Centralized Versus Decentralized Detection of Attacks in Stochastic Interconnected Systems

Rajasekhar Anguluri, *Student Member, IEEE*, Vaibhav Katewa, *Member, IEEE*, and Fabio Pasqualetti *Member, IEEE*

***Abstract*— We consider a security problem for interconnected systems governed by linear, discrete, time-invariant, stochastic dynamics, where the objective is to detect exogenous attacks by processing the measurements at different locations. We consider two classes of detectors, namely centralized and decentralized detectors, which differ primarily in their knowledge of the system model. In particular, a decentralized detector has a model of the dynamics of the isolated subsystems, but is unaware of the interconnection signals that are exchanged among subsystems. Instead, a centralized detector has a model of the entire dynamical system. We characterize the performance of the two detectors and show that, depending on the system and attack parameters, each of the detectors can outperform the other. In particular, it may be possible for the decentralized detector to outperform its centralized counterpart, despite having less information about the system dynamics, and this surprising property is due to the nature of the considered attack detection problem. To complement our results on the detection of attacks, we propose and solve an optimization problem to design attacks that maximally degrade the system performance while maintaining a pre-specified degree of detectability. Finally, we validate our findings via numerical studies on an electric power system.**

I. Introduction

Cyber-physical systems are becoming increasingly more complex and interconnected. In fact, different cyber-physical systems typically operate in a connected environment, where the performance of each system is greatly affected by neighboring units. An example is the smart grid, which arises from the interconnection of smaller power systems at different geographical locations, and whose performance depends on other critical infrastructures including the transportation network and the water system. Given the interconnected nature of large cyber-physical systems, and the fact that each subsystem usually has only partial knowledge or measurements of other interconnected units, the security question arises as to whether sophisticated attackers can hide their action to the individual subsystems while inducing system-wide critical perturbations.

In this work we investigate whether, and to what extent, coordination among different subsystems and knowledge of the global system dynamics is necessary to detect attacks in interconnected systems. In fact, while existing approaches for the detection of faults and attacks typically rely on a centralized detector [1]–[3], the use of local detectors would not only be computationally convenient, but it would also prevent the subsystems from disclosing private information about their plants. As a counterintuitive result, we will show that local and decentralized detectors can, in some cases, outperform a centralized detector, thus supporting the development of distributed and localized theories and tools for the security of cyber-physical systems.

Related work: Centralized attack detectors have been the subject of extensive research in the last years [4]–[12], where the detector has complete knowledge of the system dynamics and all measurements. Furthermore, these studies use techniques from various disciplines including game theory, information theory, fault detection and signal processing, and have a wide variety of applications [2]. Instead, decentralized attack detectors, where each local detector decides on attacks based on partial information and measurements about the system, and local detectors cooperate to improve their detection capabilities, have received only limited and recent attention [13]–[17].

Decentralized detection schemes have also been studied for fault detection and isolation (FDI). In such schemes, multiple local detectors make inferences about either the global or local process, and transmit their local decisions to a central entity, which uses appropriate fusion rules to make the global decision [18]–[22]. Methods to improve the detection performance by exchanging information among the local detectors have also been proposed [23]–[25]. These decentralized algorithms are typically complex [1], their effectiveness in detecting unknown and unmeasurable attacks is difficult to characterize, and their performance is believed to be inferior when compared to their centralized counterparts. To the best of our knowledge, a rigorous comparison of centralized and decentralized attack detection schemes is still lacking, which prevents us from assessing whether, and to what extent, decentralized and distributed schemes should be employed for attack detection and identification.

Main contributions:[1] This paper features three main contributions. First, we propose centralized and decentralized schemes to detect unknown and unmeasurable sensor attacks

This material is based upon work supported by the awards ARO 71603NSYIP and NSF ECCS1405330, and by the UC-LAB Center for Electricity Distribution Cybersecurity. The authors are with the Department of Mechanical Engineering, University of California, Riverside, {rangu003,vkatewa,fabiopas}@engr.ucr.edu.

[1] In a preliminary version of this paper [26], we used asymptotic approximations to compare the detectors performance. Instead, in this paper we provide stronger, tight, and non-asymptotic results without using any approximations. Further, it contains new results on the design of optimal undetectable attacks, and a characterization of the performance degradation induced by such attacks. In addition, an illustration of the results using electrical power grid is also presented.

in stochastic interconnected systems. Our detection schemes are based on the statistical decision theoretic framework that falls under the category of simple versus composite hypotheses testing. We characterize the probability of false alarm and the probability of detection for both detectors, as a function of the system and attack parameters. Second, we compare the performance of the centralized and decentralized detectors, and show that each detector can outperform the other for certain system and attack configurations. We discuss that this counterintuitive phenomenon is inherent with the simple versus composite nature of the considered attack detection problem, and provide numerical examples of this behavior. Third, we formulate and solve an optimization problem to design attacks against interconnected systems that maximally affect the system performance as measured by the mean square deviation of the state while remaining undetected by the centralized and decentralized detectors with a pre-selected probability. Finally, we validate our theoretical findings on the IEEE RTS-96 power system model.

Paper organization: The rest of the paper is organized as follows. Section II contains our problem formulation. In Section III, we present our local, decentralized, and centralize detectors, and characterize their performance. Section IV contains our main results regarding the comparison of the performance of centralized and decentralized detectors. Section V contains the design of optimal undetectable attacks. Finally, Section VI contains our numerical studies, and Section VII concludes the paper.

Mathematical notation: The following notation will be adopted throughout the paper. Let $X_1, \dots, X_N$ be arbitrary sets, then $\bigcup_{i=1}^N X_i$ and $\bigcap_{i=1}^N X_i$ denotes the union and intersection of the sets, respectively. $\mathrm{Trace}(\cdot)$, $\mathrm{Rank}(\cdot)$, and $\mathrm{Null}(\cdot)$ denote the trace, rank, and null space of a matrix, respectively. $Q > 0$ ($Q \geq 0$) denotes that Q is a positive definite (positive semi definite) matrix. $\otimes$ denotes the Kronecker product for matrices. $\mathrm{blkdiag}(A_1, \cdots, A_N)$ denotes the block diagonal matrix with $A_1, \cdots, A_N$ as diagonal entries. The identity matrix is denoted by I (or I_{dim} to denote dimension explicitly). $\Pr[\mathcal{E}]$ denotes the probability of the event $\mathcal{E}$. The mean and covariance of a real or vector valued random variable Y is denoted by $\mathbb{E}[Y]$ and $\mathrm{Cov}[Y]$. Further, for a real valued random variable Y, we denote the standard deviation as $\mathrm{SD}[Y]$. If Y follows a Gaussian distribution, we denote it by $Y \sim \mathcal{N}(\mathbb{E}[Y], \mathrm{Cov}[Y])$. Instead, if Y follows a noncentral chi-squared distribution, we denote it by $Y \sim \chi^2(p, \lambda)$, where p is the degrees of freedom and λ is the non-centrality parameter. For $Y \sim \chi^2(p, \lambda)$ and $\tau \geq 0$, $Q(\tau; p, \lambda)$ denotes the complementary cumulative distribution function of Y.

II. Problem Setup and Preliminary Notions

We consider an interconnected system with N subsystems, where each subsystem obeys the discrete-time linear dynamics

$$\begin{aligned} x_i(k+1) &= A_{ii}x_i(k) + B_i u_i(k) + w_i(k), \\ y_i(k) &= C_i x_i(k) + v_i(k), \end{aligned} \tag{1}$$

with $i \in \{1, \dots, N\}$. In the above equation, the vectors $x_i \in \mathbb{R}^{n_i}$ and $y_i \in \mathbb{R}^{r_i}$ are the state and measurement of the $i-$th subsystem, respectively. The process noise $w_i(k) \sim \mathcal{N}(0, \Sigma_{w_i})$ and the measurement noise $v_i(k) \sim \mathcal{N}(0, \Sigma_{v_i})$ are independent stochastic processes, and w_i is assumed to be independent of v_i, for all $k \geq 0$. Further, the noise vectors across different subsystems are assumed to be independent at all times. The $i-$th subsystem is coupled with the other subsystems through the term $B_i u_i$, which takes the form

$$\begin{aligned} B_i &= \begin{bmatrix} A_{i1} & \cdots & A_{i,i-1} & A_{i,i+1} & \cdots & A_{iN} \end{bmatrix}, \text{ and} \\ u_i &= \begin{bmatrix} x_1^\mathsf{T} & \cdots & x_{i-1}^\mathsf{T} & x_{i+1}^\mathsf{T} & \cdots & x_N^\mathsf{T} \end{bmatrix}^\mathsf{T}. \end{aligned}$$

The input $B_i u_i = \sum_{j \neq i}^N A_{ij} x_j$ represents the cumulative effect of subsystems j on subsystem i. Hence, we refer to B_i as to the interconnection matrix, and to u_i as to the interconnection signal, respectively.

We allow for the presence of attacks compromising the dynamics of the subsystems, and model such attacks as exogenous unknown inputs. In particular, the dynamics of the $i-$th subsystem under the attack u_i^a with matrix B_i^a read as

$$x_i(k+1) = A_{ii}x_i(k) + B_i u_i(k) + B_i^a u_i^a(k) + w_i(k), \tag{2}$$

where $u_i^a \in \mathbb{R}^{m_i}$. In vector form, the dynamics of the interconnected system under attack read as

$$\begin{aligned} x(k+1) &= Ax(k) + B^a u^a(k) + w(k), \\ y(k) &= Cx(k) + v(k), \end{aligned} \tag{3}$$

where $\phi = \begin{bmatrix} \phi_1^\mathsf{T} & \dots & \phi_N^\mathsf{T} \end{bmatrix}$, with ϕ standing for $x \in \mathbb{R}^n$, $w \in \mathbb{R}^n$, $u^a \in \mathbb{R}^m$, $y \in \mathbb{R}^r$, $v \in \mathbb{R}^r$, $n = \sum_{i=1}^N n_i$, $m = \sum_{i=1}^N m_i$, and $r = \sum_i^N r_i$. Moreover, as the components of the vectors w and v are independent and Gaussian, $w \sim \mathcal{N}(0, \Sigma_w)$ and $v \sim \mathcal{N}(0, \Sigma_v)$, respectively, where $\Sigma_w = \mathrm{blkdiag}\left(\Sigma_{w_1}, \dots, \Sigma_{w_N}\right)$ and $\Sigma_v = \mathrm{blkdiag}\left(\Sigma_{v_1}, \dots, \Sigma_{v_N}\right)$. Further,

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{bmatrix}, B^a = \begin{bmatrix} B_1^a & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_N^a \end{bmatrix},$$

and $C = \mathrm{blkdiag}\left(C_1, \dots, C_N\right)$.

We assume that each subsystem is equipped with a *local detector*, which uses the local measurements and knowledge of the local dynamics to detect the presence of local attacks. In particular, the $i-$th local detector has access to the measurements y_i in (1), knows the matrices A_{ii}, B_i, and C_i, and the statistical properties of the noise vectors w_i and v_i. Yet, the $i-$th local detector does not know or measure the interconnection input u_i, and the attack parameters B_i^a and u_i^a. Based on this information, the $i-$th local detector aims to detect whether $B_i^a u_i^a \neq 0$. The decisions of the local detectors are then processed by a *decentralized detector*, which aims to detect the presence of attacks against the whole interconnected system based on the local decisions. Finally, we assume the presence of a *centralized detector*, which has access to the measurements y in (3), and knows the matrix A and the statistical properties of the overall noise vectors w and v. Similarly to the local detectors, the centralized detector does not know or measure the attack parameters B^a and u^a, and aims to detect whether $B^a u^a \neq 0$. We postpone a detailed description of our detectors to Section III. To conclude this

section, note that the decentralized and centralized detectors have access to the same measurements. Yet, these detectors differ in their knowledge of the system dynamics, which determines their performance as explained in Section IV.

Remark 1: ***(Control input and initial state)*** The system setup in (2) and (3) typically includes a control input. However, assuming that each subsystem knows its control input, it can be omitted without affecting generality. Further, as the detectors do not have information about the initial state, we assume without loss of generality, that the initial state is deterministic and unknown to the detectors. □

III. Local, Decentralized, and Centralized Detectors

In this section we formally describe our local, decentralized, and centralized detectors, and characterize their performance as a function of the available measurements and knowledge of the system dynamics. To this aim, let $T > 0$ be an arbitrary time horizon and define the vectors

$$Y_i = \begin{bmatrix} y_i^{\mathsf{T}}(1) & y_i^{\mathsf{T}}(2) & \cdots & y_i^{\mathsf{T}}(T) \end{bmatrix}^{\mathsf{T}}, \tag{4}$$

which contains the measurements available to the $i-$th detector, and

$$Y_c = \begin{bmatrix} y^{\mathsf{T}}(1) & y^{\mathsf{T}}(2) & \cdots & y^{\mathsf{T}}(T) \end{bmatrix}^{\mathsf{T}}, \tag{5}$$

which contains the measurements available to the centralized detector. Both the local and centralized detectors perform the following three operations in order:

1) Collect measurements as in (4) and (5), respectively;
2) Process measurements to filter unknown variables; and
3) Perform statistical hypotheses testing to detect attacks (locally or globally) using the processed measurements.

The decisions of the local detectors are then used by the decentralized detector, which triggers an alarm if any of the local detectors does so. We next characterize how the detectors process their measurements and perform attack detection via statistical hypothesis testing.

A. Processing of measurements

The measurements (4) and (5) depend on parameters that are unknown to the detectors, namely, the system initial state and the interconnection signal (although the process and measurement noises are also unknown, the detectors know their statistical properties). Thus, to test for the presence of attacks, the detectors first process the measurement vectors to eliminate their dependency on the unknown parameters. To do so, using equations (1) and (2), define the observability matrix and the attack, interconnection, and noise forced response matrices of the $i-$th subsystem as

$$\mathcal{O}_i = \begin{bmatrix} C_i A_{ii} \\ \vdots \\ C_i A_{ii}^T \end{bmatrix}, \ \mathcal{F}_i^{(a)} = \begin{bmatrix} C_i B_i^a & \dots & 0 \\ \vdots & \ddots & \vdots \\ C_i A_{ii}^{T-1} B_i^a & \dots & C_i B_i^a \end{bmatrix},$$

$$\mathcal{F}_i^{(u)} = \begin{bmatrix} C_i B_i & \dots & 0 \\ \vdots & \ddots & \vdots \\ C_i A_{ii}^{T-1} B_i & \dots & C_i B_i \end{bmatrix}, \ \mathcal{F}_i^{(w)} = \begin{bmatrix} C_i & \dots & 0 \\ \vdots & \ddots & \vdots \\ C_i A_{ii}^{T-1} & \dots & C_i \end{bmatrix}.$$

Analogously, for the system model (3) define the matrices $\mathcal{O}_c$, $\mathcal{F}_a^{(w)}$, and $\mathcal{F}_c^{(w)}$, which are constructed as above by replacing A_i, B_i^a, and C_i with A, B^a, and C, respectively. The measurements (4) and (5) can be written as follows:

$$Y_i = \mathcal{O}_i x_i(0) + \mathcal{F}_i^{(u)} U_i + \mathcal{F}_i^{(a)} U_i^a + \mathcal{F}_i^{(w)} W_i + V_i, \tag{6}$$

$$Y_c = \mathcal{O}_c x(0) + \mathcal{F}_c^{(a)} U^a + \mathcal{F}_c^{(w)} W + V, \tag{7}$$

where $U_i = \begin{bmatrix} u_i^{\mathsf{T}}(0) & u_i^{\mathsf{T}}(1) & \cdots & u_i^{\mathsf{T}}(T-1) \end{bmatrix}^{\mathsf{T}}$. The vectors U_i^a, U^a, W_i and W are the time aggregated signals of u_i^a, u^a, w_i, and w, respectively, and are defined similarly to U_i. Instead, $V_i = \begin{bmatrix} v_i^{\mathsf{T}}(1) & v_i^{\mathsf{T}}(2) & \cdots & v_i^{\mathsf{T}}(T) \end{bmatrix}^{\mathsf{T}}$, and V is defined similarly to V_i. To eliminate the dependency from the unknown variables, let N_i and N_c be bases of the left null spaces of the matrices $\begin{bmatrix} \mathcal{O}_i & \mathcal{F}_i^{(u)} \end{bmatrix}$ and $\mathcal{O}_c$, respectively, and define the processed measurements as

$$\begin{aligned} \widetilde{Y}_i &= N_i Y_i = N_i \left[\mathcal{F}_i^{(a)} U_i^a + \mathcal{F}_i^{(w)} W_i + V_i \right], \\ \widetilde{Y}_c &= N_c Y_c = N_c \left[\mathcal{F}_c^{(a)} U^a + \mathcal{F}_c^{(w)} W + V \right], \end{aligned} \tag{8}$$

where the expressions for $\widetilde{Y}_i$ and $\widetilde{Y}_c$ follows from (6) and (7). Notice that, in the absence of attacks ($U^a = 0$), the measurements $\widetilde{Y}_i$ and $\widetilde{Y}_c$ depend only on the system noise. Instead, in the presence of attacks, such measurements also depend on the attack vector, which may leave a signature for the detectors.[2] We now characterize the statistical properties of $\widetilde{Y}_i$ and $\widetilde{Y}_c$.

Lemma 3.1: ***(Statistical properties of the processed measurements)*** The processed measurements $\widetilde{Y}_i$ and $\widetilde{Y}_c$ satisfy

$$\begin{aligned} &\widetilde{Y}_i \sim \mathcal{N}(\beta_i, \Sigma_i), \text{ for all } i \in \{1, \dots, N\}, \text{ and} \\ &\widetilde{Y}_c \sim \mathcal{N}(\beta_c, \Sigma_c), \end{aligned} \tag{9}$$

where

$$\begin{aligned} \beta_i &= N_i \mathcal{F}_i^{(a)} U_i^a, \\ \beta_c &= N_c \mathcal{F}_c^{(a)} U^a, \\ \Sigma_i &= N_i \left[\left(\mathcal{F}_i^{(w)} \right) (I_T \otimes \Sigma_{w_i}) \left(\mathcal{F}_i^{(w)} \right)^{\mathsf{T}} + (I_T \otimes \Sigma_{v_i}) \right] N_i^{\mathsf{T}}, \\ \Sigma_c &= N_c \left[\left(\mathcal{F}_c^{(w)} \right) (I_T \otimes \Sigma_w) \left(\mathcal{F}_c^{(w)} \right)^{\mathsf{T}} + (I_T \otimes \Sigma_v) \right] N_c^{\mathsf{T}}. \end{aligned} \tag{10}$$

A proof of Lemma 3.1 is postponed to the Appendix. From Lemma 3.1, the mean vectors β_i and β_c depend on the attack vector, while the covariance matrices Σ_i and Σ_c are independent of the attack. This observation motivates us to develop a detection mechanism based on the mean of the processed measurements, rather the covariance matrices.

B. Statistical hypothesis testing framework

In this section we detail our attack detection mechanism, which we assume to be the same for all local and centralized detectors, and we characterize its false alarm and detection probabilities. We start by analyzing the test procedure of the

[2]If $\mathrm{Im}(B_i^a) \subseteq \mathrm{Im}(B_i)$, then $N_i \mathcal{F}_i^{(a)} = 0$ and the processed measurements do not depend on the attack. Thus, our local detection technique can only be successful against attacks that do not satisfy this condition.

$i-$th local detector. Let H_0 be the null hypothesis, where $\beta_i = 0$ and the system is not under attack, and let H_1 be the alternative hypothesis, where $\beta_i \neq 0$ and the system is under attack. To decide which hypothesis is true, or equivalently whether the mean value of the processed measurements is zero, we resort to the generalized log-likelihood ratio test (GLRT):

$$\Lambda_i \triangleq \widetilde{Y}_i^{\mathsf{T}} \Sigma_i^{-1} \widetilde{Y}_i \underset{H_0}{\overset{H_1}{\gtrless}} \tau_i, \tag{11}$$

where the threshold $\tau_i \geq 0$ is selected based on the desired false alarm probability of the test (11) [27]. For a statistical hypothesis testing problem, the false alarm probability equals the probability of deciding for H_1 when H_0 is true, while the detection probability equals the probability of deciding for H_1 when H_1 is true. While the former is used for tuning the threshold, the latter is used for measuring the performance of the test. Formally, the false alarm and detection probabilities of (11) are the probabilities that are conditioned on the hypothesis H_0 and H_1, respectively, and are symbolically denoted as

$$P_i^F = \Pr\left[\Lambda_i \geq \tau_i | H_0\right] \text{ and } P_i^D = \Pr\left[\Lambda_i \geq \tau_i | H_1\right].$$

Similarly, the centralized detector test is defined as

$$\Lambda_c \triangleq \widetilde{Y}_c^{\mathsf{T}} \Sigma_c^{-1} \widetilde{Y}_c \underset{H_0}{\overset{H_1}{\gtrless}} \tau_c, \tag{12}$$

where $\tau_c \geq 0$ is a preselected threshold, and its false alarm and detection probabilities are denoted as P_c^F and P_c^D. We next characterize the false alarm and detection probabilities of the detectors with respect to the system and attack parameters.

Lemma 3.2: ***(False alarm and detection probabilities of local and centralized detectors)*** The false alarm and the detection probabilities of the tests (11) and (12) are, respectively,

$$\begin{aligned} P_i^F &= Q(\tau_i; p_i, 0),\ P_i^D = Q(\tau_i; p_i, \lambda_i),\ \text{and} \\ P_c^F &= Q(\tau_c; p_c, 0),\ P_c^D = Q(\tau_c; p_c, \lambda_c), \end{aligned} \tag{13}$$

where

$$\begin{aligned} p_i &= \mathrm{Rank}(\Sigma_i),\ p_c = \mathrm{Rank}(\Sigma_c), \\ \lambda_i &= (U_i^a)^{\mathsf{T}} M_i (U_i^a),\ \lambda_c = (U^a)^{\mathsf{T}} M_c (U^a), \end{aligned} \tag{14}$$

and

$$\begin{aligned} M_i &= \left(N_i \mathcal{F}_i^{(a)}\right)^{\mathsf{T}} \Sigma_i^{-1} \left(N_i \mathcal{F}_i^{(a)}\right), \\ M_c &= \left(N_c \mathcal{F}_c^{(a)}\right)^{\mathsf{T}} \Sigma_c^{-1} \left(N_c \mathcal{F}_c^{(a)}\right). \end{aligned} \tag{15}$$

Lemma 3.2, whose proof is postponed to the Appendix, allows us to compute the false alarm and detection probabilities of the detectors using the decision thresholds, the system parameters, and the attack vector. Moreover, for fixed P_i^F and P_c^F, the detection thresholds are computed as $\tau_c = Q^{-1}(P_c^F; p_c, 0)$ and $\tau_i = Q^{-1}(P_i^F; p_i, 0)$, where $Q^{-1}(\cdot)$ is the inverse of the complementary Cumulative Distribution Functions (CDF) that is associated with a central chi-squared distribution. The parameters p_i, p_c and λ_i, λ_c in Lemma 3.2 are referred to as *degrees of freedom* and *non-centrality* parameters of the detectors.

Remark 2: ***(System theoretic interpretation of detection probability parameters)*** The degrees of freedom and the non-centrality parameters quantify the knowledge of the detectors about the system dynamics and the energy of the attack signal contained in the processed measurements. In particular:

(Degrees of freedom p_i) The detection probability and the false alarm probability are both increasing functions of the degrees of freedom p_i, because the Q function in (13) is an increasing function of p_i. Thus, increasing p_i by, for instance, increasing the number of sensors or the horizon T, does not necessarily lead to an improvement of the detector performance.

(Non-centrality parameter λ_i) The non-centrality parameter λ_i measures the energy of the attack signal contained in the processed measurements. In the literature of communication and signal processing, the non-centrality parameter is often referred to as signal to noise ratio (SNR) [27]. For fixed τ_i and p_i, the detection probability increases monotonically with λ_i, and approaches the false alarm probability as λ_i tends to zero.

(Decision threshold τ_i) For fixed λ_i and p_i, the probability of detection and the false alarm probability are monotonically decreasing functions of the detection threshold τ_i. This is due to the fact that the complementary CDFs, which define the false alarm and detection probabilities, are decreasing functions of τ_i. As we show later, because of the contrasting behaviors of the false alarm and detection probabilities with respect to all individual parameters, the decentralized detector can outperform the centralized detector. □

We now state a result that provides a relation between the degrees of freedom (p_i and p_c) and the non-centrality parameters (λ_i and λ_c) of the local and the centralized detectors. This result plays a central role in comparing the performance of these centralized and decentralized detectors.

Lemma 3.3: ***(Degrees of freedom and non-centrality parameters)*** Let p_i, p_c and λ_i, λ_c be the degrees of freedom and non-centrality parameters of the $i-$th local and centralized detectors, respectively. Then, $p_i \leq p_c$ and $\lambda_i \leq \lambda_c$ for all $i \in \{1, \ldots, N\}$.

A proof of Lemma 3.3 is postponed to the Appendix. In loose words, given the interpretation of the degrees of freedom and noncentrality parameters in Remark 2, Lemma (3.3) states that a centralized detector has more knowledge about the system dynamics ($p_i \leq p_c$) and its measurements contain a stronger attack signature ($\lambda_i \leq \lambda_c$) than any of the $i-$th local detector. Despite these properties, we will show that the decentralized detector can outperform the centralized one.

IV. Comparison of Centralized and Decentralized Detection of Attacks

In this section we characterize the detection probabilities of the decentralized and centralized detectors, and we derive sufficient conditions for each detector to outperform the other. Recall that the decentralized detector triggers an alarm if any of the local detectors detects an alarm. In other words,

$$\begin{aligned} P_d^F &= \Pr\left[\Lambda_i \geq \tau_i,\ \text{for some } i \in \{1, \ldots, N\} \,|\, H_0\right], \\ P_d^D &= \Pr\left[\Lambda_i \geq \tau_i,\ \text{for some } i \in \{1, \ldots, N\} \,|\, H_1\right], \end{aligned} \tag{16}$$

where P_d^F and P_d^D denote the false alarm and detection probabilities of the decentralized detector, respectively.

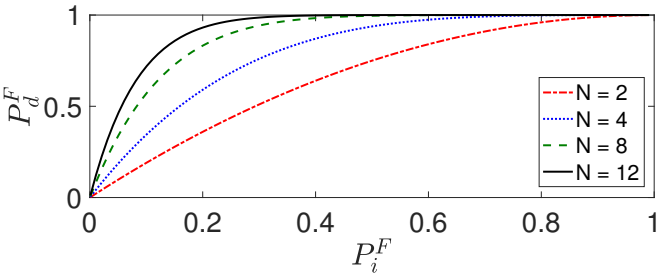

Fig. 1. This figure shows the false alarm probability of the decentralized detector, $\boldsymbol{P_d^F}$, as a function of the identical false alarm probabilities of the local detectors, $\boldsymbol{P_i^F}$, for different numbers of local detectors.

Lemma 4.1: ***(Performance of the decentralized detector)*** The false alarm and detection probabilities in (16) satisfy

$$P_d^F = 1 - \prod_{i=1}^{N} \left(1 - P_i^F\right), \text{ and } P_d^D = 1 - \prod_{i=1}^{N} \left(1 - P_i^D\right). \tag{17}$$

A proof of Lemma 4.1 is postponed to the Appendix. As shown in Fig. 1, for the case when $P_i^F = P_j^F$, for all $i, j \in \{1, \ldots, N\}$, P_d^F increases with increase in P_i^F and N. To allow for a fair comparison between the decentralized and centralized detectors, we assume that $P_c^F = P_d^F$. Consequently, for a fixed false alarm probability P_c^F, the probabilities P_i^F satisfy

$$P_c^F = 1 - \prod_{i=1}^{N} \left(1 - P_i^F\right).$$

We now derive a sufficient condition for the centralized detector to outperform the decentralized detector.

Theorem 4.2: ***(Sufficient condition for $P_c^D \geq P_d^D$)*** Let $P_c^F = P_d^F$, and assume that the following condition is satisfied:

$$\tau_c \leq p_c + \lambda_c - \sqrt{4N(p_c + 2\lambda_c) \ln\left(\frac{1}{1 - P_{\max}^D}\right)}, \tag{18}$$

where $P_{\max}^D = \max\{P_1^D, \ldots, P_N^D\}$. Then, $P_c^D \geq P_d^D$.

A proof of Theorem 4.2 is postponed to the Appendix. We next derive a sufficient condition for the decentralized detector to outperform the centralized detector.

Theorem 4.3: ***(Sufficient condition for $P_d^D \geq P_c^D$)*** Let $P_c^F = P_d^F$, and assume that the following condition is satisfied:

$$\tau_c \geq p_c + \lambda_c + \sqrt{4\left(p_c + 2\lambda_c\right) \ln\left(\frac{1}{1 - (1 - P_{\min}^D)^N}\right)} + \\ + 2\ln\left(\frac{1}{1 - (1 - P_{\min}^D)^N}\right), \tag{19}$$

where $P_{\min}^D = \min\{P_1^D, \ldots, P_N^D\}$. Then $P_d^D \geq P_c^D$.

A proof of Theorem 4.2 is postponed to the Appendix. Theorems 4.2 and 4.3 provide sufficient conditions on the detectors and attack parameters that result in one detector outperforming the other. In particular, from (18) and (19) we note that, depending on decision threshold τ_c, a centralized detector may or may not outperform a decentralized detector. This is intuitive as the Q function, which quantifies the detection probability, is a decreasing function of the detection threshold (see Remark 2). To clarify the effect of attack and detection parameters on the performance trade-offs of the detectors, we now express (18) and (19) using the mean and standard deviation of the test statistic Λ_c in (12). Let

$$\mu_c \triangleq \mathbb{E}\left[\Lambda_c\right] = \lambda_c + p_c, \text{ and}$$
$$\sigma_c \triangleq \text{SD}[\Lambda_c] = \sqrt{2(p_c + 2\lambda_c)}.$$

where the expectation and standard deviation (SD) of Λ_c follows from the fact that under H_1, $\Lambda_c \sim \chi^2(p_c, \lambda_c)$ (see proof of Lemma 3.2). Hence, (18) and (19) can be rewritten, respectively, as

$$\tau_c \leq \mu_c - \sigma_c \underbrace{\sqrt{2N \ln\left(\frac{1}{1 - P_{\max}^D}\right)}}_{\triangleq \kappa_c}, \text{ and} \tag{20a}$$

$$\tau_c \geq \mu_c + \sigma_c \underbrace{\sqrt{2 \ln\left(\frac{1}{1 - (1 - P_{\min}^D)^N}\right)}}_{\triangleq \kappa_d} + \kappa_d^2. \tag{20b}$$

From (20a) and (20b) we note that a centralized detector outperforms the decentralized one if τ_c is κ_c standard deviations smaller than the mean μ_c. Instead, a decentralized detector outperforms the centralized detector if τ_c is at least κ_d standard deviations larger than the mean μ_c. See Fig. 2 for a graphical illustration of this interpretation.

Theorems 4.2 and 4.3 are illustrated in Fig. 3 as a function of the non-centrality parameters. It can be observed that (i) each of the detectors can outperform the other depending on the values of the noncentrality parameter values, (ii) the provided bounds qualitatively capture the actual performance of the centralized and decentralized detectors as the non-centrality parameters increase, and (iii) the provided bounds are rather tight over a large range of non-centrality parameters. In Fig. 4 we show that the difference of the detection probabilities of the centralized and decentralized detectors can be

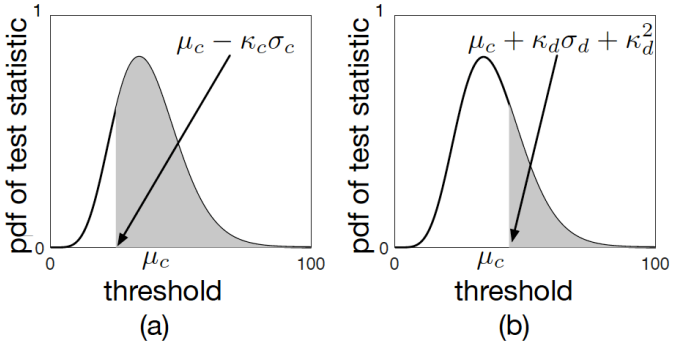

Fig. 2. This figure shows the probability density function (pdf) of $\boldsymbol{\Lambda_c}$ under $\boldsymbol{H_1}$, as a function of threshold $\boldsymbol{\tau_c}$. For $\boldsymbol{\tau_c = \mu_c - \kappa_c\sigma_c}$ and $\boldsymbol{\tau_c = \mu_c + \kappa_d\sigma_d + \sigma_d^2}$, the shaded area in panels (a) and (b) indicates the detection probability of the centralized detector. As seen in panels (a) and (b), an increase in $\boldsymbol{\kappa_c}$ results in larger area (larger detection probability) while a increase in $\boldsymbol{\kappa_d}$ results in smaller area (smaller detection probability).

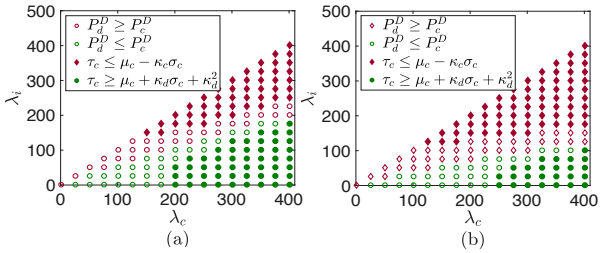

Fig. 3. This figure shows when the decentralized, which comprises identical local detectors, and centralized detectors outperform their counterpart, as a function of the non-centrality parameters. The regions identified by solid markers correspond to the conditions in Theorems 4.2 and 4.3. Instead, regions identified by empty markers are identified numerically. Since $\boldsymbol{\lambda_i \leq \lambda_c}$, the white region (top left) is not admissible. For a fixed $\boldsymbol{P_c^F = P_d^F = 0.01}$, (a) corresponds to the case of $\boldsymbol{N = 2}$ and (b) corresponds to the case of $\boldsymbol{N = 4}$. When $\boldsymbol{N = 4}$, the decentralized detector outperforms the centralized one for a larger set of noncentrality parameters.

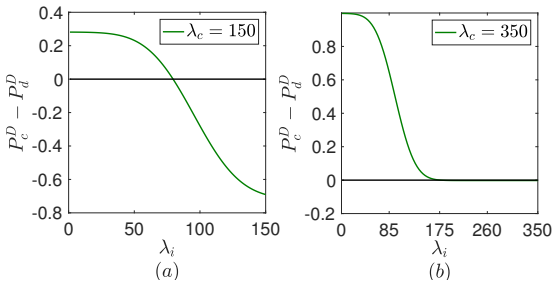

Fig. 4. This figure shows the difference of the centralized and decentralized detection probabilities as a function of $\boldsymbol{\lambda_i}$ for different values of $\boldsymbol{\lambda_c}$. For small values of $\boldsymbol{\lambda_c}$, the detection probability of the decentralized detector can be substantially larger than its centralized counterpart.

large, especially when the non-centrality parameters are small and satisfy $\lambda_c \approx \lambda_i$, as evident in panel (a) of Fig. 4 .

V. Design of optimal attacks

In this section we consider the problem of designing attacks that deteriorate the performance of the interconnected system (1) while remaining undetected from the centralized and decentralized detectors. We measure the degradation induced by an attack with the expected value of the deviation of the state trajectory from the origin. We assume that the attack is a deterministic signal, and thus independent of the noise affecting the system dynamics and measurements. In particular, for a fixed value of the probability P_c^F and a threshold $P_c^F \leq \delta_c \leq 1$, we consider the optimization problem

$$\begin{aligned} \text{(P.1)} \quad \max_{U^a} \quad & \mathbb{E}\left[\sum_{k=1}^{T} x(k)^\mathsf{T} x(k)\right], \\ \text{subject to} \quad & P_c^D \leq \delta_c, \\ & x(k+1) = Ax(k) + B^a u^a(k) + w(k), \end{aligned}$$

where U^a is the deterministic attack input over time horizon T (see (7)). Notice that, because the attack is deterministic, the objective function in (P.1) can be simplified by bringing the expectation inside the summation, and replacing the state equation constraint with the mean state response. Further, because the system parameters and P_c^F are fixed, τ_c and p_c are also fixed, which ensures that P_d^D only depends on noncentrality parameter. This observation along with the fact that $Q(\cdot)$ is increasing function in noncentrality parameter (see Remark 2) allows us to express the detection constraint in terms of λ_c. Specifically, the optimization problem (P.1) can be rewritten as

$$\begin{aligned} \text{(P.2)} \quad \max_{U^a} \quad & \sum_{k=1}^{T} \overline{x}(k)^\mathsf{T} \overline{x}(k) \\ \text{subject to} \quad & (U^a)^\mathsf{T} M_c(U^a) \leq \widetilde{\delta}_c, \\ & \overline{x}(k+1) = A\overline{x}(k) + B^a u^a(k), \end{aligned}$$

where we have used that $\mathrm{Cov}\,[x(k)]$ is independent of the attack $u^a(k)$, and

$$\mathbb{E}[x(k)^\mathsf{T} x(k)] = \overline{x}(k)^\mathsf{T}\overline{x}(k) + \mathrm{Trace}\left(\mathrm{Cov}\,[x(k)]\right),$$

with $\overline{x}(k) = \mathbb{E}[x(k)]$. Further, we have $\widetilde{\delta}_c = Q^{-1}_{p_c,\tau_c}(\delta_c)$, where $Q^{-1}_{p_c,\tau_c}(\alpha) : [0,1] \to [0,\infty]$ denotes the inverse of $Q(\tau_c; p_c, \lambda_c)$ for fixed p_c and τ_c, and $\lambda_c = (U^a)^\mathsf{T} M_c(U^a)$, with M_c as in (15). It should be noticed that the attack constraint in (P.2) essentially limits the (weighted) energy of the attack signal. We next characterize the solution to the optimization problem (P.2).

Theorem 5.1: ***(Optimal attack vectors)*** Let U_c^* be any solution of (P.2). Then, there exist a $\gamma_c > 0$ such that the pair (U_c^*, γ_c) solves the following optimality equations:

$$\left[\mathcal{B}_a^\mathsf{T}\mathcal{B}_a - \gamma_c M_c\right] U_c^* + \mathcal{B}_a^\mathsf{T}\mathcal{A}x(0) = 0, \tag{21a}$$

$$\left(U_c^*\right)^\mathsf{T} M_c(U_c^*) = \widetilde{\delta}_c, \tag{21b}$$

where

$$\mathcal{A} = \begin{bmatrix} A \\ \vdots \\ A^T \end{bmatrix} \text{ and } \mathcal{B}_a = \begin{bmatrix} B^a & \cdots & 0 \\ \vdots & \ddots & \vdots \\ A^{T-1}B^a & \cdots & B^a \end{bmatrix}. \tag{22}$$

A proof of Theorem 5.1 is postponed to the Appendix. Theorem 5.1 not only guarantees the existence of optimal attacks, but it also provides us with necessary conditions to verify if an attack is (locally) optimal. When the system initial state is zero, we can also quantify the performance degradation induced by an optimal attack. Let $\rho_{\max}(A,B)$ and $\nu_{\max}(A,B)$ denote a largest generalized eigenvalue of a matrix pair (A,B) and one of its associated generalized eigenvectors [28].

Lemma 5.2: ***(System degradation with zero initial state)*** Let $x(0) = 0$. Then, the optimal solution to (P.2) is

$$U_c^* = \left(\sqrt{\frac{\widetilde{\delta}_c}{(\nu^*)^\mathsf{T} M_c(\nu^*)}}\right)\nu^*, \tag{23}$$

and its associated optimal cost is

$$J_c^* = \widetilde{\delta}_c\, \rho_{\max}\left(\mathcal{B}_a^\mathsf{T}\mathcal{B}_a, M_c\right), \tag{24}$$

where $\nu^* = \nu_{\max}\left(\mathcal{B}_a^\mathsf{T}\mathcal{B}_a, M_c\right)$.

A proof of Lemma 5.2 is postponed to the Appendix. From (24), notice that the system degradation caused by an optimal attack depends on the detector's tolerance, as measured by $\widetilde{\delta}_c$, and the system dynamics, as measured by $\rho_{\max}(\cdot)$. See

Remark 4 for the influence of processed measurement's noise uncertainty on the system degradation due to optimal attacks.

Remark 3: ***(Optimal attack vector against decentralized detector)*** To characterize the performance degradation of the system analytically, we consider a relaxed form of detection constraint. Specifically, we design optimal attacks subjected to $\overline{P}_d^D \leq \delta_d$ instead of $P_d^D \leq \delta_d$, where $\overline{P}_d^D$ is an upper bound on P_d^D (see Lemma A.2). The design of optimal attacks that are undetectable from the decentralized detector can be formulated in the following way:

$$\begin{aligned} \text{(P.3)} \quad \max_{U^a} \quad & \sum_{k=1}^{T} \overline{x}(k)^{\mathsf{T}} \overline{x}(k) \\ \text{subject to} \quad & \sum_{i=1}^{N} (U_i^a)^{\mathsf{T}} M_i (U_i^a) \leq \widetilde{\delta}_d, \\ & \overline{x}(k+1) = A\overline{x}(k) + B^a u^a(k), \end{aligned}$$

where the summation in the detectability constraint follows from Lemma A.2 and the fact that $\overline{P}_d^D \leq \delta_d$ becomes equivalent to $\sum_{i=1}^N \lambda_i \leq \widetilde{\delta}_d$, where $\widetilde{\delta}_d = Q^{-1}_{p_{\text{sum}},\tau_{\min}}(\delta_d)$, $p_{\text{sum}} = \sum_{i=1}^N p_i$, and $\tau_{\min} = \min_{1\leq i\leq N} \tau_i$. Let Π_i be a permutation matrix such that $U_i^a = \Pi_i U^a$, and let $\Pi = \left[\Pi_1^{\mathsf{T}}, \ldots, \Pi_N^{\mathsf{T}}\right]^{\mathsf{T}}$ and $M_d = \Pi^{\mathsf{T}} \text{blkdiag}(M_1, \ldots, M_N)\Pi$. For any solution U_d^* of (P.2), there exist $\gamma_d > 0$ such that the pair (U_d^*, γ_d) solves the following optimality equations:

$$\begin{aligned} \left[\mathcal{B}_a^{\mathsf{T}} \mathcal{B}_a - \gamma_d M_d\right] U_d^* + \mathcal{B}_a^{\mathsf{T}} \mathcal{A} x(0) = 0, \text{ and} \\ (U_d^*)^{\mathsf{T}} M_d (U_d^*) = \widetilde{\delta}_d. \end{aligned}$$

Further, if $x(0) = 0$, then the largest degradation is $J_d^* = \widetilde{\delta}_d \, \rho_{\max}\left(\mathcal{B}_a^{\mathsf{T}} \mathcal{B}_a, M_d\right)$. □

Remark 4: ***(Maximum degradation of the system performance with respect to system noise)*** To see the role of noise level, in the processed measurements, on the system degradation, we consider the following covariance matrices: $\Sigma_{w_i} = \sigma^2 I_{n_i}$ and $\Sigma_{v_i} = \sigma^2 I_{r_i}$, for $i \in \{1, \ldots, N\}$. Then, from (24) we have

$$J_c^* = \sigma^2 \, \widetilde{\delta}_c \left[\rho_{\max}\left(\mathcal{B}_a^{\mathsf{T}} \mathcal{B}_a, \widetilde{M}_c\right)\right], \tag{25}$$

where $\widetilde{M}_c = \left(N_c \mathcal{F}_c^{(a)}\right)^{\mathsf{T}} \left[\mathcal{F}_c^{(w)} \left(\mathcal{F}_c^{(w)}\right)^{\mathsf{T}} + I\right]^{-1} \left(N_c \mathcal{F}_c^{(a)}\right)$. From (25) we note that the system degradation increases with the increase in the noise level, i.e., σ^2. □

VI. Numerical Comparison of Centralized and Decentralized Detectors

In this section, we demonstrate our theoretical findings on the IEEE RTS-96 power network model [29], which we partition into three subregions as shown in Fig. 5. We followed the approach in [30] to obtain a linear time-invariant model of the power network, and then discretized it using a sampling time of 0.01 seconds. For a false alarm probability $P_c^F = P_d^F = 0.05$, we consider the family of attacks $U^a = \sqrt{\theta/(\mathbf{1}^{\mathsf{T}} M_c \mathbf{1})}\mathbf{1}$, where $\mathbf{1}$ is the vector of all ones and $\theta > 0$. It can be shown that the noncentrality parameters satisfy $\lambda_c = \theta$ and $\lambda_i = \theta(\mathbf{1}^{\mathsf{T}} M_i \mathbf{1})/(\mathbf{1}^{\mathsf{T}} M_c \mathbf{1})$, and moreover, the choice of vector $\mathbf{1}$ is arbitrary and it does not affecting the following results.

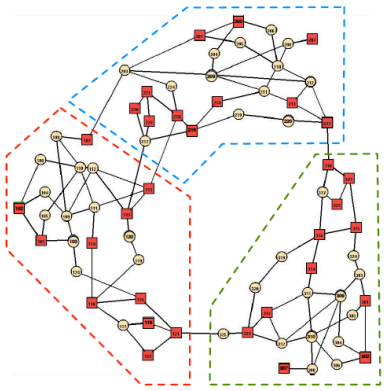

Fig. 5. The figure shows a single-line diagram of IEEE RTS96 power network, which is composed of three weakly-coupled areas (subsystems). The square nodes denote the generators, while the circular nodes denotes the load buses of the network [30].

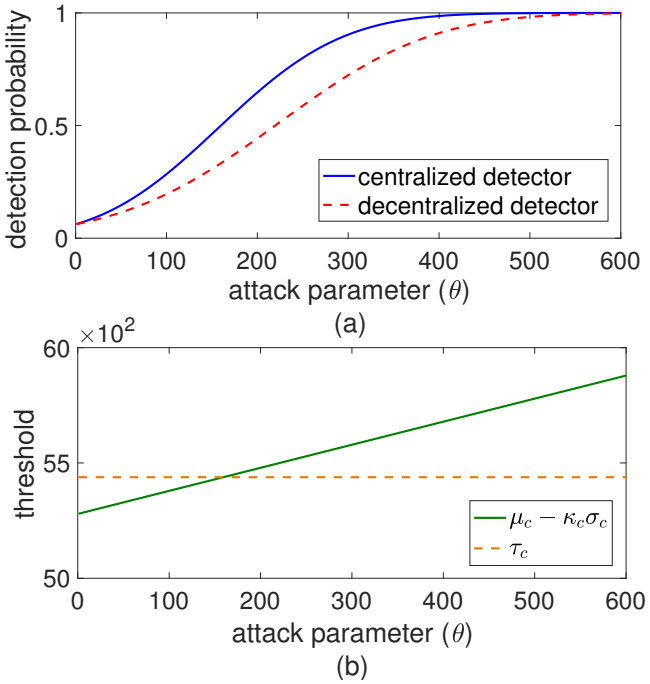

Fig. 6. An illustration of a scenario in which the centralized detector outperforms the decentralized detector for the IEEE RTS-96 power network. In panel (a), we plot the detection probabilities of the detectors with respect to the attack parameter $\boldsymbol{\theta}$. Instead, in panel (b) we plot the right (solid line) and left hand expressions (dashed line) of the inequality in (20a) as a function of $\boldsymbol{\theta}$. For attacks such that $\boldsymbol{\theta > 200}$, the sufficient condition (20a) holds true, it guarantees that $\boldsymbol{P_c^D \geq P_d^D}$.

(Illustration of Theorem 4.2) For the measurement horizon of $T = 100$ seconds, the values of p_c and τ_c are 5130 and 5480.6, respectively. Fig. 6 show that the detection probabilities of the centralized and decentralized detectors increase monotonically with the attack parameter θ. As predicted by the sufficient condition (20a) and shown in Fig. 6, the centralized detector is guaranteed to outperform the decentralized detector when $\theta > 173$. This figure also shows that our condition is conservative,

because $P_c^D \geq P_d^D$ for all values of θ as shown in Fig. 6.

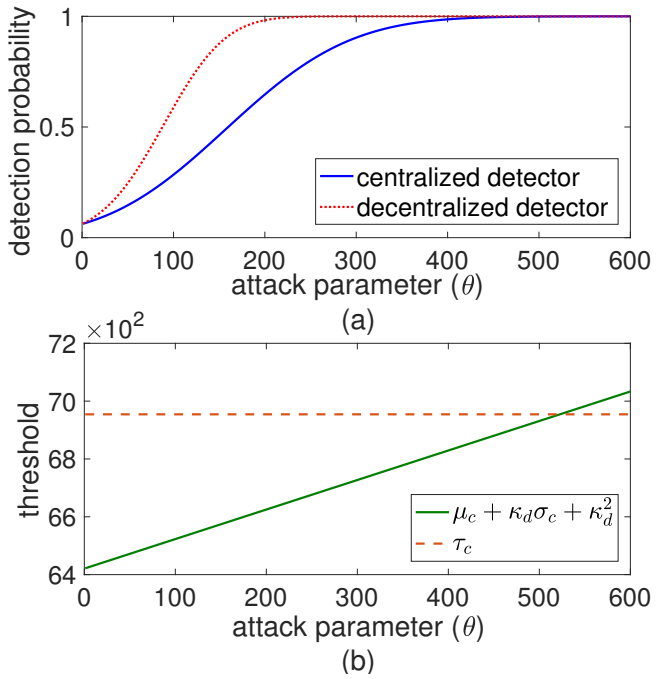

Fig. 7. An illustration of a scenario in which the decentralized detector outperforms the centralized detector for the IEEE RTS-96 power network. In panel (a), we plot the detection probabilities of the detectors with respect to the attack parameter $\boldsymbol{\theta}$. Instead, in panel (b) we plot the right (solid line) and left hand expressions (dashed line) of the inequality in (20b) as a function of $\boldsymbol{\theta}$. For attacks such that $\boldsymbol{\theta < 500}$, the sufficient condition (20b) holds true, and it guarantees that $\boldsymbol{P_c^D \leq P_d^D}$.

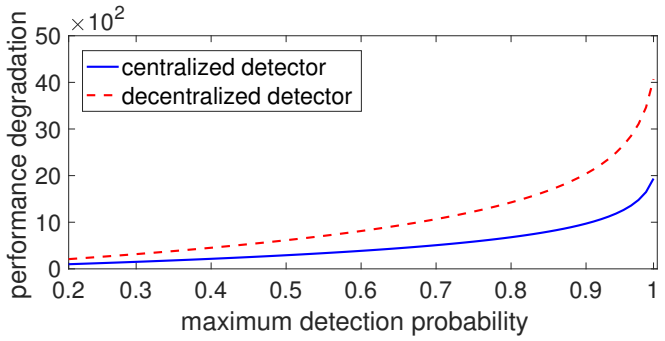

Fig. 8. This figure shows the performance degradation induced by undetectable optimal attacks on the IEEE RTS-96 power network. The performance degradation is computed using the optimal cost $\boldsymbol{J_c^*}$ and $\boldsymbol{J_d^*}$ derived in Lemma 5.2 and Remark 3, respectively. Instead, the maximum detection probability is given by the tuning parameters $\boldsymbol{\delta_c}$ and $\boldsymbol{\delta_d}$ in the detection probability constraints of the optimization problems (P.2) and (P.3), respectively.

(Illustration of Theorem 4.3) Contrary to the previous example, by letting $T = 125$ seconds, we obtain $p_c = 6755$ and $\tau_c = 6947.3$. For these choice of parameters, the decentralized detector is guaranteed to outperform the centralized detector when $\theta \leq 511$. This behavior is predicted by our sufficient condition (20b), and it is illustrated in Fig. 7. As in the previous example, the estimation provided by our condition (20b) is conservative, as illustrated in Fig. 7.

(Illustration of Lemma 5.2) In Fig. 8 we compare the performance degradation induced by the optimal attacks designed according to the optimization problems (P.2) and (P.3) with zero initial conditions. In particular, we plot the optimal costs J_c^* and J_d^* against the tolerance levels $\widetilde{\delta}_c$ and $\widetilde{\delta}_d$, respectively. As expected, the performance degradation is proportional to the tolerance levels and, for the considered setup, it is larger in the case of the decentralized detector.

VII. Conclusion

In this work we compare the performance of centralized and decentralized schemes for the detection of attacks in stochastic interconnected systems. In addition to quantifying the performance of each detection scheme, we prove the counterintuitive result that the decentralized scheme can, at times, outperform its centralized counterpart, and that this behavior results due to the simple versus composite nature of the attack detection problem. We illustrate our findings through academic examples and a case study based on the IEEE RTS-96 power system. Several questions remain of interest for future investigation, including the characterization of optimal detection schemes, an analytical comparison of the degradation induced by undetectable attacks as a function of the detection scheme, and the analysis of iterative detection strategies.

References

[1] F. Pasqualetti, F. Drfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," IEEE Transactions on Automatic Control, vol. 58, no. 11, pp. 2715–2729, Nov 2013.

[2] Y. Lun, A. D'Innocenzo, I. Malavolta, M. Domenica, and D. Benedetto, "Cyber-physical systems security: a systematic mapping study," arxiv, 2016, available at https://arxiv.org/pdf/1605.09641.pdf.

[3] J. Chen and R. Patton, Robust Model-Based Fault Diagnosis for Dynamic Systems. Springer-Verlag New York, 1999.

[4] Y. Yuan, Q. Zhu, F. Sun, Q. Wang, and T. Basar, "Resilient control of cyber-physical systems against denial-of-service attacks," in 2013 6th International Symposium on Resilient Control Systems (ISRCS), Aug 2013, pp. 54–59.

[5] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," IEEE Transactions on Automatic Control, vol. 59, no. 6, pp. 1454–1467, 2014.

[6] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," IEEE Transactions on Smart Grid, vol. 5, no. 2, pp. 580–591, March 2014.

[7] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," IEEE Transactions on Smart Grid, vol. 5, no. 2, pp. 612–621, March 2014.

[8] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," IEEE Transactions on Automatic Control, vol. 60, no. 11, pp. 3023–3028, Nov 2015.

[9] Y. Mo and B. Sinopoli, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," IEEE Transactions on Automatic Control, vol. 61, no. 9, pp. 2618–2624, Sept 2016.

[10] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," IEEE Transactions on Control of Network Systems, vol. 4, no. 1, pp. 106–117, March 2017.

[11] 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, vol. 62, no. 9, pp. 4618–4624, Sept 2017.

[12] C. Kwon and I. Hwang, "Reachability analysis for safety assurance of cyber-physical systems against cyber attacks," IEEE Transactions on Automatic Control, vol. 63, no. 7, pp. 2272–2279, July 2018.

[13] 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, Allerton, IL, USA, Sep. 2011, pp. 1486–1491.

[14] H. Nishino and H. Ishii, "Distributed detection of cyber attacks and faults for power systems," IFAC Proceedings Volumes, vol. 47, no. 3, pp. 1932 – 11 937, 2014.

[15] M. N. Kurt, Y. Yılmaz, and X. Wang, "Distributed quickest detection of cyber-attacks in smart grid," IEEE Transactions on Information Forensics and Security, vol. 13, no. 8, pp. 2015–2030, Aug 2018.

[16] J. Zhao and L. Mili, "Power system robust decentralized dynamic state estimation based on multiple hypothesis testing," IEEE Transactions on Power Systems, vol. 33, no. 4, pp. 4553–4562, July 2018.

[17] E. M. Hammad, A. K. Farraj, and D. Kundur, "A resilient feedback linearization control scheme for smart grids under cyber-physical disturbances," in 2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), Feb 2015, pp. 1–5.

[18] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," IEEE Transactions on Aerospace and Electronic Systems, vol. 17, no. 4, pp. 501–510, July 1981.

[19] P. Varshney, Distributed Detection and Data Fusion. Springer-Verlag New York, 1997.

[20] J. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," IEEE Transactions on Signal Processing, vol. 51, no. 2, pp. 407–416, Feb 2003.

[21] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Energy-efficient detection in sensor networks," IEEE Journal on Selected Areas in Communications, vol. 23, no. 4, pp. 693–702, April 2005.

[22] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Distributed sensor fault diagnosis for a network of interconnected cyberphysical systems," IEEE Transactions on Control of Network Systems, vol. 2, no. 1, pp. 11–23, March 2015.

[23] X.-G. Yan and C. Edwards, "Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer," Int. Journal of Control, vol. 81, no. 4, pp. 591–606, 2008.

[24] I. Shames, A. M. H. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed fault detection for interconnected second-order systems," Automatica, vol. 47, no. 12, pp. 2757 – 2764, 2011.

[25] R. M. G. Ferrari, T. Parisini, 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, vol. 57, no. 2, Feb 2012.

[26] A. Rajasekhar, V. Katewa, and F. Pasqualetti, "Attack detection in stochastic interconnected systems: Centralized vs decentralized detectors," in 57th IEEE Conference on Decision and Control (CDC), 2018, To appear.

[27] H. V. Poor, An Introduction to Signal Detection and Estimation. Springer-Verlag New York, 1994.

[28] K. Daniel, Numerical Methods for General and Structure Eigenvalue Problems. Springer-Verlag Berlin Heidelberg, 2005.

[29] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour, and C. Singh, "The IEEE reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee," IEEE Transactions on Power Systems, vol. 14, no. 3, pp. 1010–1020, Aug 1999.

[30] F. Dörfler, F. Pasqualetti, and F. Bullo, "Continuous-time distributed observers with discrete communication," IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 2, pp. 296–304, 2013.

[31] T. Anderson, An Introduction to Multivariate Statistical Analysis. Wiley, New York, 1958.

[32] E. K. Chong and S. H. Zak., An Introduction to Optimization. Wiley, New York, 2013.

[33] N. L. Johnson, S. Kotz, and N. Balakrishnan, Continuous univariate distributions, Volume 2. Wiley & Sons, 1995.

[34] L. Birg, "An alternative point of view on Lepski's method," Institute of Mathematical Statistics, vol. 36, pp. 113–133, 2001.

APPENDIX

Proof of Lemma 3.1:

Since the attack vectors U_i^a and U^a are deterministic, and W_i, V_i, V, and W are zero mean random vectors, from the linearity of the expectation operator it follows from (8) that

$$\beta_i \triangleq \mathbb{E}[\widetilde{Y}_i] = N_i \mathcal{F}_i^{(a)} U_i^a, \text{ and}$$
$$\beta_c \triangleq \mathbb{E}[\widetilde{Y}_c] = N_c \mathcal{F}_c^{(a)} U_c^a.$$

Further, from the properties of $\mathrm{Cov}[\cdot]$, we have the following:

$$\begin{aligned}
\Sigma_i &\triangleq \mathrm{Cov}\left[\widetilde{Y}_i\right] \\
&= N_i \mathrm{Cov}\left[Y_i\right] N_i^{\mathsf{T}} \\
&\overset{(a)}{=} N_i \left[\mathrm{Cov}\left[\mathcal{F}_i^{(w)} W_i\right] + \mathrm{Cov}[V_i]\right] N_i^{\mathsf{T}} \\
&\overset{(b)}{=} N_i \left[\left(\mathcal{F}_i^{(w)}\right) \mathrm{Cov}\left[W_i\right] \left(\mathcal{F}_i^{(w)}\right)^{\mathsf{T}} + \mathrm{Cov}[V_i]\right] N_i^{\mathsf{T}} \\
&= N_i \left[\left(\mathcal{F}_i^{(w)}\right) \left(I_T \otimes \Sigma_{w_i}\right) \left(\mathcal{F}_i^{(w)}\right)^{\mathsf{T}} + \left(I_T \otimes \Sigma_{v_i}\right)\right] N_i^{\mathsf{T}},
\end{aligned}$$

where (a) follows because the measurement and process noises are independent of each other. Instead, (b) is due to the fact that the noise vectors are independent and identically distributed. Similar analysis also results in the expression of Σ_c, and hence the details are omitted. Finally, by invoking the fact that linear transformations preserve Gaussianity, the distribution of $\widetilde{Y}_i$ and $\widetilde{Y}_c$ is Gaussian as well. ■

Proof of Lemma 3.2:

From the statistics and distributional form of $\widetilde{Y}_i$ and $\widetilde{Y}_c$ (see (9)), and threshold tests defined in (11) and (12), it follows from [31, Theorem 3.3.3] that, under

1) null hypothesis H_0: $\Lambda_i \sim \chi^2(p_i)$ and $\Lambda_c \sim \chi^2(p_c)$, where p_i and p_c are defined in (14).
2) alternative hypothesis H_1: $\Lambda_i \sim \chi^2(p_i, \lambda_i)$ and $\Lambda_c \sim \chi^2(p_c, \lambda_c)$, where $\lambda_i = \beta_i^{\mathsf{T}} \Sigma_i^{-1} \beta_i$ and $\lambda_c = \beta_c^{\mathsf{T}} \Sigma_c^{-1} \beta_c$.

By substituting $\beta_i = N_i \mathcal{F}_i^{(a)} U_i^a$ and $\beta_c = N_c \mathcal{F}_c^{(a)} U_c^a$ (see Lemma 3.1) and rearranging the terms, we get the expressions of λ_i and λ_c in (14). Finally, from the aforementioned distributional forms of Λ_i and Λ_c, it now follows that the false alarm and the detection probabilities of the tests (11) and (12) are the right tail probabilities (represented by $Q(\cdot)$ function) of the central and noncentral chi-squared distributions, respectively. Hence, the expressions in (13) follow. ■

Proof of Lemma 3.3:

Without loss of generality let $i = 1$. Thus, it suffices to show that a) $p_1 \leq p_c$ and b) $\lambda_1 \leq \lambda_c$.

Case (a): For brevity, define

$$\begin{aligned}
\widetilde{\Sigma}_i &= \left[\left(\mathcal{F}_i^{(w)}\right) \left(I_T \otimes \Sigma_{w_i}\right) \left(\mathcal{F}_i^{(w)}\right)^{\mathsf{T}} + \left(I_T \otimes \Sigma_{v_i}\right)\right] \text{ and} \\
\widetilde{\Sigma}_c &= \left[\left(\mathcal{F}_c^{(w)}\right) \left(I_T \otimes \Sigma_w\right) \left(\mathcal{F}_c^{(w)}\right)^{\mathsf{T}} + \left(I_T \otimes \Sigma_v\right)\right],
\end{aligned} \tag{26}$$

and note that $\widetilde{\Sigma}_i > 0$ and $\widetilde{\Sigma}_c > 0$. From Lemma 3.1, Lemma 3.2, and (26), we have

$$\begin{aligned}
p_c = \mathrm{Rank}(\Sigma_c) &= \mathrm{Rank}\left(\left(N_c \widetilde{\Sigma}_c^{1/2}\right)\left(N_c \widetilde{\Sigma}_c^{1/2}\right)^{\mathsf{T}}\right) \\
&= \mathrm{Rank}\left(N_c \widetilde{\Sigma}^{1/2}\right) = \mathrm{Rank}\left(N_c\right).
\end{aligned}$$

Similarly, $p_1 = \mathrm{Rank}\left(N_1\right)$. Since, N_1^{T} and N_c^{T} are a basis vectors of the null spaces $\mathcal{N}_1^L$ and $\mathcal{N}_c^L$ (see (37)) respectively, it follows from Proposition A.1 that $p_1 \leq p_c$.

Case (b): As the proof for this result is rather long and tedious, we break it down in to multiple steps:

- *Step 1*: Express λ_1 and λ_c using the statistics of a permuted version of Y_c.
- *Step 2*: Obtain lower bound on λ_c, which depends on the statistics of the measurements pertaining to Subsystem 1.
- *Step 3*: Show that λ_1 is less than bound in Step 2.

Step 1 (alternative form of λ_1 and λ_c):

Notice that λ_1 and λ_c in (14) can be expressed as $\lambda_1 = \beta_1^{\mathsf{T}}\Sigma_1^{-1}\beta_1$ and $\lambda_c = \beta_c^{\mathsf{T}}\Sigma_c^{-1}\beta_c$, respectively, where β_1, β_c, Σ_1, and Σ_c are defined in Lemma 3.1. For convenience, we express λ_1 and λ_c in an alternative way. Let $i \in \{1,\dots,N\}$ and consider the $i-$th sensor measurements of (3)

$$y_{c,i}(k) = \underbrace{\begin{bmatrix}0 & \cdots & C_i & \cdots & 0\end{bmatrix}}_{\triangleq C_{c,i}} x(k) + v_i(k). \tag{27}$$

Also, define $Y_{c,i} = \begin{bmatrix}y_{c,i}^{\mathsf{T}}(1) & \dots & y_{c,i}^{\mathsf{T}}(T)\end{bmatrix}^{\mathsf{T}}$ and $\widehat{Y}_c = \begin{bmatrix}Y_{c,1}^{\mathsf{T}} & \dots & Y_{c,N}^{\mathsf{T}}\end{bmatrix}$. Now, from (27) and state equation in (3), $Y_{c,i}$ can be expanded as

$$Y_{c,i} = \mathcal{O}_{c,i}x(0) + \mathcal{F}_{c,i}^{(a)}U^a + \mathcal{F}_{c,i}^{(w)}W + V_i,$$

where the matrices $\mathcal{O}_{c,i}$, $\mathcal{F}_{c,i}^{(a)}$, and $\mathcal{F}_{c,i}^{(w)}$ are similar to the matrices defined in Section II-A. By substituting the above decomposition of $Y_{c,i}$ in $\widehat{Y}$ we have

$$\widehat{Y}_c = \underbrace{\begin{bmatrix}\mathcal{O}_{c,1}\\ \vdots \\ \mathcal{O}_{c,N}\end{bmatrix}}_{\widehat{\mathcal{O}}_c} x(0) + \underbrace{\begin{bmatrix}\mathcal{F}_{c,1}^{(a)}\\ \vdots \\ \mathcal{F}_{c,N}^{(a)}\end{bmatrix}}_{\widehat{\mathcal{F}}_c^a} U^a + \underbrace{\begin{bmatrix}\mathcal{F}_{c,1}^{(w)}\\ \vdots \\ \mathcal{F}_{c,N}^{(w)}\end{bmatrix}}_{\widehat{\mathcal{F}}_c^w} W + V.$$

Moreover, from the distributional assumptions on W and V, it readily follows that (similarly to the proof of Lemma 3.1),

$$\widehat{Y}_c \sim \mathcal{N}\left(\widehat{\mathcal{O}}_c x(0) + \widehat{\mathcal{F}}_c^a U^a, \Sigma\right), \tag{28}$$

where $\Sigma = \left(\widehat{\mathcal{F}}_c^w\right)(I_T \otimes \Sigma_w)\left(\widehat{\mathcal{F}}_c^w\right)^{\mathsf{T}} + (I_T \otimes \Sigma_v)$, and Σ_w and Σ_v are defined same as in Lemma 3.1.

Now, consider the measurement equation $y_i(k)$ in (1) and note that $C_{c,i}x(k) = C_i x_i(k)$. Thus, $y_i(k) = y_{c,i}(k)$, for all $i \in \{1,\dots,N\}$ and $k \in \mathbb{N}$. From this observation it follows that $Y_i = Y_{c,i} = \Pi_i \widehat{Y}_c$, where Π_i is a selection matrix. Let $\widetilde{N}_i = N_i\Pi_i$ and note that $\widetilde{N}_i\widehat{\mathcal{O}} = N_i\mathcal{O}_{c,i}$. Further from Proposition A.1 we have $N_i\mathcal{O}_{c,i} = 0$. With these facts in place, from Lemma 3.1 we now have

$$\begin{aligned}\beta_i &= \widetilde{N}_i\widehat{\mathcal{F}}_c^a U^a, \text{ and}\\ \Sigma_i &= \widetilde{N}_i\Sigma\widetilde{N}_i^{\mathsf{T}}.\end{aligned} \tag{29}$$

Similarly, since $\widehat{Y}_c$ is just a rearrangement of Y_c (see (5)), there exists a permutation matrix Q such that $Y_c = Q\widehat{Y}_c$, and, ultimately $\widetilde{Y}_c = N_cY_c = N_cQ\widehat{Y}_c$. Thus,

$$\begin{aligned}\beta_c &= N_cQ\widehat{\mathcal{F}}_c^a U^a, \text{ and}\\ \Sigma_c &= N_cQ\Sigma(N_cQ)^{\mathsf{T}}.\end{aligned} \tag{30}$$

Let $z = \widehat{\mathcal{F}}_c^a U^a$. From (29) and (30) we have

$$\begin{aligned}\lambda_1 &= z^{\mathsf{T}}\widetilde{N}_1^{\mathsf{T}}\left[\widetilde{N}_1\Sigma\widetilde{N}_1^{\mathsf{T}}\right]^{-1}\widetilde{N}_1,\\ \lambda_c &= z^{\mathsf{T}}(N_cQ)^{\mathsf{T}}\left[(N_cQ)\Sigma(N_cQ)^{\mathsf{T}}\right]^{-1}(N_cQ).\end{aligned} \tag{31}$$

Step 2 (lower bound on λ_c):

Since, $Y_c = N_cY_c = N_cQ\widehat{Y}_c$, it follows that N_cQ is the basis of the null space $\widehat{\mathcal{O}}_c$. Further, the row vectors of $\mathcal{O}_{c,i}$ and $\mathcal{O}_{c,j}$ are linearly independent, whenever $i \neq j$. Using these facts we can define $N_{c,i} = \begin{bmatrix}N_{c,i}^i & \cdots & N_{c,i}^N\end{bmatrix}$ such that $N_cQ = \begin{bmatrix}N_{c,1}^{\mathsf{T}} & \cdots & N_{c,N}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$, where $N_{c,i}^i\mathcal{O}_{c,i} = 0$. Let $P_1 = \begin{bmatrix}(N_{c,2})^{\mathsf{T}} & \cdots & (N_{c,N})^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$ and note that

$$\begin{aligned}(N_cQ)\,\Sigma\,(N_cQ)^{\mathsf{T}} &= \begin{bmatrix}N_{c,1}\\ P_1\end{bmatrix}\Sigma\begin{bmatrix}N_{c,1}^{\mathsf{T}} & P_1^{\mathsf{T}}\end{bmatrix}\\ &= \begin{bmatrix}N_{c,1}\Sigma N_{c,1}^{\mathsf{T}} & N_{c,1}\Sigma P_1^{\mathsf{T}}\\ N_{c,1}^{\mathsf{T}}\Sigma P_1 & P_1^{\mathsf{T}}\Sigma P_1\end{bmatrix}.\end{aligned}$$

Let $S_1 = N_{c,1}\Sigma N_{c,1}^{\mathsf{T}}$. Since $\Sigma > 0$, it follows that both the matrices S_1 and $P_1^{\mathsf{T}}\Sigma P_1$ are invertible. Hence, from Schur's complement, there exists a matrix $X \geq 0$ such that

$$\left[(N_cQ)\,\Sigma\,(N_cQ)^{\mathsf{T}}\right]^{-1} = \begin{bmatrix}S_1^{-1} & 0\\ 0 & 0\end{bmatrix} + X. \tag{32}$$

Similarly, consider the following partition of Σ:

$$\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix},$$

where $\Sigma_{11} > 0$ and $\Sigma_{22} > 0$, and let $S_2 = (N_{c,1}^1)\Sigma_{11}(N_{c,1}^1)^{\mathsf{T}}$. Invoking Schur's complement, we have the following:

$$S_1^{-1} = \begin{bmatrix}S_2^{-1} & 0\\ 0 & 0\end{bmatrix} + Y, \tag{33}$$

where $Y \geq 0$. Substituting(32) and (33) in (31), we have

$$\begin{aligned}\lambda_c &= z^{\mathsf{T}}(N_cQ)^{\mathsf{T}}\begin{bmatrix}S_1^{-1} & 0\\ 0 & 0\end{bmatrix}(N_cQ)z + \underbrace{z^{\mathsf{T}}(N_cQ)^{\mathsf{T}}X(N_cQ)z}_{\geq 0}\\ &\geq \left[(N_{c,1}z)^{\mathsf{T}}\,(P_1z)^{\mathsf{T}}\right]\begin{bmatrix}S_1^{-1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}N_{c,1}z\\ P_1z\end{bmatrix}\\ &= z^{\mathsf{T}}\left(N_{c,1}^{\mathsf{T}}S_1^{-1}N_{c,1}\right)z\\ &= z^{\mathsf{T}}(N_{c,1})^{\mathsf{T}}\begin{bmatrix}S_2^{-1} & 0\\ 0 & 0\end{bmatrix}(N_{c,1})z + \underbrace{z^{\mathsf{T}}(N_{c,1})^{\mathsf{T}}Y(N_{c,1})z}_{\geq 0}\\ &\geq z^{\mathsf{T}}(N_{c,1})^{\mathsf{T}}\begin{bmatrix}S_2^{-1} & 0\\ 0 & 0\end{bmatrix}(N_{c,1})z\\ &= z^{\mathsf{T}}\begin{bmatrix}(N_{c,1}^1)^{\mathsf{T}}S_2^{-1}N_{c,1}^1 & 0\\ 0 & 0\end{bmatrix}z.\end{aligned} \tag{34}$$

Instead, λ_1 in (31) can be shown as

$$\lambda_1 = z^{\mathsf{T}}\begin{bmatrix}N_1^{\mathsf{T}}\left[N_1\Sigma_{11}N_1^{\mathsf{T}}\right]^{-1}N_1 & 0\\ 0 & 0\end{bmatrix}z, \tag{35}$$

where we used the fact that $\widetilde{N}_1 = N_1\Pi_1$.

Step 3 ($\lambda_c \geq \lambda_1$):

For $\lambda_c \geq \lambda_1$ to hold true, it suffices to show the following:

$$(N_{c,1}^1)^{\mathsf{T}}S_2^{-1}N_{c,1}^1 \geq N_1^{\mathsf{T}}\left[N_1\Sigma_{11}N_1^{\mathsf{T}}\right]^{-1}N_1.$$

By invoking Proposition A.1, we note that there exists a full row rank matrix F_1, such that $N_1 = F_1N_{c,1}^1$. Since F_1^{T} is a full column rank matrix, we can define an invertible matrix

$\widetilde{F}_1^{\mathsf{T}} \triangleq [F_1^{\mathsf{T}}\, M_1^{\mathsf{T}}]$, where M_1 forms a basis for null space of F_1, such that the following holds

$$S_2^{-1} = \widetilde{F}_1^{\mathsf{T}} \left[\widetilde{F}_1 S_2 \widetilde{F}_1^{\mathsf{T}}\right]^{-1} \widetilde{F}_1 = \widetilde{F}_1^{\mathsf{T}} \begin{bmatrix} F_1 S_2 F_1^{\mathsf{T}} & F_1 S_2 M_1^{\mathsf{T}} \\ M_1 S_2 F_1^{\mathsf{T}} & M_1 S_2 M_1^{\mathsf{T}} \end{bmatrix}^{-1} \widetilde{F}_1.$$

By invoking Schur's complement, it follows that

$$\begin{bmatrix} F_1 S_2 F_1^{\mathsf{T}} & F_1 S_2 M_1^{\mathsf{T}} \\ M_1 S_2 F_1^{\mathsf{T}} & M_1 S_2 M_1^{\mathsf{T}} \end{bmatrix}^{-1} = \begin{bmatrix} \left(F_1 S_2 F_1^{\mathsf{T}}\right)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + Y,$$

where $Z \geq 0$. Hence,

$$(N_{c,1}^1)^{\mathsf{T}} S_2^{-1} N_{c,1}^1 = (\widetilde{F}_1 N_{c,1}^1)^{\mathsf{T}} \begin{bmatrix} \left(F_1 S_2 F_1^{\mathsf{T}}\right)^{-1} & 0 \\ 0 & 0 \end{bmatrix} (\widetilde{F}_1 N_{c,1}^1) + (\widetilde{F}_1 N_{c,1}^1)^{\mathsf{T}} Z (\widetilde{F}_1 N_{c,1}^1).$$

By substituting $\widetilde{F}_1^{\mathsf{T}} = [F_1^{\mathsf{T}}\, M_1^{\mathsf{T}}]$ in the above expression, and rearranging the terms we have

$$(N_{c,1}^1)^{\mathsf{T}} S_2^{-1} N_{c,1}^1 = (F_1 N_{c,1}^1)^{\mathsf{T}} \left(F_1 S_2 F_1^{\mathsf{T}}\right)^{-1} (F_1 N_{c,1}^1) + (\widetilde{F}_1 N_{c,1}^1)^{\mathsf{T}} Z (\widetilde{F}_1 N_{c,1}^1).$$

The required inequality follows by substituting $S_2 = (N_{c,1}^1)\Sigma_{11}(N_{c,1}^1)^{\mathsf{T}}$ and $N_1 = F_1 N_{c,1}$, and recalling the fact that the sum of two positive semi definite matrices is greater than or equal to either of the matrices. ■

Proof of Lemma 4.1:

Let $\mathcal{E}_i$ be an event that the $i-$th local detector decides H_1 when the true hypothesis is H_0. Then, $P_i^F = \Pr[\mathcal{E}_i]$. Let $\mathcal{E}_i^{\complement}$ be the complement of $\mathcal{E}_i$. Then, from (16) it follows that

$$P_d^F = \Pr\left(\bigcup_{i=i}^{N} \mathcal{E}_i\right) = 1 - \Pr\left(\bigcap_{i=i}^{N} \mathcal{E}_i^{\complement}\right) \overset{(a)}{=} 1 - \prod_{i=1}^{N} \Pr\left(\mathcal{E}_i^{\complement}\right) = 1 - \prod_{i=1}^{N} \left(1 - \Pr\left(\mathcal{E}_i\right)\right) = 1 - \prod_{i=1}^{N} \left(1 - P_i^F\right),$$

where for the (a) we used the fact that the events $\mathcal{E}_i$ are mutually independent for all $i \in \{1, \dots N\}$. To see this fact, notice that the event $\mathcal{E}_i$ is defined on $\widetilde{Y}_i$ (see (8)). Further, $\widetilde{Y}_i$ depends only on the deterministic attack signal U_i^a and the noise vectors V_i and W_i, but not on the interconnection signal U_i (see (6)). Now, by invoking the fact that noises variables across different subsystems are independent, it also follows that the events $\mathcal{E}_i$ are also mutually independent. Similar procedure will lead to the analogous expression for P_d^D and hence, the details are omitted. ■

Proof of Theorem 4.2:

Let $\mu_c = p_c + \lambda_c$ and $\sigma_c = \sqrt{2(p_c + 2\lambda_c)}$, and assume that (18) holds true. Then, from the monotonicity property of the CDF associated with the test statistic Λ_c, which follows $\chi^2(p_c, \lambda_c)$, we have the following inequality

$$\Pr[\Lambda_c \leq \tau_c] \leq \Pr\left[\Lambda_c \leq \mu_c - \sigma_c \sqrt{2N \ln\left(\frac{1}{1 - P_{\text{max}}^D}\right)}\right].$$

From the inequality (41b), it now follows that

$$\Pr[\Lambda_c \leq \tau_c] \leq \exp\left(-N \ln\left(\frac{1}{1 - P_{\text{max}}^D}\right)\right) = \exp\left(\ln\left(1 - P_{\text{max}}^D\right)^N\right) \leq \prod_{i=1}^{N} \left(1 - P_i^D\right),$$

where for the last inequality we used the fact that $P_i^D \leq P_{\text{max}}^D$ for all $i \in \{1, \dots, N\}$. By using the above inequality and Lemma 3.2, under hypothesis H_1, we have

$$P_c^D = 1 - \Pr[\Lambda_c \leq \tau_c | H_1] \geq 1 - \prod_{i=1}^{N} \left(1 - P_i^D\right) = P_d^D. \quad ■$$

Proof of Theorem 4.3

Let $\mu_c = p_c + \lambda_c$ and $\sigma_c = \sqrt{2(p_c + 2\lambda_c)}$, and assume that (19) holds true. Then, from the monotonicity property of the CDF associated with the test statistic Λ_c, which follows $\chi^2(p_c, \lambda_c)$, we have the following inequality

$$\Pr[\Lambda_c \leq \tau_c] \geq \Pr\left[\Lambda_c \leq \mu_c + \sigma_c \sqrt{2 \ln\left(\frac{1}{1 - (1 - P_{\text{min}}^D)^N}\right)} + 2 \ln\left(\frac{1}{1 - (1 - P_{\text{min}}^D)^N}\right)\right].$$

From the inequality (41a), it now follows that

$$\Pr[\Lambda_c \leq \tau_c] \geq 1 - \exp\left(-\ln\left(\frac{1}{1 - (1 - P_{\text{min}}^D)^N}\right)\right), = 1 - \exp\left(\ln\left(1 - \left(1 - P_{\text{min}}^D\right)^N\right)\right), \geq \prod_{i=1}^{N} \left(1 - P_i^D\right) = 1 - \underbrace{\left[1 - \prod_{i=1}^{N} \left(1 - P_i^D\right)\right]}_{P_d^D}.$$

The result follows by substituting $P_c^D = 1 - \Pr[\Lambda_c \leq \tau_c | H_1]$ in the above inequality. ■

Proof of Theorem 5.1

By recursively expanding the equality constraint of the optimization problem (P.2) we have

$$\begin{bmatrix} \overline{x}(1) \\ \vdots \\ \overline{x}(T) \end{bmatrix} = \mathcal{A} x(0) + \mathcal{B}_a U^a$$

By using the above identity, (P.2) can also be expressed as

$$\max_{U^a} \quad \underbrace{[\mathcal{A}x(0) + \mathcal{B}_a U^a]^{\mathsf{T}} [\mathcal{A}x(0) + \mathcal{B}_a U^a]}_{f(U^a)} \qquad \text{subject to} \quad (U^a)^{\mathsf{T}} M_c(U^a) \leq \widetilde{\delta}_c.$$

From the first-order necessary conditions [32] we now have

$$\nabla\left(f(U_c^*) - \gamma (U_c^*)^{\mathsf{T}} M_c(U_c^*)\right) = 0, \tag{36a}$$

$$\gamma\left(\widetilde{\delta}_c - (U_c^*)^{\mathsf{T}} M_c(U_c^*)\right) = 0, \tag{36b}$$

$$\gamma \geq 0, \tag{36c}$$

$$(U_c^*)^{\mathsf{T}} M_c(U_c^*) \leq \widetilde{\delta}_c, \tag{36d}$$

where the gradient ∇ is with respect to U^a.

Case (i): Suppose $(U_c^*)^\mathsf{T} M_c(U_c^*) < \widetilde{\delta}_c$. Then $\gamma = 0$ should hold true to ensure the complementarity slackness condition (36b). Using these observations in the KKT conditions we now have $\nabla f(U_c^*) = 0$. Further, since, $f(U^a)$ is a convex function of U^a, by evaluating the second derivative of $f(U^a)$ at U_c^*, it can be easily seen that the obtained U_c^* results in minimum value of (P.2) rather than the maximum.Thus, for any U_c^* of (P.2), the condition $(U_c^*)^\mathsf{T} M_c(U_c^*) < \widetilde{\delta}_c$ cannot hold true.

Case (ii): Suppose $(U_c^*)^\mathsf{T} M_c(U_c^*) = \widetilde{\delta}_c$. Then the KKT conditions can be simplified to the following:

$$\nabla \left(f(U_c^*) - \gamma (U_c^*)^\mathsf{T} M_c(U_c^*) \right) = 0,$$
$$(U_c^*)^\mathsf{T} M_c(U_c^*) = \widetilde{\delta}_c.$$

The result now follows by evaluating the derivative on the left hand side of the first equality. ■

Proof of Lemma 5.2

By substituting $x(0) = 0$ in (21a), we note that any optimal attack U_c^* is of the form $k\nu$, where ν is the generalized eigenvector of the pair $(\mathcal{B}_a^\mathsf{T}\mathcal{B}, M_c)$ [28], and the scalar $k = \sqrt{\widetilde{\delta}_c / \nu^\mathsf{T} M_c \nu}$ is obtained from (21b). Let J_c be the optimal cost associated with an attack of the form $U_c^* = k\nu$. Then,

$$J_c = (k\nu)^\mathsf{T} \mathcal{B}_a^\mathsf{T} \mathcal{B}_a (k\nu) = \gamma (k\nu)^\mathsf{T} M_c (k\nu) = \gamma \widetilde{\delta}_c,$$

where the first equality follows from the fact that the objective function $\sum_{k=1}^{T} \overline{x}^\mathsf{T}(k)\overline{x}(k)$ in (P.2) can be expressed as $(U_c^*)^\mathsf{T} \mathcal{B}_a^\mathsf{T} \mathcal{B}_a (U_c^*)$, and the second equality follows from (21a). Since ν is a generalized eigenvector of the pair $(\mathcal{B}_a^\mathsf{T}\mathcal{B}, M_c)$, it follows that γ is the eigenvalue corresponding to ν and hence, J_c is maximized when γ is maximum, which is obtained for $v = v^\star$. The result follows since, $\gamma = \rho_{\max}$, for $v = v^\star$. ■

Proposition A.1: Let $\mathcal{O}_i$ $\mathcal{F}_i^{(u)}$ be the observability and impulse response matrices defined in (6). Define

$$\begin{aligned} \mathcal{N}_i^L &= \left\{ z : z^\mathsf{T} \begin{bmatrix} \mathcal{O}_i & \mathcal{F}_i^{(u)} \end{bmatrix} = 0^\mathsf{T} \right\}, \\ \mathcal{N}_{c,i}^L &= \left\{ z : z^\mathsf{T} \mathcal{O}_{c,i} = 0^\mathsf{T} \right\}, \text{ and} \\ \mathcal{N}_c^L &= \bigcup_{i=1}^{N} \mathcal{N}_{c,i}^L. \end{aligned} \tag{37}$$

where $\mathcal{O}_{c,i} = \begin{bmatrix} (C_{c,i}A)^\mathsf{T} & \cdots & \left(C_{c,i}A^T\right)^\mathsf{T} \end{bmatrix}^\mathsf{T}$ and $C_{c,i} = \begin{bmatrix} 0 & \cdots & C_i & \cdots & 0 \end{bmatrix}$. Then, $\mathcal{N}_i^L \subseteq \mathcal{N}_{c,i}^L \subseteq \mathcal{N}_c^L$, for all $i \in \{1, \ldots, N\}$.

Proof: Without loss of generality, let $i = 1$. By definition, the set inclusion $\mathcal{N}_{c,1}^L \subseteq \mathcal{N}_c^L$ is trivial. For the other inclusion, consider the system defined in (3) without the attack and noise, i.e., $x(k+1) = Ax(k)$. Let $x(k) = \begin{bmatrix} x_1^\mathsf{T}(k) & u_1^\mathsf{T}(k) \end{bmatrix}^\mathsf{T}$, where $x_1(k)$ and $u_1(k)$ are the state and the interconnection signal of Subsystem 1. Also, let

$$A = \begin{bmatrix} A_{11} & B_1 \\ \widetilde{B}_1 & \widetilde{A}_{11} \end{bmatrix}. \tag{38}$$

Notice that, $x(k+1) = Ax(k)$ can be decomposed as

$$\begin{aligned} x_1(k+1) &= A_{11} x_1(k) + B_1 u_1(k), \\ u_1(k+1) &= \widetilde{A}_{11} u_1(k) + \widetilde{B}_1 x_1(k). \end{aligned} \tag{39}$$

By letting $\widetilde{C}_1 = \begin{bmatrix} C_1 A_{11} & C_1 B_1 \end{bmatrix}$ and recursively expanding $x_1(k)$ using (39), we have

$$\begin{aligned} C_{c,1} A^k x(0) &= \begin{bmatrix} C_1 & 0 \end{bmatrix} A A^{k-1} x(0) \\ &= \widetilde{C}_1 A^{k-1} x(0) \\ &= \widetilde{C}_1 \begin{bmatrix} x_1(k-1) \\ u_1(k-1) \end{bmatrix} \\ &= \widetilde{C}_1 \begin{bmatrix} A_{11}^{k-1} x_1(0) + \sum_{j=0}^{k-2} A_{11}^{k-2-j} B_1 u_1(j) \\ u_1(k-1) \end{bmatrix} \\ &= C_1 A_{11}^k x_1(0) + \sum_{j=0}^{k-1} C_1 A_{11}^{k-1-j} B_1 u_1(j), \end{aligned} \tag{40}$$

where the second, third, and fourth equalities follows from (38), system $x(k+1) = Ax(k)$, and (39), respectively. By recalling that $\mathcal{O}_{c,1} x(0) = \begin{bmatrix} (C_{c,1}A)^\mathsf{T} & \cdots & \left(C_{c,1}A^T\right)^\mathsf{T} \end{bmatrix}^\mathsf{T} x(0)$, it follows from (40) that

$$\mathcal{O}_{c,1} x(0) = \mathcal{O}_1 x_1(0) + \mathcal{F}_1^{(u)} \begin{bmatrix} u_1^\mathsf{T}(0) & \cdots & u_1^\mathsf{T}(T-1) \end{bmatrix}^\mathsf{T}.$$

Let z be any vector such that $z^\mathsf{T} \begin{bmatrix} \mathcal{O}_1 & \mathcal{F}_1^{(u)} \end{bmatrix} = 0^\mathsf{T}$. Then, z also satisfies $z^\mathsf{T} \mathcal{O}_{c,1} = 0^\mathsf{T}$. Thus, $\mathcal{N}_1^L \subseteq \mathcal{N}_{c,1}^L$. ■

Lemma A.2: ***(Upper bound on P_d^D)*** Let p_i and λ_i be defined as in (14), and τ_i be defined as in (11). Let $p_{\text{sum}} = \sum_{i=1}^N p_i$, $\lambda_{\text{sum}} = \sum_{i=1}^N \lambda_i$, and $\tau_{\min} = \min_{1 \le i \le N} \tau_i$. Then,

$$P_d^D \le \underbrace{\Pr\left[S_d > \tau_{\min} \right]}_{\triangleq \overline{P}_d^D},$$

where $S_d \sim \chi^2(p_{\text{sum}}, \lambda_{\text{sum}})$.

Proof: Consider the following events:

$$\mathcal{V}_i = \left\{ \widetilde{Y}_i^\mathsf{T} \Sigma_i^{-1} \widetilde{Y}_i \ge \tau_i \right\} \text{ for all } i \in \{1, \ldots, N\}, \text{ and}$$
$$\mathcal{V} = \left\{ \sum_{i=1}^{N} \widetilde{Y}_i^\mathsf{T} \Sigma_i^{-1} \widetilde{Y}_i \ge \tau_{\min} \right\},$$

where the event $\mathcal{V}_i$ is associated with the $i-$th local detector's threshold test. From the definition of the above events, it is easy to note that $\bigcup_{i=1}^N \mathcal{V}_i \subseteq \mathcal{V}$. By the monotonicity of the probability measures, it follows that

$$P_d^D \triangleq \Pr\left[\bigcup_{i=1}^{N} \mathcal{V}_i \mid H_1 \right] \le \Pr\left[\mathcal{V} \mid H_1 \right].$$

From the reproducibility property of the noncentral chi-squared distribution [33], it now follows that $\sum_{i=1}^N \widetilde{Y}_i^\mathsf{T} \Sigma_i^{-1} \widetilde{Y}_i$ equals S_d in distribution and hence, $\Pr[\mathcal{V}|H_1] = \Pr[S_d > \tau_{\min}]$. ■

Lemma A.3: ***(Exponential bounds on the tails of $\chi^2(p,\lambda)$)*** Let $Y \sim \chi^2(p, \lambda)$, $\mu = p + \lambda$, $\sigma = \sqrt{2(p+2\lambda)}$. For all $x > 0$,

$$\Pr\left[Y \ge \mu + \sigma\sqrt{2x} + 2x \right] \le \exp(-x) \tag{41a}$$
$$\Pr\left[Y \le \mu - \sigma\sqrt{2x} \right] \le \exp(-x) \tag{41b}$$

Proof: See [34]. ■