Deflection-based Attack Detection for Network Systems

Rajasekhar Anguluri and Fabio Pasqualetti

Abstract— **This paper considers a deflection-based detector for random attacks compromising the inputs of a network system, using measurements from nodes non-collocated with the input nodes. We derive the decision rule of our deflection-based detector, and characterize its performance as a function of the edge weights, attack and noise statistics, and the locations of input and output nodes. In the asymptotic measurement regime, we show that the detector's performance is governed by the singular values of the system's transfer function matrix. Finally, for a given input and output node locations, we numerically solve an optimization problem to find the optimal network edge weights that maximize the detector's performance. Numerical examples are presented to validate the theoretical results.**

I. Introduction

For safe, reliable, and efficient operation of cyber-physical systems (e.g., power and transportation systems, and medical devices), there is an increasing need to develop sophisticated detectors to decide against attacks with complex behaviors. A few notable works in this direction include [1]–[3], where the authors study the role of system dynamics, number of compromised sensors/actuators, and attack magnitude and type on the detection performance. As many cyber-physical systems also have an in-built network structure, a few studies also focused on understanding the role of network topology (edge connectivity and magnitude of weights) and location of the input and output nodes on the attack detector's performance from a graph-theoretic perspective [4]–[6]. These studies highlight what structural properties of a network (e.g., directed or undirected edges, positive or negative edge weights, or presence of community structures) play a crucial role in modulating the detection performance.

Motivated by the above studies, in this paper we propose a deflection based attack detector for detecting random attacks compromising inputs of a network (dynamical) system. Without assuming proper knowledge on the probability distribution of the attacks, we characterize the performance of the deflection-based detector and show that its performance can be improved by designing the edge weights of the network. Overall, for random attacks on inputs, our results show that the deflection-based detector can be used as an alternative to the existing generalized likelihood ratio-based detectors, which depends on the probability distribution of the attack and also the nominal system dynamics.

Related work: In the last few years, with the increasing need for detecting attacks with different characteristics (eg., replay attack, false data injection attack, integrity attack, and stealthy attack), researchers have proposed several attack detection strategies and studied them extensively [1]–[3]. Notice that these strategies primarily differs depending on the presence or absence of stochastic noise in the system. For a comprehensive summary on various detection methods, see [7] and the references therein.

In the context of attack detection in linear dynamical systems with Gaussian noise, *chi-squared* detectors are well studied [8]; however, these detectors only work for deterministic attacks, and cannot be extended for random attacks. Studies on network design for attack detection are limited. A notable work in this direction is [9], where the authors design an optimal communication network to implement global detectors using local detectors. Instead, we consider network design for maximizing the performance of a global input attack detector. Finally, we note that there are several studies on network design for non-detection based applications, including disturbance rejection [10], network coherence [11], and sensor placement [12].

Contribution: The contribution of our work is two-fold. First, we develop a deflection-based detector to detect random attacks compromising the inputs of a network (dynamical) system. Both in the finite and asymptotic measurement regimes, we characterize the attack detector's performance as a function of the network edge weights, attack and noise statistics, and the input and output node locations. Our attack detector relies on the attack's covariance matrix, thereby making it amenable to the attacks following arbitrary probability distributions. Second, for fixed input and output node locations, we formulate and numerically solve a non-convex optimization problem to choose the network weights that maximize the detector's performance. Finally, via numerical examples we show that the detection performance on directed networks is higher than that of the undirected ones.

Organization: The rest of the paper is organized as follows. Section II introduces the network system and our deflection-based attack detector. Section III characterizes the deflection-based detector's performance, and proposes an optimization framework for network design. Section VI contains illustrative numerical examples. Section V concludes the paper.

Notation: For any two $n \times n$ real matrices M and N, the operation $M \geq N$ denotes $M_{ij} \geq N_{ij}$, for all $i, j \in \{1, \dots, n\}$. A $n \times n$ symmetric positive (resp. semi) definite matrix M is denoted by $M \succ 0$ (resp. $M \succeq 0$), and $M \succeq N$ if $M - N \succeq 0$. The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ denotes the standard basis vectors of $\mathbb{R}^n$. The probability of an event $\mathcal{E}$ is denoted by $\Pr[\mathcal{E}]$. The expectation and variance of a random variable (vector) X is denoted by $\mathbb{E}[X]$ and $\mathrm{Var}[X]$, respectively.

This material is based upon work supported in part by awards UCOP-LFR-18-548175, ARO-71603NSYIP, and AFOSR-FA9550-19-1-0235. Rajasekhar Anguluri is with the School of Electrical, Computer and Energy Engineering, Tempe, Arizona State University, rangulur@asu.edu. Fabio Pasqualetti is with the Department of Mechanical Engineering, University of California, Riverside, fabiopas@engr.ucr.edu.

II. Problem setup and preliminary notions

A. Attacked system model

Consider a network of n nodes represented by the directed graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where the node and edge sets are given by $\mathcal{V} := \{1, \dots, n\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, respectively. Let $a_{ij} \in \mathbb{R}$ be the weight assigned to the edge $(i, j) \in \mathcal{E}$, and define the *weighted adjacency matrix* of $\mathcal{G}$ as $A := [a_{ij}]$, where $a_{ij} = 0$ whenever $(i, j) \notin \mathcal{E}$. If $A = A^{\mathsf{T}}$, we refer the underlying graph as un-directed network. Let $\mathcal{K} := \{k_1, \dots, k_r\} \subseteq \mathcal{V}$ be the set of input nodes, which receive r inputs. For the i-th node, we assign a state $x_i \in \mathbb{R}$, and let the network evolve with linear time-invariant dynamics

$$x[k+1] = Ax[k] + Bw[k], \tag{1}$$

where $x = [x_1 \cdots x_n]^T \in \mathbb{R}^n$ contains the states, the initial state $x[0] = 0$, and $w[k] \in \mathbb{R}^r$ is the input. The matrix $B = [\mathbf{e}_{k_1}, \dots, \mathbf{e}_{k_r}]$ indicates the nodes through which the inputs excite the network. The input $w[k]$ follows one of the two statistical hypotheses:

$$\begin{aligned} &\text{(attack absent) } H_0 : \ w[k] = u[k], \\ &\text{(attack present) } H_1 : \ w[k] = u[k] + a[k]. \end{aligned} \tag{2}$$

Here, $u[k]$ is an independent (across time) stochastic process distributed according to $\mathcal{N}(0, \Sigma_0)$, and Σ_0 is known. Thus, under H_0, a known Gaussian process drives the network.

Under the alternative H_1, the attack $a[k]$ is assumed to be a zero mean temporally uncorrelated (need not be independent) stochastic process with unknown probability distribution and marginal covariance matrix $\Sigma_a := \text{Cov}[a[k]] = \text{Cov}[a[l]]$, for all $k \neq l$. Our modeling framework allows attacks to follow any arbitrary probability distribution. In fact, $a[k]$ can be a continuous or discrete random vector (see Remark 1).

We assume that the network dynamics (1) are measured at m nodes (not necessarily collocated with the input nodes set $\mathcal{K}$). The measurements $y[k] \in \mathbb{R}^m$ from these sensors obey

$$y[k] = Cx[k] + v[k], \tag{3}$$

where $C = [\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_m}]^{\mathsf{T}}$ and $v[k] \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Sigma_v)$. Finally, under the hypothesis H_0 and H_1, we assume that any finite set of random vectors in $\{a[0], v[0], u[0], a[1], v[1], u[1], \dots\}$ are uncorrelated (not necessarily independent).

B. Deflection criterion for attack detection

As the attack's probability distribution is unknown in our setup, the likelihood ratio-based attack detectors [8], [13] are not only difficult to implement, but are hard to analyze using the detection probability. To overcome the above problem, we consider a detector based on a signal-to-noise ratio based criterion called *deflection* [14], [15]. We now formally state the deflection-based attack detector.

For compactness, let us denote the system and measurement matrices by $\Omega := (A, B, C)$. For a fixed sample size N, let $Y_N = [y[1]^{\mathsf{T}}, \dots, y[N]^{\mathsf{T}}]^{\mathsf{T}}$ be the measurements collected over the horizon $k = 1, \dots, N$. Define the threshold test :[1]

$$\begin{aligned} \delta(Y_N) < \tau : &\quad \text{attack absent } (H_0), \\ \delta(Y_N) \geq \tau : &\quad \text{attack present } (H_1). \end{aligned} \tag{4}$$

In the deflection test, one defines a test statistic $\delta(Y_N) = Y_N^{\mathsf{T}} Q Y_N$, for $Q \succ 0$, that maximizes the deflection criterion:

$$D(Q, \Omega, \Sigma_a) = \frac{\left(\mathbb{E}_1[\delta(Y_N)] - \mathbb{E}_0[\delta(Y_N)]\right)^2}{\text{Var}_0[\delta(Y_N)]} \tag{5}$$

Here, $\mathbb{E}_i[\cdot]$ and $\text{Var}_i[\cdot]$, $i \in \{0, 1\}$, are the mean and variance of $\delta(Y_M)$ evaluated under hypothesis H_i. However, as the attack covariance is unknown, we cannot evaluate $\mathbb{E}_1[\delta(Y_N)]$. Thus, we consider a robust version of (5)—referred to as the *worst case deflection criterion* (or WDC).

Assumption 2.1: The attack covariance matrix Σ_a is uncertain, but is lower bounded by a known positive definite matrix (Σ_ℓ), i.e., $0 \prec \Sigma_\ell \preceq \Sigma_a$.

Assumption 2.1 ensures that the attack signature cannot be arbitrarily low. Thus, any reasonable test (4) have, at least, some power to detect attacks. These types of assumptions are standard in the theory of robust detection [16].

Definition 1: ***(Optimal WDC based detector)*** The optimal WDC based detector is the test (4) for which the test statistic is $\delta(Y_N) = Y_N^{\mathsf{T}} Q^* Y_N$, where

$$Q^* = \arg\max_{Q \succ 0} \ D_W(Q, \Omega), \tag{6}$$

and the WDC is defined[2] as

$$D_W(Q, \Omega) = \inf_{0 \prec \Sigma_\ell \preceq \Sigma_a} D(Q, \Omega, \Sigma_a). \tag{7}$$

Thus, the performance of the optimal WDC based detector is given by $D_W(Q^*, \Omega)$. Once Q^* is obtained, the threshold (τ) needed to perform the test (4) can be computed by prefixing the false alarm probability $P_F = \Pr[\delta(Y_N) \geq \tau | H_0]$. Finally, the asymptotic optimal WDC is given by

$$\overline{D}_W(Q^*, \Omega) = \limsup_{N \to \infty} \frac{D_W(Q^*, \Omega)}{N}. \tag{8}$$

Notice that the asymptotic measure (8) is similar to that of *error exponent* in the case of the Neyman-Pearson detector.

C. The network design problem

In this paper, for a given input and output nodal locations (i.e., for fixed a matrices C and B), our goal is to select the edge weights (or matrix A) such that the optimal WDC (8) is maximized. This can be cast as an optimization problem:

$$\begin{aligned} \underset{A \in \mathbb{M}^{n \times n}}{\text{maximize}} \quad & D_W(Q^*, \Omega) \\ \text{subject to} \quad & \rho(A) < 1 \quad \text{(stability constraint)}, \\ & |\mathcal{E}| \leq K \quad \text{(budget constraint)}, \end{aligned} \tag{9}$$

where $\mathbb{M} \subseteq \mathbb{R}^{n \times n}$ denotes the set of adjacency matrices over which the optimization is carried out. This restriction

[1] Throughout the paper, we will use the terms "detector" and "threshold test" interchangeably.

[2] The expression (7) depends on the statistics of the input, attack, sensor noise, and the matrices (Q, A, B, C). However, to avoid cluttered notation, some of these parameters are suppressed in the definition of D_W.

allows us to impose structural constraints on the network. For instance, by imposing $\mathbb{M}$ to be symmetric, we get undirected networks, or by letting $\mathbb{M} = \{0,1\}^{n\times n}$, we get binary networks. Here, $\rho(A)$denotes the spectral radius of A. The budget constraint regulates the total number of connected edges. The numerical solution of this problem is deferred until Section V.

Remark 1: ***(Mean of the attack)*** The zero mean assumption of the attack is to obtain simplified and intuitive expressions of the detector's performance measure [17]. With an extra effort, the results in this paper can be extended to the non-zero mean case by considering the test statistic $\delta(Y_N)$ that contains both the linear and quadratic term [15], [18].

III. Optimal worst case deflection based detector

In this section we solve the maximization problem (6) to obtain an expression for Q^* as a function of the system matrices Ω, and the input, attack, and noise covariance matrices. We also characterize the expressions for finite and asymptotic optimal WDC as well. For a given $\Omega = (A, B, C)$, define the following impulse response matrix:

$$F = \begin{bmatrix} CB & 0 & \dots & 0 \\ CAB & CB & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-1}B & CA^{N-2}B & \dots & CB \end{bmatrix}. \quad (10)$$

The following result is instrumental for computing Q^* (6).

Proposition 3.1: ***(Mean and variance of*** $\delta(Y_N)$***)*** For $i \in \{0,1\}$, let the hypothesis H_i be as in (2). Let $\delta(Y_N) = Y_N^{\mathsf{T}} Q Y_N$ for some $Q \succ 0$. Then

$$\mathbb{E}_i[\delta(Y_N)] = \mathrm{tr}\,(QS_i) \text{ and } \mathrm{Var}_0[\delta(Y_N)] = 2\mathrm{tr}[(QS_0)^2], \quad (11)$$

where $S_0 = F\,(I \otimes \Sigma_0)\,F^{\mathsf{T}} + (I \otimes \Sigma_v)$ and $S_1 = S_0 + F\,(I \otimes \Sigma_a)\,F^{\mathsf{T}}$.

Proof: Expand the state vector $x[k]$ (1) and substitute it in measurement equation (3) to obtain $Y_N = \Theta W_N + V_N$, where $W_N = [x^{\mathsf{T}}[0], w^{\mathsf{T}}[0], \dots, w^{\mathsf{T}}[N-1]]^{\mathsf{T}}$ and $V_N = [v^{\mathsf{T}}[1]], \dots, v^{\mathsf{T}}[N]]^{\mathsf{T}}$. From the statistical description of $w[k]$ and $v[k]$, we have that $\mathbb{E}_i[Y_N] = 0$ and $\mathrm{Var}_0[Y_N] = S_0$. From these observations and the fact that $\delta(Y_N) = Y_N^{\mathsf{T}} Q Y_N$, the expressions of $\mathbb{E}_i[\delta(Y_N)]$ and $\mathrm{Var}_0[\delta(Y_N)]$ in (11) can be obtained by invoking Corollary 5.1 in [19]. ∎

Lemma 3.2: ***(Worst case deflection criterion)*** Let S_0 be defined as in the statement of Proposition 3.1. Then

$$D_W(Q,\Omega) = \frac{\left(\mathrm{tr}[QF(I\otimes\Sigma_l)F^{\mathsf{T}}]\right)^2}{2\mathrm{tr}[(QS_0)^2]}, \quad (12)$$

where Σ_ℓ are defined in Assumption 2.1.

Proof: By Substituting (11) in $D_W(Q,\Omega,\Sigma)$ (5), followed by few algebraic manipulations, we note that

$$D(Q,\Omega,\Sigma_a) = \frac{\left(\mathrm{tr}\left[QF(I\otimes\Sigma_a)F^{\mathsf{T}}\right]\right)^2}{2\mathrm{tr}[(QS_0)^2]}. \quad (13)$$

Now, consider the following inequality:

$$\begin{aligned} \frac{\left(\mathrm{tr}\left[QF(I\otimes\Sigma_\ell)F^{\mathsf{T}}\right]\right)^2}{2\mathrm{tr}[(QS_0)^2]} &\geq \frac{\inf_{\Sigma_\ell\preceq\Sigma_a}\left(\mathrm{tr}\left[QF(I\otimes\Sigma_a)F^{\mathsf{T}}\right]\right)^2}{2\mathrm{tr}[(QS_0)^2]} \\ &= \inf_{\Sigma_\ell\preceq\Sigma_a} \frac{\left(\mathrm{tr}\left[QF(I\otimes\Sigma_a)F^{\mathsf{T}}\right]\right)^2}{2\mathrm{tr}[(QS_0)^2]} \\ &= \inf_{\Sigma_\ell\preceq\Sigma_a} D(Q,\Omega,\Sigma_a) \\ &= D_W(Q,\Omega). \end{aligned}$$

The first inequality follows by noting that $QF(I\otimes\Sigma_l)F^{\mathsf{T}} \preceq QF(I\otimes\Sigma_a)F^{\mathsf{T}}$ (since $Q \succ 0$ and $\Sigma_\ell \preceq \Sigma_a$), and trace is an operator monotone function. The second equality because the term $2\mathrm{tr}[(QS_0)^2$ is independent of Σ_a. The third from the expression in (13), and the last one by definition. ∎

From (6) and (12), the optimal WDC is given by

$$D_W(Q^*,\Omega) = \underset{Q\succ 0}{\text{maximize}} \frac{\left(\mathrm{tr}\left[QF(I\otimes\Sigma_a)F^{\mathsf{T}}\right]\right)^2}{2\mathrm{tr}[(QS_0)^2]}. \quad (14)$$

Equation (14) is similar to the frame-theoretic notion of measure of quality—a qualitative measure for the controllability and observability of a linear dynamical system [20]. Instead, we note that the optimal WDC measures the ability of a linear dynamical system to distinguish between the hypotheses H_0 (attack) and H_1 (no attack).

Lemma 3.3: ***(Optimal worst case deflection criterion)*** Suppose the Assumption 2.1 holds true. Then, the optimal Q^* (6) is given by $S_0^{-1}F(I\otimes\Sigma_l)F^{\mathsf{T}}S_0^{-1}$. Moreover,

$$D_W(Q^*,\Omega) = \frac{1}{2}\,\mathrm{tr}[(F(I\otimes\Sigma_l)F^{\mathsf{T}}S_0^{-1})^2]. \quad (15)$$

Proof: The proof is known in the literature. We sketch few details for completeness. From (12) notice that

$$Q^* = \arg\max_{Q\succ 0} \frac{\left(\mathrm{tr}[QF(I\otimes\Sigma_l)F^{\mathsf{T}}]\right)^2}{2\mathrm{tr}[(QS_0)^2]}.$$

Let $H_1 = S_0^{-\frac{1}{2}}F(I\otimes\Sigma_l)F^{\mathsf{T}}S_0^{-\frac{1}{2}}$ and $H_2 = S_0^{-\frac{1}{2}}QS_0^{-\frac{1}{2}}$. Notice that $|\mathrm{tr}(H_1H_2)|^2 \leq \mathrm{tr}(H_1H_2^{\mathsf{T}})\mathrm{tr}(H_2H_2^{\mathsf{T}})$, with equality iff $H_1 = \alpha H_2$ for some $\alpha \in R$. It follows that $D_W(Q,\Omega) \leq 0.5\mathrm{tr}[(F(I\otimes\Sigma_l)F^{\mathsf{T}}S_0^{-1})^2]$, for any $Q \succ 0$. To complete the proof, let $\alpha = 1$ in $H_1 = \alpha H_2$ and then solve for Q^*. ∎

We consider specific structures of input and attack covariance matrices that results in simplified expression of (15).

Corollary 3.4: ***(Structured input and attack covariance matrices)*** Let $\Sigma_0 = \sigma_0^2 M$ and $\Sigma_\ell = \sigma_\ell^2 M$, for some $M \succ 0$. Let $\widetilde{F} = F(I\otimes M^{\frac{1}{2}})$ and $d = \min\{m, r\}$. Then

$$D_W(Q^*,\Omega) = \frac{1}{2}\sum_{i=1}^{Nd}\left[\frac{\sigma_\ell^2\, s_i^2(\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F})}{1+\sigma_0^2\, s_i^2(\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F})}\right]^2, \quad (16)$$

where $\widetilde{\Sigma}_v^{-\frac{1}{2}} = I\otimes\Sigma_v^{-\frac{1}{2}}$ and $\Sigma^{\frac{1}{2}}$ is the unique square root of Σ_v, and $s_i(\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F})$ denotes the i-th singular value of $\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F}$.

Proof: Let $\widetilde{\Sigma}_v = I\otimes\Sigma_v$. $S_0 = \sigma_0^2\widetilde{F}\widetilde{F}^{\mathsf{T}} + \widetilde{\Sigma}_v$ (see Proposition 3.1). Let $U\Lambda U^{\mathsf{T}}$ be the eigenvalue decomposition of $\widetilde{F}\widetilde{F}^{\mathsf{T}}\widetilde{\Sigma}_v^{-1}$, where $UU^{\mathsf{T}} = U^{\mathsf{T}}U = I$. We claim that $F(I\otimes$

$\Sigma_l)F^\mathsf{T}S_0^{-1}$ and $\sigma_\ell^2[\sigma_0^2 I+\Lambda^{-1}]^{-1}$ are similar matrices. To see this, we begin with the following identity:

$$\begin{aligned} F(I\otimes\Sigma_l)F^\mathsf{T}S_0^{-1} &= \sigma_\ell^2\widetilde{F}\widetilde{F}^\mathsf{T}(\sigma_0^2\widetilde{F}\widetilde{F}^\mathsf{T}+\widetilde{\Sigma}_v)^{-1} \\ &= \sigma_\ell^2\widetilde{F}\widetilde{F}^\mathsf{T}\widetilde{\Sigma}_v^{-1}(\sigma_0^2\widetilde{F}\widetilde{F}^\mathsf{T}\widetilde{\Sigma}_v^{-1}+I)^{-1}. \end{aligned}$$

The claim follows by substituting $\widetilde{F}\widetilde{F}^\mathsf{T}\widetilde{\Sigma}_v^{-1}=U\Lambda U^\mathsf{T}$ in the right hand side of above expression. From (15), we have

$$\begin{aligned} D_W(Q^*,\Omega) &= 0.5\,\mathrm{tr}[(F(I\otimes\Sigma_l)F^\mathsf{T}S_0^{-1})^2] \\ &= 0.5\,\mathrm{tr}[U^\mathsf{T}(F(I\otimes\Sigma_l)F^\mathsf{T}S_0^{-1})^2U] \\ &= 0.5\,\mathrm{tr}[\sigma_\ell^4[\sigma_0^2 I+\Lambda^{-1}]^{-2}] \\ &= 0.5\sum_{i=1}^{Nd}\left[\frac{\sigma_\ell^2\lambda_i}{\sigma_0^2\lambda_i+1}\right]^2, \end{aligned}$$

where the second equality follows from the cyclic property of trace operator and the fact that $UU^\mathsf{T}=I$, the third because $F(I\otimes\Sigma_l)F^\mathsf{T}S_0^{-1}$ and $\sigma_\ell^2[\sigma_0^2 I+\Lambda^{-1}]^{-1}$ are similar matrices. The proof follows by noticing that $s_i^2(\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F}) = \lambda_i((\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F})(\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F})^\mathsf{T}) = \lambda_i(\widetilde{F}\widetilde{F}^\mathsf{T}\widetilde{\Sigma}_v^{-1})$. ■

We shall drop the argument in $s_i(\widetilde{\Sigma}_v^{-\frac{1}{2}}\widetilde{F})$ when the context is clear. Corollary 3.4 states that when the nominal input and the attack lies in the same subspace, the optimal WDC (16) is governed by the singular values of the (weighted) impulse response matrix $\widetilde{\Sigma}_v^{-\frac{1}{2}}F(I\otimes M^{\frac{1}{2}})$. Let $M=I$ and $\Sigma_v=\sigma_v^2 I$, and consider the following special cases:

Static system ($A=0$ and $CB=I$): From (16), it follows that $D_W(Q^*,\Omega) = 0.5[Nd(\sigma_\ell^2/(\sigma_0^2+\sigma_v^2))^2]$, which is the square of SNR [3]. This expression says that larger the attack signature, or larger the number of spatial/temporal measurements, better the detection performance.

Dynamical system ($A\neq 0$ and $CB\neq I$): From (16), depending on s_i's being lesser or greater than one, one can show that $D_W(Q^*,\Omega)$ can be lesser or greater (respectively) than the optimal WDC (16) associated with the static system. Thus, when the network nodes interact, depending on their interaction (connectivity structure or magnitude of the edge weight), the detection performance may degrade.

For the ease of presentation, in what follows, we work with the optimal WDC characterized in Corollary 3.4.

A. Asymptotic optimal WDC

We consider the asymptotic measurement regime, i.e., we let $N\to\infty$ in the expression given by (16). To this end, let $\rho(A)<1$, and consider the following transfer function matrix associated with the state-space model in (1) and (3):

$$H(e^{j\theta}) = C(e^{j\theta}I-A)^{-1}B = C\textstyle\sum_{k=0}^{\infty}A^k e^{-jk\theta}B, \quad (17)$$

where $\theta\in[0,2\pi]$ and j is the imaginary unit. The second equality is a consequence of the Liouville-Neumann series expansion theorem [21]. We show that as $N\to\infty$, a properly scaled WDC (16) converges to a function of $H(e^{j\theta})$. We need the following definition to state the result.

[3]Signal being the lower bound on the attack variance and noise refers to the input and sensor noise

Definition 2: ***(Matrix function [22])*** Consider an $m\times m$ diagonalizable matrix $M=U\mathrm{diag}(\lambda_1(A),\ldots,\lambda_n(A))U^{-1}$. Let g be a complex function on the set $\{\lambda_1(A),\ldots,\lambda_n(A)\}$. Then, $g(M)=U\mathrm{diag}(g(\lambda_1(A)),\ldots,g(\lambda_n(A)))U^{-1}$. □

Lemma 3.5: ***(Optimal WDC: asymptotic measurements)*** Let $\Sigma_0^2=\sigma_0^2 I$, $\Sigma_\ell^2=\sigma_\ell^2 I$, and $\Sigma_v^2=\sigma_v^2 I$. The asymptotic optimal WDC (8) can be computed as

$$\overline{D}_W(Q^\star,\Omega) = \frac{1}{4\pi}\int_0^{2\pi}\mathrm{tr}\left[g(H^\mathsf{T}(e^{-j\theta})H(e^{j\theta}))\right]\,d\theta, \quad (18)$$

where the mapping $g(x)=(\sigma_\ell^2\,x/(\sigma_0^2\,x+\sigma_v^2))^2$.

Proof: Since B and C are bounded matrices and $\rho(A)<1$, matrix $H(e^{j\theta})$ is bounded for all $\theta\in[0,2\pi]$. Now, the statement of the lemma follows as a corollary to the Avram-Parter theorem in the block Toeplitz case [23]. ■

Notice that the expression (18) is finite and it also coincides with $\liminf_{N\to\infty}D_W(Q^*,\Omega)/N$. Thus, $\overline{D}_W(Q^\star,\Omega)$ (18) is the actual limit of $D_W(Q^*,\Omega)/N$ (16).

B. Network Design

In this section, for fixed input and output nodal locations, and input, attack, and noise statistics, we design networks that maximizes the performance of an asymptotic[4] optimal WDC based attack detector. The reason for choosing the asymptotic criterion is that it is easy to evaluate numerically.

We assume that the edge weights are bounded ie., $a_{\min}\le a_{ij}\le a_{\max}$ for all $i,j\in\{1,\ldots,n\}$, where $a_{\max}>0$ and $a_{\min}<0$. This assumption is weaker than those made in the prior works [10], [11], which also assume that the networks are un directed. In our numerical examples, we compare the performance of the optimal WDC based detector on directed networks to that of the undirected networks.

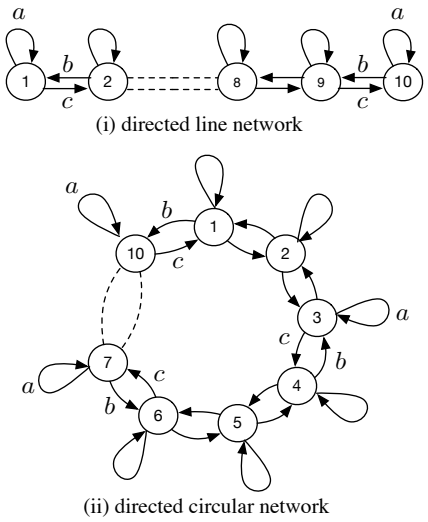
Using (15) as an objective function, the problem (9) can be reformulated as

$$\begin{aligned} \underset{A\,\in\,\mathbb{M}^{n\times n}}{\text{maximize}} \quad & \frac{1}{4\pi}\int_0^{2\pi}\mathrm{tr}\left[g(H^\mathsf{T}(e^{-j\theta})H(e^{j\theta}))\right]\,d\theta, \\ \text{subject to} \quad & \rho(A)<1 \quad \text{(stability)}, \\ & a_{\min}\mathbf{1}\mathbf{1}^\mathsf{T}\le A\le a_{\max}\mathbf{1}\mathbf{1}^\mathsf{T} \quad \text{(weight)}, \\ & |\mathcal{E}|\le K \quad \text{(connectivity)}. \end{aligned} \quad (19)$$

Notice that the preceding (19) is a non-convex optimization problem. To see this, note that the spectral radius $\rho(A)$, in general, is a non-convex function of A. Further, the optimization variable is not a symmetric matrix. Thus, our design problem may not be numerically solved using the existing off the shelf convex optimization, at least in its current form. Thus, in this paper, we use standard non-linear programming software, such as `fmincon` in MATLAB, to solve (19). As a result, the optimal solution might not be a global maximum.

Remark 2: ***(Network design: optimal WDC vs H_2 norm)*** We compare the asymptotic optimal WDC to that of the H_2

[4]Similar procedure can be used for designing networks that maximizes the optimal WDC given by (15).



(i) directed line network

(ii) directed circular network

Fig. 1: An illustration of line and circular networks. By setting $b = c$, both the networks become undirected.

norm—a widely used metric for network design—given by

$$\|H\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi} \text{tr}\left[(H^{\mathsf{T}}(e^{-j\theta})H(e^{j\theta}))\right]\, d\theta. \qquad (20)$$

If g in (18) is linear, the H_2 norm is identical (upto a scaling factor) to the asymptotic optimal WDC. This happens when all the singular values of $H(e^{j\theta})$ are zero or one (if $\sigma_0^2 = 0$). In general, the H_2-norm (20) and the asymptotic optimal WDC (18) both can outperform each other, because each of them are associated with different statistical moments of the measurements, and there is no ordering relation among the moments. Thus, the network that maximizes asymptotic WDC may not maximize the H_2 norm, and vice versa. □

IV. Numerical Examples

We present few numerical examples to illustrate the performance of asymptotic optimal WDC based detector on line and circular networks; see Fig 1. For a network of size n, the directed line and circular networks are parameterized by the weights $[a, b, c]$. By setting $b = c$, we get the un-directed network. For all the examples, we chose $n = 10$, $\sigma_\ell^2 = 1.0$, $\sigma_0^2 = 0.75$, and $\sigma_v^2 = 0.5$. The weight constraint bounds in (19) are $a_{\max} = 2$ and $a_{\min} = -2$.

A. Detection performance: line and circular networks

In Fig. 2, we plot the detection performance for the line and circular networks with no self-loops ($a = 0$), by varying the edge weights b and c. In both cases, the input and output nodes are 1 and 2, respectively. For both these networks, we see that the detection performance is a convex function in b and c. Further, the detection performance over the directed line (or circular) network is greater than its undirected counterpart. This has been also verified by solving[5] the design problem 19, whose values are reported in Table I.

[5]We used `fmincon` command in MATLAB R2019b to solve 19.

TABLE I: Optimal edge weights of line and circular networks

network	weights	asymptotic WDC
directed line	$b = 0$, $c = 2$	**0.8843**
undirected line	$b = 0.52$, $c = 0.52$	0.0269
directed circular	$b = 0$, $c = 2$	**0.2956**
undirected circular	$b = 0.52$, $c = 0.52$	0.0442

B. Detection performance: sensor placement

In Fig. 3, we plot the detection performance of optimal line and circular networks as a function of the output nodes. In particular, for an attack input at node 1, we solved 19 for every output node in $\{2, 3, \ldots, 10\}$. From panel (i), notice that the detection performance on undirected line networks deteriorates as the input-output distance increases. Instead, the detection performance on a directed line network improves as the input-output distance increases. A similar description holds for the circular networks. This counter-intuitive behavior has also been shown in the context of the optimal (when the attack input is completely specified, unlike in our current setting) *Maximum-a-Posteriori* detector [5].

V. Conclusions

This paper studies an attack detection problem for network systems. In particular, for random input attacks affecting few nodes in the network, we consider a worst-case deflection-based detector that relies only on the known lower bound of the attack's covariance matrix but not on its probability distribution. We explicitly characterize the performance of this detector in terms of the network weights, noise statistics, and input and output nodes' location. Further, for given input and output nodes, we also formulate and solve an optimization problem to find the edge weights that maximize the detection performance. Our numerical results suggest that directed line and circular networks have better detection performance than their undirected counterparts.

References

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

[2] Y. Mo, J.P.Hespanha, and B. Sinopoli. Resilient detection in the presence of integrity attacks. *IEEE Trans. Signal Process.*, 62(1):31–43, 2014.

[3] J. Liang, L. Sankar, and O. Kosut. Vulnerability analysis and consequences of false data injection attack on power system state estimation. *IEEE Transactions on Power Systems*, 31(5):3864–3872, 2016.

[4] R. Dhal, J. Abad Torres, and S. Roy. Detecting link failures in complex network processes using remote monitoring. *Physica A: Statistical Mechanics and its Applications*, 437:36 – 54, 2015.

[5] R. Anguluri, R. Dhal, S. Roy, and F. Pasqualetti. Network invariants for optimal input detection. In *American Control Conference*, pages 3776–3781, 2016.

[6] F. Pasqualetti, A. Bicchi, and F. Bullo. A graph-theoretical characterization of power network vulnerabilities. In *Proceedings of the 2011 American Control Conference*, pages 3918–3923, 2011.

[7] 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)*, 51(4):76, 2018.

[8] C. Murguia and J. Ruths. Cusum and chi-squared attack detection of compromised sensors. In *2016 IEEE Conference on Control Applications (CCA)*, pages 474–480, 2016.

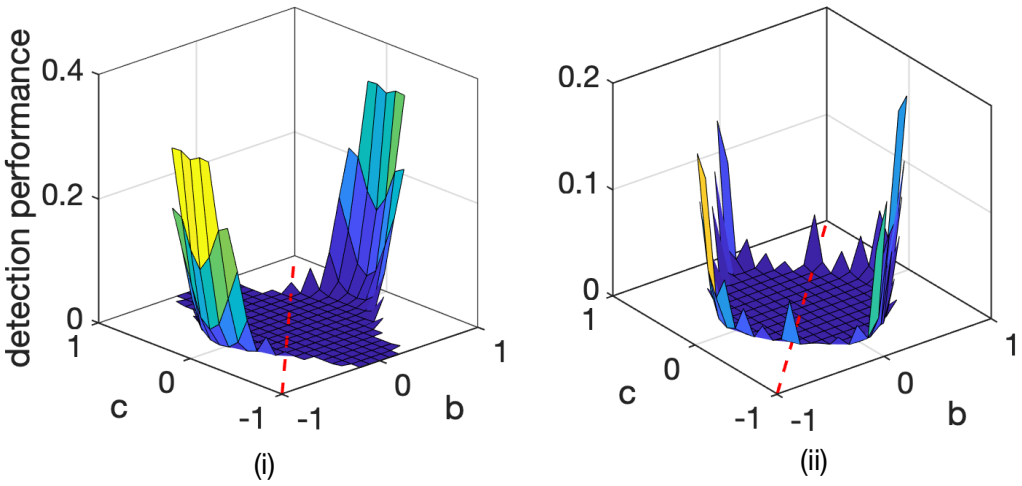


Fig. 2: Detection performance on (i) line and (ii) circular networks in the absence of self-loops. In both the plots, the detection performance associated with the directed networks ($b \neq c$) is always greater than that of the directed networks ($b = c$; indicated by dashed line). The un plotted region in the bc-plane corresponds to the adjacency matrix configuration that is unstable.

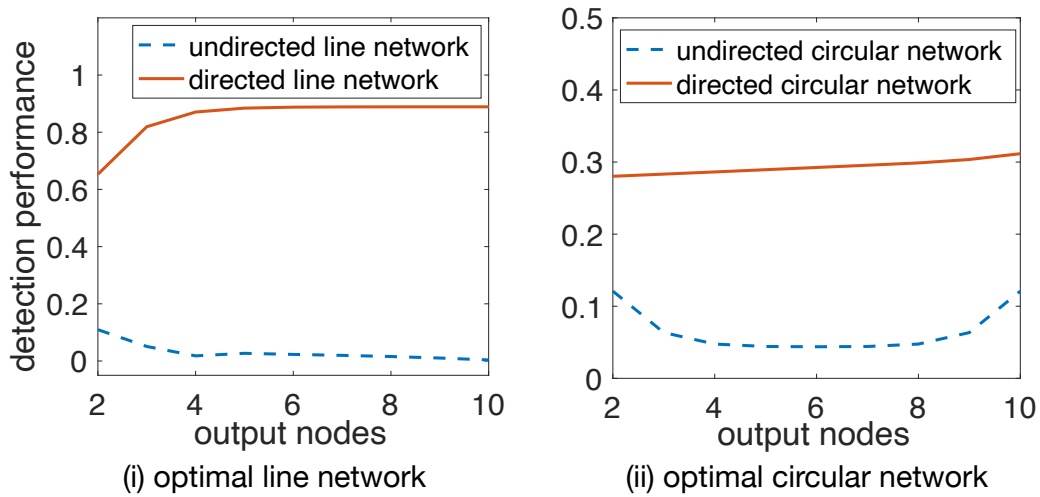


Fig. 3: Detection performance associated with the optimal networks vs sensor location (for an attack at node 1). Irrespective of the sensor location, the detection performance of an optimal directed network (either line or circular) is always greater than the undirected one. Further, the detection performances of the optimal directed networks is maximum when the sensor node is far away from the input node.

[9] S. Kar, S.Aldosari, and J.M.F. Moura. Topology for distributed inference on graphs. *IEEE Trans. Signal Process.*, 56(6):2609–2613, 2008.

[10] A. Chapman, E. Schoof, and M. Mesbahi. Distributed online topology design for network-level disturbance rejection. In *52nd IEEE Conference on Decision and Control*, pages 817–822, 2013.

[11] T. Summers, I. Shames, J. Lygeros, and F. Dörfler. Topology design for optimal network coherence. In *2015 European Control Conference (ECC)*, pages 575–580, 2015.

[12] U. A. Khan and M. Doostmohammadian. A sensor placement and network design paradigm for future smart grids. In *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 137–140, 2011.

[13] T. R., C. Murguia, and J. Ruths. Tuning windowed chi-squared detectors for sensor attacks. In *2018 Annual American Control Conference (ACC)*, pages 1752–1757, 2018.

[14] C. Baker. Optimum quadratic detection of a random vector in gaussian noise. *IEEE Trans. on Comm. Tech.*, 14(6):802–805, 1966.

[15] B. Picinbono and P. Devaut. Optimal linear-quadratic systems for detection and estimation. *IEEE Transactions on Information Theory*, 34(2):304–311, 1988.

[16] S. Kim, A. Magnani, A. Mutapcic, S. P. Boyd, and Z. Luo. Robust beamforming via worst-case sinr maximization. *IEEE Transactions on Signal Processing*, 56(4):1539–1547, 2008.

[17] W. Zhang, H. V. Poor, and Z. Quan. Frequency-domain correlation: An asymptotically optimum approximation of quadratic likelihood ratio detectors. *IEEE Trans. Signal Process.*, 58(3):969–979, 2010.

[18] J. Zhang and R. S. Blum. Asymptotically optimal truncated multivariate gaussian hypothesis testing with application to consensus algorithms. *IEEE Trans. Signal Process.*, 62(2):431–442, 2014.

[19] A. Khuri. *Linear Model Methodology*. New York: Chapman and Hall/CRC, 2009.

[20] M. R. Sheriff and D. Chatterjee. On a frame theoretic measure of quality of LTI systems. In *IEEE Conf. on Decision and Control*, pages 4012–4017, 2017.

[21] Athanasios C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Society for Industrial and Applied Mathematics, 2005.

[22] N.J. Higham. *Functions of Matrices Theory and Computation*. Society for Industrial and Applied Mathematics, 2008.

[23] A. Böttcher and B. Silbermann. *Introduction to Large Truncated Toeplitz Matrices*. New York: Springer-Verlag, 1999.