Centralized Versus Decentralized Detection of Attacks in Stochastic Interconnected Systems

Rajasekhar Anguluri, Vaibhav Katewa, and Fabio Pasqualetti

***Abstract*—We consider a security problem for interconnected systems with linear, discrete, time-invariant, stochastic dynamics, where the objective is to detect exogenous attacks by processing measurements at different locations. We consider 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. Hence, it may be possible for the decentralized detector to outperform its centralized counterpart, despite having less information about the system dynamics, and this property is due to the nature of the considered attack detection problem (that is, a simple vs composite hypothesis testing problem). Finally, we numerically validate our findings on a power system model.**

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 centralized detectors [1], [2], the use of local detectors would not only be computationally convenient, but 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 tools for the security of cyber-physical systems.

This work was supported in part by awards ARO-71603NSYIP, NSF-1405330, and UCOP-LFR-18-548175. The authors are with the Department of Mechanical Engineering, University of California, Riverside, {ranguluri,vkatewa,fabiopas}@engr.ucr.edu.

Related work: Centralized attack detectors have been the subject of extensive research in the last years [3]–[7], 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 [8]–[10].

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 [11]–[14]. Methods to improve the detection performance by exchanging information among the local detectors have also been proposed [15], [16]. 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 decentralized and distributed schemes should be employed for attack detection and identification.

Main contributions:[1] This paper features two main contributions. First, we propose particular centralized and decentralized schemes to detect unknown and unmeasurable actuator attacks in stochastic interconnected systems (Section III). 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 considered centralized and decentralized detectors, and show that each detector can outperform the other for certain system and attack configurations (Section IV). We discuss that this phenomenon is inherent with the simple versus composite nature of the considered attack detection problem, and provide numerical examples of this behavior. Finally, we validate our theoretical findings on the IEEE RTS-96 power system model.

[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 approximation. In addition, this paper includes an illustration of the results on a power grid.

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 random variable Y is denoted by $\mathbb{E}[Y]$ and $\mathrm{Cov}[Y]$. 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 $Y \sim \chi^2(p, \lambda)$, $Q(\tau; p, \lambda)$ denotes the complementary cumulative distribution function (CDF) of Y, where $\tau \geq 0$.

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 $i \in \{1, \dots, N\}$. The vectors $x_i \in \mathbb{R}^{n_i}$ and $y_i \in \mathbb{R}^{r_i}$ denote the state and measurements of the i-th subsystem, respectively. The process noise $w_i(k) \sim \mathcal{N}(0, \Sigma_{w_i})$ and the measurement noise $v_i(k) \sim \mathcal{N}(0, \Sigma_{v_i})$, with $\Sigma_{w_i} > 0$ and $\Sigma_{v_i} > 0$, 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 reads as

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

The input $B_i u_i = \sum_{j \neq i}^N A_{ij} x_j$ represents the effect of all subsystems on subsystem i. We refer to B_i and u_i as to the interconnection matrix and 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 $u_i^a \in \mathbb{R}^{m_i}$. Loosely speaking, the matrix B_i^a identifies the states compromised by the attacker in the i-th subsystem, while u_i^a denotes the i-th attack strategy. 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 $x = \begin{bmatrix} x_1^\mathsf{T} & \cdots & x_N^\mathsf{T} \end{bmatrix} \in \mathbb{R}^n$, $w \in \mathbb{R}^n$, $u^a \in \mathbb{R}^m$, $y \in \mathbb{R}^r$, $v \in \mathbb{R}^r$, $n = \sum_{i=1}^N n_i$, $m = \sum_{i=1}^N m_i$, and $r = \sum_i^N r_i$. Moreover, as the components of the vectors w and v are independent and Gaussian, it holds $w \sim \mathcal{N}(0, \Sigma_w)$ 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,

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{bmatrix},$$

$C = \mathrm{blkdiag}\left(C_1, \dots, C_N\right)$ 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)*** Without loss of generality, known control inputs have been omitted from (2) and (3), because their effect can always be subtracted from the measurements. Further, because the detectors do not have any information about the initial state and attack signals, we let these quantities be deterministic and unknown. □

III. Local, decentralized, and centralized detectors

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

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

the measurements available to detector i, 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 of the centralized detector. Local and centralized detectors perform three operations:

1) collect measurements as in (4) and (5), respectively;
2) process the measurements to filter unknown variables;
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.

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 to eliminate their dependency on the unknown parameters. Using equations (1) and (2), define the i-th observability matrix and the i-th attack, interconnection, and noise forced response matrices 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, using (3), define the matrices $\mathcal{O}_c$, $\mathcal{F}_a^w$, and $\mathcal{F}_c^w$ 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

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

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

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 $\mathcal{O}_c$,[2] respectively, and define the processed measurements as

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

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. It should be noticed that $\mathrm{Im}(B_i^a) \subseteq \mathrm{Im}(B_i)$ implies $N_i \mathcal{F}_i^a = 0$. Thus, the processed measurements do not depend on the attack, and our local detection technique cannot be successful against attacks that satisfy this condition. We now characterize the statistical properties of $\widetilde{Y}_i$ and $\widetilde{Y}_c$ (recall that the attack is a deterministic and unknown vector).

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

$$\widetilde{Y}_i \sim \mathcal{N}\left(\beta_i, \Sigma_i\right), \text{ and } \widetilde{Y}_c \sim \mathcal{N}\left(\beta_c, \Sigma_c\right), \quad (9)$$

where $\beta_i = N_i \mathcal{F}_i^a U_i^a$, $\beta_c = N_c \mathcal{F}_c^a U^a$, and

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

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, which are invertible because N_i and N_c are assumed to be of full rank, are independent of the attack. Hence, we develop a detection mechanism based on the mean of the processed measurements.

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 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):

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

where the threshold $\tau_i \geq 0$ is selected based on the desired false alarm probability of the test (11) [17]. 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 given by $P_i^F = \Pr\left[\Lambda_i \geq \tau_i | H_0\right]$ and $P_i^D = \Pr\left[\Lambda_i \geq \tau_i | H_1\right]$, respectively. Similarly, the centralized detector test is defined as

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

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.

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,

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

where

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

[2] Throughout the paper, we assume that the matrices N_i and N_c are nonzero and of full rank. In general, while N_c can be always made nonzero for a sufficiently large horizon T, N_i depends on the number and location of the interconnection signals and sensors. When N_i (resp. N_c) is zero, the detection technique developed in the paper for the i-th subsystem cannot be successful.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2955690, IEEE Transactions on Automatic Control

$M_i = (N_i\mathcal{F}_i^a)^\mathsf{T}\,\Sigma_i^{-1}\,(N_i\mathcal{F}_i^a)$, $M_c = (N_c\mathcal{F}_c^a)^\mathsf{T}\,\Sigma_c^{-1}\,(N_c\mathcal{F}_c^a)$.

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 CDF of a central chi-squared distribution. The parameters p_i, p_c and λ_i, λ_c in Lemma 3.2 are the *degrees of freedom* and *non-centrality* parameters.

Remark 2: ***(System theoretic interpretation of the 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.

(Degrees of freedom) The detection and false alarm probabilities are increasing functions of 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, may not lead to an improvement of the detector performance.

(Non-centrality parameter) The non-centrality parameter 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) [17]. 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) For fixed λ_i and p_i, the detection and false alarm probabilities decrease monotonically with the threshold τ_i. This is due to the fact that the complementary CDFs, which define the detection and false alarm probabilities, are decreasing functions of τ_i. As we show later, because of the contrasting behaviors of the detection and false alarm probabilities with respect to all individual parameters, the decentralized detector can outperform the centralized one. □

We now state a result that provides a relation between the degrees of freedom and the non-centrality parameters of the local and centralized detectors. This result plays a central role in comparing the performance of these 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 and centralized detectors. 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 Detectors

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,

$$\begin{aligned} P_d^D &= \Pr\left[\Lambda_i \geq \tau_i, \text{ for some } i \in \{1, \ldots, N\} \mid H_1\right], \\ P_d^F &= \Pr\left[\Lambda_i \geq \tau_i, \text{ for some } i \in \{1, \ldots, N\} \mid H_0\right], \end{aligned} \tag{15}$$

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

Lemma 4.1: ***(Performance of the decentralized detector)*** The detection and false alarm probabilities in (15) satisfy

$$P_d^D = 1 - \prod_{i=1}^{N}\left(1 - P_i^D\right) \text{ and } P_d^F = 1 - \prod_{i=1}^{N}\left(1 - P_i^F\right). \tag{16}$$

A proof of Lemma 4.1 is postponed to the Appendix. It can be shown that, when $P_i^F = P_j^F$ for all $i, j \in \{1, \ldots, N\}$, P_d^F increases with P_i^F and N. To allow for a fair comparison of the detectors, we assume that $P_c^F = P_d^F$. Consequently, for a fixed P_c^F, the probabilities P_i^F satisfy $P_c^F = 1 - \prod_{i=1}^{N}\left(1 - P_i^F\right)$.

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

$$\tau_c \leq p_c + \lambda_c - \sqrt{4N(p_c + 2\lambda_c)\ln\left(1 - P_{\text{max}}^D\right)^{-1}}, \tag{17}$$

where $P_{\text{max}}^D = \max\{P_1^D, \ldots, P_N^D\}$. 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 opposite behavior.

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

$$\begin{aligned} \tau_c \geq p_c + \lambda_c + \sqrt{4\left(p_c + 2\lambda_c\right)\ln\left(1 - (1 - P_{\text{min}}^D)^N\right)^{-1}} \\ + 2\ln\left(1 - (1 - P_{\text{min}}^D)^N\right)^{-1}, \end{aligned} \tag{18}$$

where $P_{\text{min}}^D = \min\{P_1^D, \ldots, P_N^D\}$. 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. From (17) and (18) we note that, depending on decision threshold τ_c, a centralized detector may or may not outperform a decentralized detector. This can be expected, as the Q function, which quantifies the detection probability, is a decreasing function of τ_c (see Remark 2).

To clarify the effect of attack and detection parameters on the detection performance, we express (17) and (18) using the mean and standard deviation of the test statistic (12). Let $\mu_c \triangleq \mathbb{E}\left[\Lambda_c\right] = \lambda_c + p_c$ 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). Thus, (17) and (18) can be rewritten as

$$\tau_c \leq \mu_c - \sigma_c \underbrace{\sqrt{2N\ln\left(1 - P_{\text{max}}^D\right)^{-1}}}_{\triangleq \kappa_c}, \text{ and} \tag{19a}$$

$$\tau_c \geq \mu_c + \sigma_c \underbrace{\sqrt{2\ln\left(1 - (1 - P_{\text{min}}^D)^N\right)^{-1}}}_{\triangleq \kappa_d} + \kappa_d^2. \tag{19b}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2955690, IEEE Transactions on Automatic Control

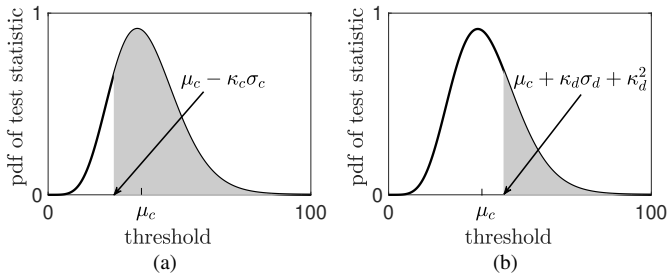

Fig. 1: Probability density function of Λ_c under H_1, as a function of threshold τ_c. For $\tau_c = \mu_c - \kappa_c\sigma_c$ and $\tau_c = \mu_c + \kappa_d\sigma_d + \sigma_d^2$, the shaded area in panels (a) and (b) indicates the detection probability of the centralized detector. As seen in panels (a) and (b), an increase in κ_c results in larger area (larger detection probability) while a increase in κ_d results in smaller area (smaller detection probability).

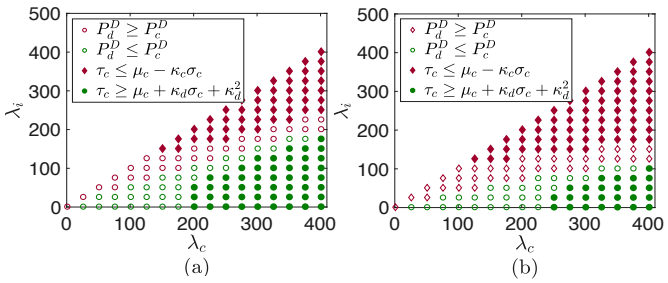

Fig. 2: This figure shows when the decentralized, comprising identical local detectors, and centralized detectors outperform their counterpart, as a function of the non-centrality parameters. The regions identified by solid markers correspond to the conditions in Theorems 4.2 and 4.3. Instead, regions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

From (19a) and (19b) 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, for a decentralized detector to outperform the centralized detector, τ_c should be at least κ_d standard deviations larger than the mean μ_c. See also Fig. 1.

Theorems 4.2 and 4.3 are illustrated in Fig. 2 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, (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. It can also be shown that the difference of the centralized and decentralized detection probabilities can be large, especially when the non-centrality parameters are small and satisfy $\lambda_c \approx \lambda_i$.

Remark 3: ***(Detectors' performance and lack of Uniformly Most Powerful (UMP) test)*** The GLRT is likely not a UMP test for our simple vs. composite attack detection problem and, in fact, a UMP test likely does not exist in this case. To see this, notice that a UMP test does not exist even when the attack vector has length 1; see [17]. Due to the lack of a UMP test for our attack detection problem, the decentralized detector outperforms the centralized one in some cases, even though the latter has more knowledge about the system. Finally, our findings are specific to the considered detectors, and different tradeoffs can be obtained for different detection schemes. □

V. Numerical comparison of centralized and decentralized detectors

In this section, we demonstrate our theoretical findings on the IEEE RTS-96 power network model [18], which we partition into three subregions as shown in [19]. We followed the approach in [19] to obtain a linear time-invariant model of the power network, and then discretized it using a sampling time of 0.01 seconds. For $P_c^F = P_d^F = 0.05$, 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 affect the following results.

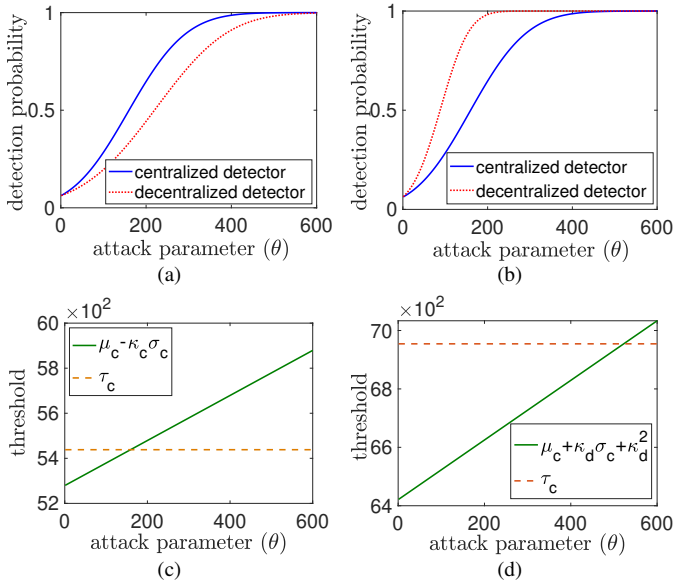

Fig. 3: Scenarios in which the centralized detector outperforms the decentralized detector (a), and vice versa (b), on the IEEE RTS-96 power network, for a range of attack parameter (θ) values. In panels (c) and (d) we plot the right (solid line) and left hand expressions (dashed line) of the inequalities (19a) and (19b), respectively, as a function of θ. For attacks such that the time horizon $T = 100$ sec and $\theta > 200$, the sufficient condition (19a) holds true, it guarantees that $P_c^D \geq P_d^D$. Instead, when $T = 125$ sec and $\theta < 500$, the sufficient condition (19a) holds true, it guarantees that $P_c^D \geq P_d^D$.

(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. 3 show that the detection probabilities of the centralized and decentralized detectors increase monotonically with the attack parameter θ. As predicted by the sufficient condition (19a) and shown in Fig. 3, 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. 3.

(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 parameters, the decentralized detector is guaranteed to outperform the centralized one when $\theta \leq 511$. This behavior is predicted by our sufficient condition (19b), and is illustrated in Fig. 3. The estimation provided by our condition (19b) is conservative, as illustrated in Fig. 3.

VI. Conclusion

In this work we compare the performance of GLRT based centralized and decentralized schemes for the detection of attacks in stochastic interconnected systems. In addition to quantifying the performance of each detector, we prove the counterintuitive result that the decentralized scheme can, at times, outperform its centralized counterpart, and that this behavior is due to the simple versus composite nature of the attack detection problem. We remark that this result holds for the proposed detectors. 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. Dörfler, 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. Z. Lun, A. D'Innocenzo, F. Smarra, I. Malavolta, and M. D. D. Benedetto, "State of the art of cyber-physical systems security: An automatic control perspective," The Journal of Systems and Software, vol. 149, no. 2019, pp. 174–216, 2019.

[3] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," ACM Computing Surveys (CSUR), vol. 51, no. 4, p. 76, 2018.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

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

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

[22] L. Birgé, "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 vectors U_i^a and U^a are deterministic, and W_i, V_i, V, and W are zero mean random vectors, due to linearity, 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) and (b) follows because the measurement and process noises are i.i.d. Similar analysis also results in the expression of Σ_c. Finally, $\widetilde{Y}_i$ and $\widetilde{Y}_c$ are Gaussian because they are result of linear transformation of the Gaussian vectors. ■

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 [20, Theorem 3.3.3] that:

1) under the 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) under the 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 distributional forms of Λ_i and Λ_c, it now follows that the false alarm and the detection probabilities of (11) and (12) are the right tail probabilities of the central and noncentral chi-squared distributions, respectively. Hence, the expressions in (13) follows. ■

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): Let

$$\begin{aligned}\widetilde{\Sigma}_i &= \left[\left(\mathcal{F}_i^w\right)\left(I_T \otimes \Sigma_{w_i}\right)\left(\mathcal{F}_i^w\right)^{\mathsf{T}} + \left(I_T \otimes \Sigma_{v_i}\right)\right] > 0,\\ \widetilde{\Sigma}_c &= \left[\left(\mathcal{F}_c^w\right)\left(I_T \otimes \Sigma_{w}\right)\left(\mathcal{F}_c^w\right)^{\mathsf{T}} + \left(I_T \otimes \Sigma_{v}\right)\right] > 0.\end{aligned} \tag{20}$$

From Lemmas 3.1, 3.2, and (20), we have $p_c = \text{Rank}(\Sigma_c) = \text{Rank}\left(N_c\widetilde{\Sigma}^{1/2}\right) = \text{Rank}\left(N_c\right)$, and, $p_1 = \text{Rank}\left(N_1\right)$. 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 (31)) respectively, from Proposition A.1, $p_1 \leq p_c$.

Case (b): *Step 1 (alternative form of λ_1 and λ_c):* From (14), λ_1 and λ_c can be expressed as $\beta_1^{\mathsf{T}}\Sigma_1^{-1}\beta_1$ and $\beta_c^{\mathsf{T}}\Sigma_c^{-1}\beta_c$, respectively, where β_1, β_c, Σ_1, and Σ_c, that are obtained using expressions in (8), are defined in Lemma 3.1. However, these parameters can be obtained using permuted representation of Y_c (5). To see this, consider i-th sensor measurements of (3)

$$y_{c,i}(k) = C_{c,i}x(k) + v_i(k), \tag{21}$$

where $C_{c,i} = \begin{bmatrix}0 & \cdots & C_i & \cdots & 0\end{bmatrix}$. Let $Y_{c,i} = \begin{bmatrix}y_{c,i}^{\mathsf{T}}(1) & \ldots & y_{c,i}^{\mathsf{T}}(T)\end{bmatrix}^{\mathsf{T}}$. Then, from (21) and (3), we have

$$Y_{c,i} = \mathcal{O}_{c,i}x(0) + \mathcal{F}_{c,i}^a U^a + \mathcal{F}_{c,i}^w W + V_i, \tag{22}$$

where $\mathcal{O}_{c,i}$, $\mathcal{F}_{c,i}^a$, and $\mathcal{F}_{c,i}^w$ are similar to the matrices defined in Section II-A. Finally, from (22), it follows that

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

From the distributional assumptions on W and V, it follows that $\widehat{Y}_c \sim \mathcal{N}\left(\widehat{\mathcal{O}}_c x(0) + \widehat{\mathcal{F}}_c^a U^a, \Sigma\right)$, where $\Sigma = \left(\widehat{\mathcal{F}}_c^w\right)\left(I_T \otimes \Sigma_w\right)\left(\widehat{\mathcal{F}}_c^w\right)^{\mathsf{T}} + \left(I_T \otimes \Sigma_v\right)$.

Now, consider the measurement equation 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 $i \in \{1, \ldots, N\}$ and $k \in \mathbb{N}$. Then, $Y_i = Y_{c,i} = \Pi_i \widehat{Y}_c$ for some matrix Π_i. Let $\widetilde{N}_i = N_i\Pi_i$ and note that $\widetilde{N}_i\widehat{\mathcal{O}} = N_i\mathcal{O}_{c,i}$. From Proposition A.1 and Lemma 3.1, $N_i\mathcal{O}_{c,i} = 0$, and

$$\begin{aligned}\beta_i &\triangleq \mathbb{E}[Y_i] = \Pi_i\mathbb{E}[\widehat{Y}_c] = \widetilde{N}_i\widehat{\mathcal{F}}_c^a U^a, \text{ and}\\ \Sigma_i &\triangleq \text{Cov}[Y_i] = \Pi_i\text{Cov}[\widehat{Y}_c]\Pi_i^{\mathsf{T}} = \widetilde{N}_i\Sigma\widetilde{N}_i^{\mathsf{T}}.\end{aligned} \tag{23}$$

Similarly, 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,

$$\beta_c = N_cQ\widehat{\mathcal{F}}_c^a U^a, \text{ and } \Sigma_c = N_cQ\Sigma(N_cQ)^{\mathsf{T}}. \tag{24}$$

Let $z = \widehat{\mathcal{F}}_c^a U^a$. From (23) and (24) we have

$$\begin{aligned}\lambda_1 &= z^{\mathsf{T}}\widetilde{N}_1^{\mathsf{T}}\left[\widetilde{N}_1\Sigma\widetilde{N}_1^{\mathsf{T}}\right]^{-1}\widetilde{N}_1, \text{ and}\\ \lambda_c &= z^{\mathsf{T}}\left(N_cQ\right)^{\mathsf{T}}\left[\left(N_cQ\right)\Sigma\left(N_cQ\right)^{\mathsf{T}}\right]^{-1}\left(N_cQ\right),\end{aligned} \tag{25}$$

which are the required alternative forms for λ_1 and λ_c in (14).

Step 2 (lower bound on λ_c): Since $Y_c = N_cY_c = N_cQ\widehat{Y}_c$, N_cQ is the basis of the null space of $\widehat{\mathcal{O}}_c$. 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 $N_{c,i} = \begin{bmatrix}N_{c,i}^1 & \cdots & N_{c,i}^N\end{bmatrix}$ 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 $P_1 = \begin{bmatrix}\left(N_{c,2}\right)^{\mathsf{T}} & \cdots & \left(N_{c,N}\right)^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$ and note that

$$\underbrace{\begin{bmatrix}N_{c,1} & P_1\end{bmatrix}\Sigma\begin{bmatrix}N_{c,1}^{\mathsf{T}} & P_1^{\mathsf{T}}\end{bmatrix}}_{(N_cQ)\Sigma(N_cQ)^{\mathsf{T}}} = \begin{bmatrix}S_1 & N_{c,1}\Sigma P_1^{\mathsf{T}}\\ N_{c,1}^{\mathsf{T}}\Sigma P_1 & R\end{bmatrix},$$

where $S_1 = N_{c,1}\Sigma N_{c,1}^{\mathsf{T}}$ and $R = P_1^{\mathsf{T}}\Sigma P_1$. Since $\Sigma > 0$, S_1 and R are invertible, and hence, there exists $X \geq 0$ such that

$$\left[\left(N_cQ\right)\Sigma\left(N_cQ\right)^{\mathsf{T}}\right]^{-1} = \begin{bmatrix}S_1^{-1} & 0\\ 0 & 0\end{bmatrix} + X. \tag{26}$$

Let, $\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$ such that $\Sigma_{11} > 0$ and $\Sigma_{22} > 0$, and define $S_2 = (N_{c,1}^1)\Sigma_{11}(N_{c,1}^1)^{\mathsf{T}}$. By substituting $N_{c,1} = \begin{bmatrix}N_{c,1}^1 & \cdots & N_{c,1}^N\end{bmatrix}$ in S_1, by means of Schur's complement, it follows that

$$S_1^{-1} = \begin{bmatrix}S_2^{-1} & 0\\ 0 & 0\end{bmatrix} + Y, \tag{27}$$

where $Y \geq 0$. Substituting(26) and (27) into (25), we have

$$\begin{aligned}\lambda_c &= z^{\mathsf{T}}(N_cQ)^{\mathsf{T}}\begin{bmatrix}S_1^{-1} & 0\\ 0 & 0\end{bmatrix}(N_cQ)z + \underbrace{z^{\mathsf{T}}(N_cQ)^{\mathsf{T}}X(N_cQ)z}_{\geq 0}\\ &\geq \begin{bmatrix}(N_{c,1}z)^{\mathsf{T}} & (P_1z)^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}S_1^{-1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}N_{c,1}z\\ P_1z\end{bmatrix}\\ &= z^{\mathsf{T}}(N_{c,1})^{\mathsf{T}}\begin{bmatrix}S_2^{-1} & 0\\ 0 & 0\end{bmatrix}(N_{c,1})z + \underbrace{z^{\mathsf{T}}(N_{c,1})^{\mathsf{T}}Y(N_{c,1})z}_{\geq 0}\\ &\geq z^{\mathsf{T}}\begin{bmatrix}(N_{c,1}^1)^{\mathsf{T}}S_2^{-1}N_{c,1}^1 & 0\\ 0 & 0\end{bmatrix}z.\end{aligned} \tag{28}$$

Instead, λ_1 in (25) can be shown to satisfy

$$\lambda_1 = z^{\mathsf{T}}\begin{bmatrix}N_1^{\mathsf{T}}\left[N_1\Sigma_{11}N_1^{\mathsf{T}}\right]^{-1}N_1 & 0\\ 0 & 0\end{bmatrix}z, \tag{29}$$

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

Step 3 (show $\lambda_c \geq \lambda_1$ using the lower bound (28)): To ensure the inequality $\lambda_c \geq \lambda_1$, from (28) and (29), it suffices to show that $(N_{c,1}^1)^{\mathsf{T}}S_2^{-1}N_{c,1}^1 \geq N_1^{\mathsf{T}}\left[N_1\Sigma_{11}N_1^{\mathsf{T}}\right]^{-1}N_1$. From 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, define the invertible matrix $\widetilde{F}_1^{\mathsf{T}} \triangleq \begin{bmatrix}F_1^{\mathsf{T}} & M_1^{\mathsf{T}}\end{bmatrix}$, where M_1 forms a basis of the null space of F_1, such that

$$\underbrace{\widetilde{F}_1^{\mathsf{T}}\left[\widetilde{F}_1S_2\widetilde{F}_1^{\mathsf{T}}\right]^{-1}\widetilde{F}_1}_{S_2^{-1}} = \widetilde{F}_1^{\mathsf{T}}\begin{bmatrix}F_1S_2F_1^{\mathsf{T}} & F_1S_2M_1^{\mathsf{T}}\\ M_1S_2F_1^{\mathsf{T}} & M_1S_2M_1^{\mathsf{T}}\end{bmatrix}^{-1}\widetilde{F}_1,$$

where the inverse term on the right hand side satisfies

$$\begin{bmatrix}F_1S_2F_1^{\mathsf{T}} & F_1S_2M_1^{\mathsf{T}}\\ M_1S_2F_1^{\mathsf{T}} & M_1S_2M_1^{\mathsf{T}}\end{bmatrix}^{-1} = \begin{bmatrix}\left(F_1S_2F_1^{\mathsf{T}}\right)^{-1} & 0\\ 0 & 0\end{bmatrix} + \underbrace{Z}_{\geq 0}.$$

By the above identities, it follows that

$$S_2^{-1} = \widetilde{F}_1^{\mathsf{T}}\begin{bmatrix}\left(F_1S_2F_1^{\mathsf{T}}\right)^{-1} & 0\\ 0 & 0\end{bmatrix}\widetilde{F}_1 + \widetilde{F}_1^{\mathsf{T}}Z\widetilde{F}_1. \tag{30}$$

Let $Z_1 \triangleq (\widetilde{F}_1 N_{c,1}^1)^{\mathsf{T}} Z (\widetilde{F}_1 N_{c,1}^1) \geq 0$, and now consider

$$\begin{aligned}(N_{c,1}^1)^{\mathsf{T}} S_2^{-1} N_{c,1}^1 &\overset{(31)}{=} (\widetilde{F}_1 N_{c,1}^1)^{\mathsf{T}} \begin{bmatrix} \left(F_1 S_2 F_1^{\mathsf{T}}\right)^{-1} 0 \\ 0 \quad 0 \end{bmatrix} \widetilde{F}_1 N_{c,1}^1 + Z_1 \\ &\overset{a}{=} (F_1 N_{c,1}^1)^{\mathsf{T}} \left(F_1 S_2 F_1^{\mathsf{T}}\right)^{-1} F_1 N_{c,1}^1 + Z_1 \\ &\overset{b}{=} N_1^{\mathsf{T}} \left[N_1 \Sigma_{11} N_1^{\mathsf{T}}\right]^{-1} N_1 + Z_1\end{aligned}$$

where (a) follows because $\widetilde{F}_1^{\mathsf{T}} = [F_1^{\mathsf{T}} \; M_1^{\mathsf{T}}]$, and (b) by substituting $S_2 = (N_{c,1}^1)\Sigma_{11}(N_{c,1}^1)^{\mathsf{T}}$ and $N_1 = F_1 N_{c,1}^1$. Since, $Z_1 \geq 0$, it follows that $(N_{c,1}^1)^{\mathsf{T}} S_2^{-1} N_{c,1}^1 > N_1^{\mathsf{T}} \left[N_1 \Sigma_{11} N_1^{\mathsf{T}}\right]^{-1} N_1$, which implies that $\lambda_c \geq \lambda_1$, as required. ∎

Proof of Lemma 4.1: Let $\mathcal{E}_i$ be an event that the i-th local detector decides H_1 when H_0 is true. Then, $P_i^F = \Pr[\mathcal{E}_i]$. Let $\mathcal{E}_i^{\complement}$ be the complement of $\mathcal{E}_i$. Then, from (15) we have

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

where (a) follows because $\mathcal{E}_i$'s are mutually independent for all $i \in \{1, \dots N\}$. To see this fact, notice that $\mathcal{E}_i$ depends only on $\widetilde{Y}_i$ (see (8)). Further, $\widetilde{Y}_i$ depends on the non-random attack U_i^a and the noise vectors V_i and W_i, but not on the interconnection signal U_i (see (6)). Since, V_i and W_i across subsystems are independent, $\mathcal{E}_i$ are also mutually independent. Similar procedure will lead to the expression for P_d^D. ∎

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 (17) holds true. Then,

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

Since $\Lambda_c \sim \chi^2(\lambda_c, p_c)$, from the inequality (34b), we have

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

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

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

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 (18) holds true. Then,

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

Since $\Lambda_c \sim \chi^2(\lambda_c, p_c)$, from the inequality (34a) we have

$$\begin{aligned}1 - P_c^D = \Pr[\Lambda_c \leq \tau_c] &\geq 1 - \exp\left(-\ln\left(1 - (1 - P_{\min}^D)^N\right)^{-1}\right) \\ &\geq \prod_{i=1}^{N} \left(1 - P_i^D\right) = 1 - P_d^D. \quad ∎\end{aligned}$$

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

$$\begin{aligned}\mathcal{N}_i^L &= \left\{z : z^{\mathsf{T}} \begin{bmatrix}\mathcal{O}_i & \mathcal{F}_i^u\end{bmatrix} = 0\right\}, \text{and} \\ \mathcal{N}_{c,i}^L &= \left\{z : z^{\mathsf{T}} \mathcal{O}_{c,i} = 0\right\},\end{aligned} \tag{31}$$

where $\mathcal{O}_{c,i} = \left[(C_{c,i}A)^{\mathsf{T}} \quad \cdots \quad \left(C_{c,i}A^T\right)^{\mathsf{T}}\right]^{\mathsf{T}}$ 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 $i \in \{1, \dots, N\}$, where $\mathcal{N}_c^L = \bigcup_{i=1}^N \mathcal{N}_{c,i}^L$.

Proof: Without loss of generality, let $i = 1$. By definition, the inclusion $\mathcal{N}_{c,1}^L \subseteq \mathcal{N}_c^L$ is trivial. For the other inclusion, consider an auxiliary system $x(k+1) = Ax(k)$. 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

$$A = \begin{bmatrix}A_{11} & B_1 \\ \widetilde{B}_1 & \widetilde{A}_{11}\end{bmatrix}. \tag{32}$$

Then, the state $x(k+1)$ is decomposed into $x_1(k+1) = A_{11}x_1(k) + B_1 u_1(k)$ and $u_1(k+1) = \widetilde{A}_{11}u_1(k) + \widetilde{B}_1 x_1(k)$. By letting $\widetilde{C}_1 = \begin{bmatrix}C_1 A_{11} & C_1 B_1\end{bmatrix}$ it follows that

$$\begin{aligned}C_{c,1}A^k x(0) &= \begin{bmatrix}C_1 & 0\end{bmatrix} A A^{k-1} x(0) = \widetilde{C}_1 \begin{bmatrix}x_1(k-1) \\ u_1(k-1)\end{bmatrix} \\ &= C_1 A_{11}^k x_1(0) + \sum_{j=0}^{k-1} C_1 A_{11}^{k-1-j} B_1 u_1(j), \quad (33)\end{aligned}$$

where the second, third, and fourth equality follows from (32), system $x(k+1) = Ax(k)$, and the decomposition equations, respectively. By invoking definition of $\mathcal{O}_{c,1}$ in (33), we have

$$\mathcal{O}_{c,1}x(0) = \mathcal{O}_1 x_1(0) + \mathcal{F}_1^u \left[u_1^{\mathsf{T}}(0) \quad \cdots \quad u_1^{\mathsf{T}}(T-1)\right]^{\mathsf{T}}.$$

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 $z^{\mathsf{T}}\mathcal{O}_{c,1} = 0^{\mathsf{T}}$, which implies $\mathcal{N}_1^L \subseteq \mathcal{N}_{c,1}^L$. ∎

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 $\tau_{\min} = \min_{1 \leq i \leq N} \tau_i$. Then, $P_d^D \leq \Pr[S_d > \tau_{\min}]$, where $S_d \sim \chi^2(p_{\text{sum}}, \lambda_{\text{sum}})$.

Proof: Consider the events $\mathcal{V}_i = \left\{\widetilde{Y}_i^{\mathsf{T}} \Sigma_i^{-1} \widetilde{Y}_i \geq \tau_i\right\}$, for all $i \in \{1, \dots, N\}$, and $\mathcal{V} = \left\{\sum_{i=1}^N \widetilde{Y}_i^{\mathsf{T}} \Sigma_i^{-1} \widetilde{Y}_i \geq \tau_{\min}\right\}$. The event $\mathcal{V}_i$ is associated with the i-th local detector's threshold test. By observing that $\bigcup_{i=1}^N \mathcal{V}_i \subseteq \mathcal{V}$, the inequality $P_d^D \triangleq \Pr\left[\bigcup_{i=1}^N \mathcal{V}_i \mid H_1\right] \leq \Pr[\mathcal{V} \mid H_1]$ is obvious. Now, from the reproducibility property of the non central chi-squared distribution [21], 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: ***(Tight 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 $x > 0$,

$$\Pr\left[Y \geq \mu + \sigma\sqrt{2x} + 2x\right] \leq \exp(-x) \tag{34a}$$

$$\Pr\left[Y \leq \mu - \sigma\sqrt{2x}\right] \leq \exp(-x) \tag{34b}$$

Proof: See [22]. ∎