exists for alternative detection schemes (for ex., unknown-input observers) and for other types of statistical tests. Also, privacy ordering of two mechanisms irrespective of their subspaces of shared measurement will be defined using suitable weighing matrix for each subspace.

Appendix. Auxiliary Results (for proofs, see Katewa et al. (2020))

Lemma A.1. *The optimal solutions of the weighted least squares problem:* $\min_x \; J(x) = (y - Hx)^{\mathsf{T}} \Sigma^{-1} (y - Hx)$*, with* $\Sigma > 0$ *are given by*

$$x^* = \tilde{H}^+ H^{\mathsf{T}} \Sigma^{-1} y + (I - \tilde{H}^+ \tilde{H}) d, \tag{32}$$

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

$$J(x^*) = y^{\mathsf{T}} (\Sigma^{-1} - \Sigma^{-1} H \tilde{H}^+ H^{\mathsf{T}} \Sigma^{-1}) y. \tag{33}$$

Lemma A.2. *Let* $\Sigma > 0 \in \mathbb{R}^{n \times n}$*,* $\Sigma_a \geq 0 \in \mathbb{R}^{m \times m}$*, with* $m \leq n$*, and let* $S \in \mathbb{R}^{n \times m}$ *be full (column) rank. Then,*

$$\Sigma^{-1} \geq S(S^{\mathsf{T}} \Sigma S + \Sigma_a)^{-1} S^{\mathsf{T}}. \tag{34}$$

Lemma A.3. *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.*

References

Akyol, E., Langbort, C., & Basar, T. (2015). Privacy constrained information processing. In *IEEE conf. on decision and control* Osaka, Japan.

Anguluri, R., Katewa, V., & Pasqualetti, F. (2018). 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.

Anguluri, R., Katewa, V., & Pasqualetti, F. (2020). Centralized versus decentralized detection of attacks in stochastic interconnected systems. *IEEE Transactions on Automatic Control*, *65*(9), 3903–3910.

Boem, F., Gallo, A. J., Ferrari-Trecate, G., & Parisini, T. (2017). A distributed attack detection method for multi-agent systems governed by consensus-based control. In *IEEE conf. on decision and control* Melbourne, Australia (pp. 5961–5966).

Cardenas, A., Amin, S., & Sastry, S. (2008). Secure control: Towards survivable cyber-physical systems. In *International Conference on Distributed Computing Systems Workshops* Beijing, China (pp. 495—500).

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

Cortes, J., Dullerud, G. E., Han, S., Le Ny, J., Mitra, S., & Pappas, G. J. (2016). Differential privacy in control and network systems. In *IEEE Conf. on Decision and Control* (pp. 4252–4272).

Cui, S., Han, Z., Kar, S., Kim, T. T., Poor, H. V., & Tajer, A. (2012). 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.

Farokhi, F., & Nair, G. (2016). Privacy-constrained communication. In *IFAC workshop on distributed estimation and control in networked systems* Tokyo, Japan, (pp. 43–48).

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

Ferrari, R. M. G., Parisian, T., & Polycarpou, M. M. (2012). 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.

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

Franco, E., Olfati-Saber, R., Parisini, T., & Polycarpou, M. M. (2006). Distributed fault diagnosis using sensor networks and consensus-based filters. In *IEEE Conf. on decision and control* San Diego, CA, USA, (pp. 386–391).

Furman, E., & Zitikis, R. (2008). 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.

Giraldo, J., Cardenas, A., & Kantarcioglu, M. (2017). Security and privacy trade-offs in CPS by leveraging inherent differential privacy. In *IEEE conference on control technology and applications* Hawaii, USA, (pp. 1313–1318).

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

Guan, Y., & Ge, X. (2018). 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.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2*. Wiley-Interscience.

Katewa, V., Anguluri, R., & Pasqualetti, F. (2020). On a security vs privacy trade-off in interconnected dynamical systems. https://arxiv.org/abs/2006.13416, arXiv:2006.13416.

Katewa, V., Pasqualetti, F., & Gupta, V. (2018). On privacy vs cooperation in multi-agent systems. *International Journal of Control*, *91*(7), 1693–1707. http://dx.doi.org/10.1080/00207179.2017.1326632.

Lehmann, E. L., & Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer-Verlag New York.

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

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

Nishino, H., & Ishii, H. (2014). Distributed detection of cyber attacks and faults for power systems. In *IFAC World Congress* Cape Town, South Africa (pp. 11932–11937).

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

Pasqualetti, F., Dörfler, F., & Bullo, F. (2015). A divide-and-conquer approach to distributed attack identification. In *IEEE Conf. on Decision and Control* (pp. 5801–5807).

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

Shames, I., Teixeira, A. M. H., Sandberg, H., & Johansson, K. H. (2011). Distributed fault detection for interconnected second-order systems. *Automatica*, *47*, 2757–2764.

Stankovic, S., Ilic, N., Djurovic, Z., Stankovic, M., & Johansson, K. H. (2010). Consensus based overlapping decentralized fault detection and isolation. In *Conference on control and fault tolerant systems* Nice, France.

Tanaka, T., Skoglund, M., Sandberg, H., & Johansson, K. H. (2017). Directed information and privacy loss in cloud-based control. In *American control conference* Seattle, USA.

Teixeira, A., Sandberg, H., & Johansson, K. H. (2010). Networked control systems under cyber attacks with applications to power networks. In *American Control Conference*.

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

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

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

Vaibhav Katewa is an Assistant Professor in the Department of Electrical Communication Engineering and a full-time associate faculty member of the Robert Bosch Center for Cyber–Physical Systems at the Indian Institute of Science, Bangalore. He was a Postdoctoral Scholar in the department of Mechanical Engineering at the University of California, Riverside from 2017 to 2019. He received his M.S. and Ph.D. degrees from University of Notre Dame in 2012 and 2016, and his Bachelor's degree from Indian Institute of Technology, Kanpur in 2007, all in Electrical Engineering. His research interests include analysis and design of security and privacy methods for cyber–physical systems and complex networks, decentralized and sparse feedback control, and protocol design for networked control systems.

Rajasekhar Anguluri received the B.Tech. degree in electrical engineering from the National Institute of Technology Warangal, India, in 2013, and both the M.S. degree in statistics and the Ph.D. degree in mechanical engineering from the University of California at Riverside, CA, USA, in 2019. He is currently a postdoctoral research scholar with the School of Electrical, Computer, and Energy Engineering at Arizona State University, Tempe, AZ, USA. His current research interests include statistical signal processing, stochastic control, and security of cyber–physical systems

Fabio Pasqualetti (SM'07, M'13) is an Associate Professor in the Department of Mechanical Engineering, University of California, Riverside. He completed a Doctor of Philosophy degree in Mechanical Engineering at the University of California, Santa Barbara, in 2012, a Laurea Magistrale degree (M.Sc. equivalent) in Automation Engineering at the University of Pisa, Italy, in 2007, and a Laurea degree (B.Sc. equivalent) in Computer Engineering at the University of Pisa, Italy, in 2004. His main research interests are in the areas of security for cyber–physical systems, distributed systems and networks, computational neuroscience, optimization, and robotic patrolling and persistent surveillance.