Causality Countermeasures for Anomaly Detection in Cyber-Physical Systems

Dawei Shi, Ziyang Guo, Karl Henrik Johansson, *Fellow, IEEE*, and Ling Shi

***Abstract*—The problem of attack detection in cyber-physical systems is considered in this paper. Transfer-entropy-based causality countermeasures are introduced for both sensor measurements and innovation sequences, which can be evaluated in a data-driven fashion without relying on a model of the underlying dynamic system. The relationships between the countermeasures and the system parameters as well as the noise statistics are investigated, based on which conditions that guarantee the time convergence of the countermeasures are obtained. The effectiveness of the transfer entropy countermeasures in attack detection is evaluated via theoretical analysis, numerical demonstrations, as well as comparative simulations with classical χ^2 detectors. Four types of attacks are considered: denial-of-service, replay, innovation-based deception, and data injection attacks. Abnormal behavior of the transfer entropy can be observed after the occurrence of each of these attacks.**

***Index Terms*—Anomaly detection, causality countermeasures, cyber-physical systems, transfer entropy.**

I. Introduction

THE increased importance of communication networks in control systems and the emergence of cyber-physical systems have reinforced safety and security requirements in control system design. Recently reported security related accidents (e.g., the Maroochy water bleach [1], the StuxNet malfare [2], smart grid attacks [3]) evidently indicate the impendence of these requirements, as vulnerabilities in civil infrastructures and industrial processes may cause devastating consequences to economy, national security, and even human life.

Attack detection and secure control system design has received a lot of attention during the past few years. A number of results focus on specific types of attacks, and the goals are to design detection schemes or to build attack-resilient controllers based on the feature of the attacks considered. False data injection attacks were analyzed in [4]–[7]. Specifically, the effect of false data injection attacks on state estimation of a discrete-time linear time-invariant Gaussian system was investigated in [6], and a quantitative measure of the resilience of the system to the attacks was proposed by characterizing the set of abnormal measurements that could blind a failure detector. In [7], two scenarios of false data injection attacks in electric power grids were considered, and the stealthiness of the attacks was demonstrated numerically. Replay attacks were considered in [8]–[10]. For this type of attacks, the feasibility condition and countermeasures were considered in [8] and [9], while a stochastic game theoretic approach was investigated to solve the attack detection problem in [10]. Denial-of-service (DoS) attacks were studied in [11]–[15]. In [11], a problem of security constrained optimal control under DoS attacks was investigated. The worst case attack policies against remote state estimation were investigated in [12] and [13]. In [14], a problem of risk-sensitive stochastic control under a Markov DoS model was investigated using the reference probability measure approach, and a separation principle for the stochastic control problem was proved. In [15], the frequency and duration of DoS attacks to maintain input-to-state stability of the closed-loop system was characterized.

In many scenarios, however, it is difficult to know *a priori* what type of attacks may be inserted into the system. For example, almost all deception attacks that prevent the controller from knowing the real sensor measurements share the same aim of deteriorating the system performance without being detected, but the specific policies used to generate the false measurement data are different and difficult to be distinguished until the attacks are identified. In fact, identification of the attack type is not always necessary, since the ultimate goal is to detect the existence of the attack (rather than its type) and, then, eliminate it to ensure safe and secure operation. So systematic approaches to attack detection and secure estimation policies applicable to different attack scenarios have been developed as well. In [16], the problem of attack detection and identification was solved utilizing system- and graph-theoretic approaches; centralized and distributed attack detection and identification monitors were proposed. In [17], the maximum number of tolerable attacks that allow accurate reconstruction of the system state was characterized, and an efficient algorithm capturing the sparsity pattern of

Manuscript received January 3, 2017; revised May 15, 2017; accepted May 31, 2017. Date of publication June 12, 2017; date of current version January 26, 2018. The work of D. Shi was supported by the Natural Science Foundation of China under Grant 61503027. The work of Z. Guo and L. Shi was supported by a Hong Kong RGC GRF Grant 16210015. The work of K. H. Johansson was supported in part by the Knut and Alice Wallenberg Foundation, in part by the Swedish Strategic Research Foundation, and in part by the Swedish Research Council. Recommended by Associate Editor R. M. Jungers. *(Corresponding author: Dawei Shi.)*

D. Shi is with the State Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: dawei.shi@outlook.com).

Z. Guo and L. Shi are with Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: zguoae@ust.hk; eesling@ust.hk).

K. H. Johansson is with the ACCESS Linnaeus Centre and Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm 114 28, Sweden (e-mail: kallej@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2017.2714646

the attack policies was proposed and analyzed; based on these results, a procedure for attack-resilient state estimation of systems with noise and modeling errors was developed in [18]. In [19], adversary models were proposed for different attack policies and the impact of the attacks was quantified by the concept of safe sets. A finite-state stochastic modeling framework was introduced in [20], based on which the problem of secure estimation was investigated. For further results and developments, see also [21]–[28] and the recent special issue on cyber-security of networked control systems [29].

It is important to consider the effect of process and measurement noises in secure controller design, due to the inevitable existence of disturbances and unmodeled dynamics. When noises are considered, χ^2 detectors or residue detectors are normally utilized for attack detection [5], [7], [9], [30]; these detectors have been extensively utilized in process monitoring and fault detection [31], but may not be able to capture the stealthy attacks. Therefore, it is necessary to figure out efficient countermeasures that are capable of detecting the existence of different attacks for dynamic systems affected by noises or disturbances. In this paper, a type of countermeasures that potentially satisfy these requirements is investigated. The main results obtained are summarized as follows.

First, transfer entropy countermeasures are introduced for anomaly detection in cyber-physical systems. For these countermeasures, the common causation feature of different attacks is characterized. The intuition is that in a large number of scenarios, the intervention of malicious attacks invariably disturb the cause-effect relationships between different variables and affect the underlying causality properties of the original dynamic system. The notion of transfer entropy was originally introduced in [32] as a measure of information transfer, and has been extensively utilized for causality analysis in different disciplines [33]–[38]. In this paper, transfer entropy measures based on sensor measurements and innovation sequences are used to detect the existence of attacks for discrete-time linear Gaussian systems. The qualitative relationships between the transfer entropy measures and the parameter matrices of the system dynamics as well as the noise statistics are explored for the unattacked system (see Proposition 1). Conditions that guarantee the convergence of the transfer entropy measures with respect to time are provided (see Theorems 1 and 2), and we show that the condition for the innovation-based transfer entropy is weaker than that for the transfer entropy based on sensor measurements.

Second, the effect of different attacks on the transfer entropy countermeasures are investigated through both theoretical analysis and numerical verifications. Four types of attacks are considered: the DoS, replay, innovation-based deception, and data injection attacks. For all attacks considered, we show that the changes in transfer entropy are expected when attacks are deployed. The feasibility of the approach is not only showed theoretically, but also numerically observed in simulation examples, where χ^2 detectors are included for comparison. Another benefit of the transfer entropy measure is that it allows data-driven implementation [36], which is preferable in attack detection, as the model developed based on unattacked data may not remain effective in characterizing the dynamics of the attacked system.

In this paper, the transfer entropy countermeasures are utilized for anomaly detection of cyber-physical systems. The hindsight is that when a signal is affected by attacks, it is likely to lose the authenticity required to maintain the original causal relationship with other signals in a control system, as is validated by the extensive simulation studies presented in this paper—The existence of attacks do change the transfer entropy readings in a certain way in many scenarios, indicating the change of causal relationships of the signals. On the other hand, causality is a fundamental property that is associated with the signals in a dynamic system, a property that has been investigated for a long time in different disciplines but is not extensively considered in control system analysis, which intrigues us to utilize this property in detecting the attacks launched by an intelligent attacker. It is also interesting to note that in a very recent paper [39], a related idea exploiting the notion of "information flow" was employed in cyber-security design of control systems. To our best knowledge, the notion of information flow also builds on the idea of analyzing the cause-effect relationship for different processes (objects). The notion of information flow in [39] originated from software security [40], while the transfer entropy notion utilized in our work was originally proposed to detect asymmetry in the interaction of subsystems for causality analysis [32]. On the other hand, we note that the information flow in [39] was quantified via the Kullback–Leibler (KL) divergence between the distribution of the output under attack and the distribution of the output under normal operation; transfer entropy can also be regarded as a special form of KL divergence, although the transfer entropy countermeasure considered in this paper attempts to achieve anomaly detection by analyzing the cause-effect relationship of two different signals in a control system.

The rest of the paper is organized as follows: Section II presents the system description and problem formulation. The transfer entropy countermeasures are introduced and analyzed in Section III. Section IV investigates the effect of different attacks on transfer entropy. Discussions on application issues are presented in Section V. Numerical examples and comparisons are provided in Section VI. Some concluding remarks and discussions on future work are provided in Section VII. Discussions on implementation issues of transfer entropy are summarized in Appendix A.

Notation: For $i, j \in \mathbb{N}$ and $i \leq j$, we use the shorthand notation $x_{i:j} := \{x_i, \ldots, x_j\}$. For a probability measure P, we use $E(\cdot)$ to represent the expectation operator, use $\mathrm{Cov}(\cdot)$ to represent the covariance of random processes, and use $f(\cdot)$ to denote a probability distribution density function. Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, then $A \otimes B$ denotes the Kronecker product of A with B, and $\mathrm{vec}\, A$ denotes the vectorization of matrix A; when $m = n$, $\sigma(A)$ denotes the set of eigenvalues of A. Finally, for $x \in \mathbb{R}$ and $x > 0$, $\log x$ denotes natural logarithm of x, namely, the logarithm of x to base e.

II. System Description and Problem Formulation

Consider the attack detection scheme in Fig. 1. The nominal process is first assumed to have the following form:

$$x_{k+1} = Ax_k + w_k \tag{1}$$

where $x_k \in \mathbb{R}^n$ denotes the set of process state variables, and $w_k \in \mathbb{R}^n$ denotes the process noise. The information of the process is measured using a number of sensors of the form

$$y_k^i = C^i x_k + v_k^i \tag{2}$$

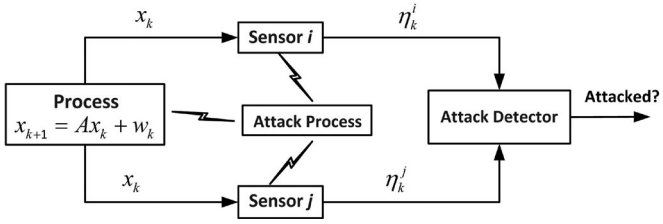


Fig. 1. Attack detection scheme based on causality countermeasures.

where v_k^i denotes the measurement noise of sensor i. We assume the number of sensors equals M. We assume the initial state x_0 is zero-mean Gaussian with covariance P_0, w_k and v_k^i are zero-mean Gaussian white noises with covariance matrices Q and R^i, respectively, and in addition, x_0, w_k, and v_k^i are mutually uncorrelated. Although the system setup considered for linear Gaussian processes cannot exactly describe the nonlinear dynamics of a complex system, it serves as a reasonable approximation of the systems considered in many cases; in terms of attack detection, the simple structure of the model considered helps understand how the countermeasures manage to detect the changes caused by the stealthy attacks. Due to attacks on sensor measurements, the received measurement η_k^i may be different from the actual measurement process y_k^i:

$$\eta_k^i = \mathcal{F}(y_k^i, \theta_k^i) \tag{3}$$

indicating that η_k^i may depend on the sensor measurement information as well as another parameter θ_k^i; in fact, in certain attack scenarios, η_k^i does not reflect the true measurement y_k^i at all, e.g., replay attacks. The attacks may directly insert data into the process as well; in this case, we assume the attacked process inserted has the following form:

$$x_{k+1} = Ax_k + Ba_k + w_k \tag{4}$$

where a_k is the attack signal.

The goal of this paper is to propose efficient generic countermeasures that are able to detect the existence of attacks, and to evaluate the effectiveness of the countermeasures in attack detection. A baseline assumption in countermeasure design is that the type of attack on the system is unknown, so the proposed measure should be effective to different attack policies. Specifically, the countermeasures considered in this paper take advantage of the causality or connectivity relationships between signals in Fig. 1, as will be introduced in detail in the next section.

III. Causality Countermeasures and Transfer Entropy

In this section, transfer entropy countermeasures are introduced. For the considered system model, the countermeasures are derived and their convergence properties are analyzed.

The process structural properties and the relationships between sensor measurements are captured by causality countermeasures. Causality analysis has been extensively utilized in econometric inference, biosciences, climatology, functional magnetic resonance imaging, and more recently industrial alarm system design [36], [41], [42], to identify topological properties of the underlying complex networks. For two sensor measurement processes y^i and y^j, the transfer entropy quantifies the causal relationship between the two variables; to be specific, for positive integer-valued parameters τ, μ, l, and $k \geq \max\{\mu, l\}$, the transfer entropy from y^j to y^i at time instant $k+\tau$ is defined as

$$\begin{aligned}\mathcal{T}_{y^j \to y^i}(k+\tau) = & \int f(y_{k+\tau}^i, y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \\ & \times \log \frac{f(y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j)}{f(y_{k+\tau}^i | y_{k-\mu+1:k}^i)} \\ & \times \mathrm{d}y_{k+\tau}^i \mathrm{d}y_{k-\mu+1:k}^i \mathrm{d}y_{k-l+1:k}^j \end{aligned} \tag{5}$$

where we recall that f denotes relevant probability density functions. Obviously, this measure is asymmetric, namely,

$$\mathcal{T}_{y^j \to y^i}(k+\tau) \neq \mathcal{T}_{y^i \to y^j}(k+\tau)$$

in general. In the following, we analyze several properties of this causality measure based on the nominal process models in (1) and (2), which help provide benchmarks for different attack scenarios studied in the next section.

By Bayes' rule, we have

$$\begin{aligned} & \int f(y_{k+\tau}^i, y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \\ & \quad \times \log f(y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \\ & \quad \times \mathrm{d}y_{k+\tau}^i \mathrm{d}y_{k-\mu+1:k}^i \mathrm{d}y_{k-l+1:k}^j \\ = & \int f(y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \\ & \quad \times \log f(y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \mathrm{d}y_{k+\tau}^i \\ & \quad \times f(y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \mathrm{d}y_{k-\mu+1:k}^i \mathrm{d}y_{k-l+1:k}^j. \end{aligned}$$

For the nominal process, since x_0 is Gaussian, and w_k and v_k^i are Gaussian white noises, $y_{k+\tau}^i, y_{k-\mu+1:k}^i, y_{k-l+1:k}^j$ are jointly Gaussian. Therefore, $y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j$ is also Gaussian; for notational brevity, we write

$$y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j \sim \mathcal{N}(\bar{y}_{k+\tau}^i, \Phi_{k+\tau}^i).$$

We have

$$\begin{aligned} & \int f(y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \\ & \quad \times \log f(y_{k+\tau}^i | y_{k-\mu+1:k}^i, y_{k-l+1:k}^j) \mathrm{d}y_{k+\tau}^i \\ = & -\frac{1}{2} \int \frac{1}{\sqrt{(2\pi)^m \det \Phi_{k+\tau}^i}} \\ & \quad \times \exp\left[-\frac{1}{2} (y_{k+\tau}^i - \bar{y}_{k+\tau}^i)^\top (\Phi_{k+\tau}^i)^{-1} (y_{k+\tau}^i - \bar{y}_{k+\tau}^i) \right] \\ & \quad \times \left[(y_{k+\tau}^i - \bar{y}_{k+\tau}^i)^\top (\Phi_{k+\tau}^i)^{-1} (y_{k+\tau}^i - \bar{y}_{k+\tau}^i) \right. \\ & \quad \left. + \log[(2\pi)^m \det \Phi_{k+\tau}^i] \right] \mathrm{d}y_{k+\tau}^i \\ = & -\frac{m}{2} \log(2\pi e) - \frac{1}{2} \log \det \Phi_{k+\tau}^i \end{aligned} \tag{6}$$

which implies

$$\int f\left(y^i_{k+\tau}, y^i_{k-\mu+1:k}, y^j_{k-l+1:k}\right) \times \log f(y^i_{k+\tau}|y^i_{k-\mu+1:k}, y^j_{k-l+1:k})\mathrm{d}y^i_{k+\tau}\mathrm{d}y^i_{k-\mu+1:k}\mathrm{d}y^j_{k-l+1:k} = -\frac{m}{2}\log(2\pi e) - \frac{1}{2}\log\det\Phi^i_{k+\tau}. \quad (7)$$

Similarly, $y^i_{k+\tau}|y^i_{k-\mu+1:k} \sim \mathcal{N}(\hat{y}^i_{k+\tau}, \Psi^i_{k+\tau})$, so we have

$$\int f\left(y^i_{k+\tau}, y^i_{k-\mu+1:k}, y^j_{k-l+1:k}\right) \times \log f(y^i_{k+\tau}|y^i_{k-\mu+1:k})\mathrm{d}y^i_{k+\tau}\mathrm{d}y^i_{k-\mu+1:k}\mathrm{d}y^j_{k-l+1:k} = -\frac{m}{2}\log(2\pi e) - \frac{1}{2}\log\det\Psi^i_{k+\tau}. \quad (8)$$

Therefore, we have

$$\mathcal{T}_{y^j\to y^i}(k+\tau) = 0.5\log\left(\det\Psi^i_{k+\tau}/\det\Phi^i_{k+\tau}\right). \quad (9)$$

The transfer entropy $\mathcal{T}_{y^j\to y^i}(k+\tau)$ measures the amount of improvement when using past information of both y^i and y^j to predict the future of y^i, compared with that obtained using y^i only. It reflects the causation link from y^j to y^i. To analyze how the transfer entropy measure reflects connectivity properties of the process (1)–(2), it suffices to derive the expressions for $\Psi^i_{k+\tau}$ and $\Phi^i_{k+\tau}$. To do this, we introduce a few matrix operators. For notational brevity, write

$$C = [(C^i)^\top (C^j)^\top]^\top,\ R = \mathrm{diag}\{R^i, R^j\}.$$

Define the matrix operators $h(\cdot)$: $\mathbb{S}^n_+ \to \mathbb{S}^n_+$, $\tilde{g}_i(\cdot)$, $\tilde{g}(\cdot)$, $g_i(\cdot)$ and $g(\cdot)$: $\mathbb{S}^n_+ \to \mathbb{S}^n_+$ as follows:

$$h(X) := AXA^\top + Q \quad (10)$$

$$\tilde{g}_i(X) := X - X(C^i)^\top[C^iX(C^i)^\top + R^i]^{-1}C^iX \quad (11)$$

$$\tilde{g}(X) := X - XC^\top(CXC^\top + R)^{-1}CX \quad (12)$$

$$g_i(X) := \tilde{g}_i(h(X)),\ g(X) := \tilde{g}(h(X)). \quad (13)$$

Let

$$\vartheta := \begin{cases} i, & \text{if } \mu \geq l \\ j, & \text{otherwise} \end{cases}. \quad (14)$$

Since w_k and x_0 are zero-mean Gaussian, $x_{k-\max\{\mu,l\}}$ is zero-mean Gaussian as well with covariance $P_{k-\max\{\mu,l\}}$ satisfying

$$P_{k-\max\{\mu,l\}} = h^{k-\max\{\mu,l\}}(P_0). \quad (15)$$

Since v^i_k and v^j_k are Gaussian white noises and are mutually uncorrelated with w_k and x_0, it is easy to verify that $x_k|y^i_{k-\mu+1:k}, y^j_{k-l+1:k}$ is Gaussian with covariance

$$\bar{P}_k = g^{\max\{\mu,l\}-|\mu-l|} \circ g_\vartheta^{|\mu-l|}(P_{k-\max\{\mu,l\}}). \quad (16)$$

Therefore, $x_{k+\tau}|y^i_{k-\mu+1:k}, y^j_{k-l+1:k}$ is Gaussian with covariance

$$\bar{P}_{k+\tau} = h^\tau \circ g^{\max\{\mu,l\}-|\mu-l|} \circ g_\vartheta^{|\mu-l|}(P_{k-\max\{\mu,l\}}). \quad (17)$$

Finally, since v^i_k and w_k are mutually uncorrelated, we have

$$\Phi^i_{k+\tau} = C^i\bar{P}_{k+\tau}(C^i)^\top + R^i. \quad (18)$$

Following a similar argument, we have that $x_{k+\tau}|y^i_{k-\mu+1:k}$ is Gaussian with covariance

$$\hat{P}_{k+\tau} = h^\tau \circ g_i^\mu \circ h^{k-\mu}(P_0). \quad (19)$$

Since w_k and v^i_k are uncorrelated, we have

$$\Psi^i_{k+\tau} = C^i\hat{P}_{k+\tau}(C^i)^\top + R^i. \quad (20)$$

From the above discussions, we observe that the transfer entropy measure not only points out the direction of causality, but also quantitatively reflects the connectivity properties of the underlying dynamic system determined by the state-space model parameters and the noise statistics.

In many remote estimation scenarios, the sensor measurements are not directly transmitted to the remote estimator; instead, the sensor measurement y^i_k is preprocessed and only the innovation

$$z^i_k := y^i_k - \mathrm{E}(y^i_k|y^i_{0:k-1}) \quad (21)$$

is transmitted. In this case, it is necessary to evaluate the transfer entropy measure for the innovation processes; for positive integer-valued parameters τ, μ, l, and $k \geq \max\{\mu, l\}$, it is defined as

$$\mathcal{T}_{z^j\to z^i}(k+\tau) = \int f\left(z^i_{k+\tau}, z^i_{k-\mu+1:k}, z^j_{k-l+1:k}\right) \times \log\frac{f\left(z^i_{k+\tau}|z^i_{k-\mu+1:k}, z^j_{k-l+1:k}\right)}{f\left(z^i_{k+\tau}|z^i_{k-\mu+1:k}\right)} \times \mathrm{d}z^i_{k+\tau}\mathrm{d}z^i_{k-\mu+1:k}\mathrm{d}z^j_{k-l+1:k}. \quad (22)$$

Now, we derive (22) for the process (1)–(2). First we notice that since the process model is linear, $\mathrm{E}(y^i_k|y^i_{0:k-1})$ is a linear combination of y_0, ..., y_{k-1}. Since $y^i_{0:k}$ and $y^j_{0:k}$ are jointly Gaussian, $z^i_{0:k}$ and $z^j_{0:k}$ are jointly Gaussian as well. In this way, following the line of argument for the derivation of (5), we have

$$\mathcal{T}_{z^j\to z^i}(k+\tau) = \frac{1}{2}\log\frac{\det\Pi^i_{k+\tau}}{\det\Upsilon^i_{k+\tau}} \quad (23)$$

where

$$\Pi^i_{k+\tau} := \mathrm{Cov}(z^i_{k+\tau}|z^i_{k-\mu+1:k}) \quad (24)$$

$$\Upsilon^i_{k+\tau} := \mathrm{Cov}(z^i_{k+\tau}|z^i_{k-\mu+1:k}, z^j_{k-l+1:k}). \quad (25)$$

Since the innovation sequence $\{z^i_k\}$ is a zero-mean random process and satisfies $\mathrm{E}[z^i_k(z^i_t)^\top] = 0$, we have

$$\Pi^i_{k+\tau} = \mathrm{Cov}(z^i_{k+\tau}) = C^ig_i^{k+\tau-1}(P_0)(C^i)^\top + R^i. \quad (26)$$

The calculation of $\Upsilon^i_{k+\tau}$ is more complicated, as $\mathrm{E}[z^i_k(z^j_t)^\top] \neq 0$ holds in general. We start from

$$\Gamma_k := \mathrm{Cov}([z^i_{k+\tau}, z^i_{k-\mu+1:k}, z^j_{k-l+1:k}]^\top). \quad (27)$$

This matrix has the following structure:

$$\Gamma_k = \begin{bmatrix} \Gamma^{1,1}_k & \Gamma^{1,2}_k \\ (\Gamma^{1,2}_k)^\top & \Gamma^{2,2}_k \end{bmatrix} \quad (28)$$

where $\Gamma^{1,1}_k := \mathrm{Cov}([z^i_{k+\tau}, z^i_{k-\mu+1:k}]^\top)$ and $\Gamma^{2,2}_k := \mathrm{Cov}([z^j_{k-l+1:k}]^\top)$ are block-diagonal matrices and easy

to calculate based on the orthogonality between z_k^i and z_t^i for $k \neq t$, and $\Gamma_k^{1,2} := \mathrm{E}([z_{k+\tau}^i, z_{k-\mu+1:k}^i]^\top [z_{k-l+1:k}^j])$. To calculate Γ_k, it suffices to evaluate $\Gamma_k^{1,2}$, which requires us to find the expression of $\mathrm{Cov}[z_k^i (z_t^j)^\top]$. Without loss of generality, we assume $k \leq t$. For z_k^i, we observe that

$$z_k^i = y_k^i - \mathrm{E}(y_k^i | y_{0:k-1}^i) = C^i x_k - C^i \hat{x}_k^i + v_k^i$$

where $\hat{x}_k^i$ satisfies

$$\hat{x}_k^i = A \bar{x}_{k-1}^i$$
$$\bar{x}_k^i = \hat{x}_k^i + \hat{P}_k^{i,i} (C^i)^\top [C^i \hat{P}_k^{i,i} (C^i)^\top + R^i]^{-1} (y_k^i - C^i \hat{x}_k^i)$$

and $\hat{P}_k^{i,i} = \mathrm{Cov}(x_k - \hat{x}_k^i)$ satisfies $\hat{P}_k^{i,i} = h(\tilde{g}_i(\hat{P}_{k-1}^{i,i}))$. Writing $\hat{e}_k^i := x_k - \hat{x}_k^i$, we have

$$\hat{e}_{k+1}^i = \hat{A}_k^i \hat{e}_k^i + w_k + K_k^i v_k^i \tag{29}$$
$$z_k^i = C^i \hat{e}_k^i + v_k^i \tag{30}$$
$$\hat{A}_k^i = A - A \hat{P}_k^{i,i} (C^i)^\top [C^i \hat{P}_k^{i,i} (C^i)^\top + R^i]^{-1} C^i \tag{31}$$
$$K_k^i = A \hat{P}_k^{i,i} (C^i)^\top [C^i \hat{P}_k^{i,i} (C^i)^\top + R^i]^{-1}. \tag{32}$$

In this way, we have

$$\mathrm{E}[z_k^i (z_t^j)^\top] = C^i \mathrm{E}[\hat{e}_k^i (\hat{e}_t^j)^\top] (C^j)^\top. \tag{33}$$

Define

$$P_{k,t}^{i,j} := \mathrm{E}[\hat{e}_k^i (\hat{e}_t^j)^\top] \tag{34}$$

for brevity, we write $P_k^{i,j} := P_{k,k}^{i,j}$. To calculate $P_{k,t}^{i,j}$, we first find an expression for $P_k^{i,j} = \mathrm{E}[\hat{e}_k^i (\hat{e}_k^j)^\top]$. By assumption, for $k = 0$, we have $P_0^{i,j} = P_0$. Since w_k, v_k^i, and v_k^j are mutually uncorrelated, from (29), we have

$$P_{k+1}^{i,j} = \hat{A}_k^i P_k^{i,j} (\hat{A}_k^j)^\top + Q. \tag{35}$$

Starting from $P_k^{i,j}$, we further find a recursive expression for $P_{k,t}^{i,j}$ in terms of t. Obviously, we have $P_{k,t}^{i,j} = P_{k,k}^{i,j}$ for $t = k$, which can be calculated based on (35). For $t+1$, we observe

$$\hat{e}_{t+1}^j = \hat{A}_t^j \hat{e}_t^j + w_t + K_t^j v_t^j$$

and, thus, we have

$$P_{k,t+1}^{i,j} = \mathrm{E}[\hat{e}_k^i (\hat{e}_{t+1}^j)^\top] = P_{k,t}^{i,j} (\hat{A}_t^j)^\top. \tag{36}$$

Summarizing the above discussions, the expression of Γ_k can be obtained by combining (27), (28), (33), (35), and (36).

To calculate $\Upsilon_{k+\tau}^i$, we recall the following lemma.

Lemma 1 (see [43]): Let x and y be jointly Gaussian with covariance

$$\mathrm{Cov}([x^\top y^\top]^\top) = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^\top & \Sigma_{yy} \end{bmatrix}.$$

The covariance of the conditional distribution of x on y is $\Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy}^\top$.

To calculate $\Upsilon_{k+\tau}^i$, we consider an alternative parameterization of Γ_k:

$$\Gamma_k = \begin{bmatrix} \check{\Gamma}_k^{1,1} & \check{\Gamma}_k^{1,2} \\ (\check{\Gamma}_k^{1,2})^\top & \check{\Gamma}_k^{2,2} \end{bmatrix} \tag{37}$$

where $\check{\Gamma}_k^{1,1} := \mathrm{Cov}([z_{k+\tau}^i]^\top)$ and $\check{\Gamma}_k^{2,2} := \mathrm{Cov}([z_{k-\mu+1:k}^i, z_{k-l+1:k}^j]^\top)$ are block-diagonal matrices easy to calculate based on the orthogonality between z_k^i and z_t^i for $k \neq t$, and $\check{\Gamma}_k^{1,2} := \mathrm{E}([z_{k+\tau}^i]^\top [z_{k-\mu+1:k}^i, z_{k-l+1:k}^j])$. From Lemma 1, we have

$$\Upsilon_{k+\tau}^i = \check{\Gamma}_k^{1,1} - \check{\Gamma}_k^{1,2} (\check{\Gamma}_k^{2,2})^{-1} \check{\Gamma}_k^{2,1}. \tag{38}$$

We summarize the above derivations in the following proposition.

Proposition 1:

1) For the transfer entropy measure in (5), we have
$$\mathcal{T}_{y^j \to y^i}(k+\tau) = \log \left(\det \Psi_{k+\tau}^i / \det \Phi_{k+\tau}^i \right)^{\frac{1}{2}}$$
with $\Psi_{k+\tau}^i$ determined by (15), (17) and (18), and $\Phi_{k+\tau}^i$ by (19) and (20);
2) For the transfer entropy measure in (22), we have
$$\mathcal{T}_{z^j \to z^i}(k+\tau) = \log \left(\det \Pi_{k+\tau}^i / \det \Upsilon_{k+\tau}^i \right)^{\frac{1}{2}}$$
where $\Pi_{k+\tau}^i$ is determined by (26), and $\Upsilon_{k+\tau}^i$ by (27), (28), (33)–(38).

The above proposition completes the exploration of the theoretic expressions for the transfer entropy measures defined in (5) and (22) in terms of the model parameters and noise statistics of the system in (1)–(2), which provides the basis to analyze the asymptotic properties of the transfer entropy countermeasures. In the following, we analyze the convergence property of the two transfer entropy measures introduced above. We first present a technical lemma.

Lemma 2 (see [44, Theorem 4.2.12]): Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$. Let $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$ and $\sigma(B) = \{\xi_1, \ldots, \xi_m\}$. Then

$$\sigma(A \otimes B) = \{\lambda_i \xi_j | i = 1, \ldots, n, j = i, \ldots, m\}.$$

Now, we are ready to present the results on the convergence of the transfer entropy measures.

Theorem 1: Assume the values of τ, μ, and l are fixed. Then,

1) there exists $\bar{\mathcal{T}}_{y^i \to y^j} \in [0, \infty)$ such that $\lim_{k \to \infty} \mathcal{T}_{y^i \to y^j}(k+\tau) = \bar{\mathcal{T}}_{y^i \to y^j}$ if A is stable;
2) there exists $\bar{\mathcal{T}}_{z^i \to z^j} \in [0, \infty)$ such that $\lim_{k \to \infty} \mathcal{T}_{z^i \to z^j}(k+\tau) = \bar{\mathcal{T}}_{z^i \to z^j}$ if (A, Q) is stabilizable and (C^i, A) and (C^j, A) are detectable.

Proof: First we analyze the convergence of $\mathcal{T}_{y^i \to y^j}(k+\tau)$. From (17)–(20), since the values of τ, μ, and l are fixed and finite, we focus on the property of $\lim_{k \to \infty} h^k(P_0)$. As $h(\cdot)$ is a Lyapunov operator, it follows that $\lim_{k \to \infty} h^k(P_0)$ exists if A is stable, which completes the proof of the first part of the theorem.

To prove the second part of the theorem, it suffices to analyze the convergence of Γ_k. To do this, we first provide a more explicit form of the structure of Γ_k. Since $\Gamma_k^{1,1} := \mathrm{Cov}([z_{k+\tau}^i, z_{k-\mu+1:k}^i]^\top)$,

$$\Gamma_k^{1,1} = \mathrm{diag}\left\{ \mathrm{Cov}(z_{k+\tau}^i), \mathrm{Cov}(z_{k-\mu+1}^i), \ldots, \mathrm{Cov}(z_k^i) \right\}$$
$$= \mathrm{diag}\Big\{ C^i \hat{P}_{k+\tau}^{i,i} (C^i)^\top + R^i,$$
$$C^i \hat{P}_{k-\mu+1}^{i,i} (C^i)^\top + R^i, \ldots, C^i \hat{P}_k^{i,i} (C^i)^\top + R^i \Big\}.$$

Similarly, we have

$$\Gamma_k^{2,2} = \operatorname{diag}\left\{\operatorname{Cov}(z_{k-l+1}^j), \ldots, \operatorname{Cov}(z_k^j)\right\}$$
$$= \operatorname{diag}\left\{C^j \hat{P}_{k-l+1}^{j,j}(C^j)^\top + R^j, \ldots, C^j \hat{P}_k^{j,j}(C^j)^\top + R^j\right\}.$$

Therefore, the convergence of $\Gamma_k^{1,1}$ and $\Gamma_k^{2,2}$ with respect to k is determined by the convergence of $\hat{P}_k^{i,i}$ and $\hat{P}_k^{j,j}$, which can be guaranteed by the stabilizability of the pair (A, Q) and the detectability of the pairs (C^i, A) and (C^j, A) according to the theory of Riccati equations [45]. On the other hand, since

$$\Gamma_k^{1,2} = \mathrm{E}([z_{k+\tau}^i, z_{k-\mu+1:k}^i]^\top [z_{k-l+1:k}^j])$$

it is composed of entries in the form of $\mathrm{E}[z_{k+\alpha}^i (z_{k+\beta}^j)^\top]$, where $\alpha, \beta \in [-\max\{l, \mu\} + 1, \tau]$. We observe that $\mathrm{E}[z_{k+\alpha}^i (z_{k+\beta}^j)^\top]$ can be calculated based on the recursive equations in (35)–(36) and (33). Since τ, μ and l are finite, the convergence of the entries in $\Gamma_k^{1,2}$ is determined by the convergence of $P_k^{i,j}$ in (35), namely,

$$P_{k+1}^{i,j} = \hat{A}_k^i P_k^{i,j} (\hat{A}_k^j)^\top + Q. \tag{39}$$

Since (A, Q) is stabilizable and (C^i, A) and (C^j, A) are detectable, $\lim_{k\to\infty} \hat{A}_k^i$ and $\lim_{k\to\infty} \hat{A}_k^j$ exist. We write

$$\bar{A}^i := \lim_{k\to\infty} \hat{A}_k^i,\ \bar{A}^j := \lim_{k\to\infty} \hat{A}_k^j$$

and only focus on the steady-state form of (39), namely,

$$P_{k+1}^{i,j} = \bar{A}^i P_k^{i,j} (\bar{A}^j)^\top + Q. \tag{40}$$

Before continuing, we note that the stabilizability of the pair (A, Q) and the detectability of the pairs (C^i, A) and (C^j, A) ensure that $\bar{A}^i$ and $\bar{A}^j$ are stable matrices. Write

$$X_k^{i,j} = \operatorname{vec}(P_k^{i,j}).$$

From (40), we have

$$X_{k+1}^{i,j} = (\bar{A}^i \otimes \bar{A}^j) X_k^{i,j} + \operatorname{vec}(Q). \tag{41}$$

Since $\bar{A}^i$ and $\bar{A}^j$ are stable, from Lemma 2, $\bar{A}^i \otimes \bar{A}^j$ is a stable matrix as well. Furthermore, since Q is a constant matrix, there exists a bounded vector $\bar{X}^{i,j}$ such that

$$\lim_{k\to\infty} X_k^{i,j} = \bar{X}^{i,j}.$$

This implies that $\lim_{k\to\infty} \operatorname{vec}(P_k^{i,j})$ exists as well, which proves the convergence of $\{P_k^{i,j}\}$. In this way, the convergence of $\{\Gamma_k\}$ is achieved. Based on Lemma 1, we obtain the convergence of Υ_k^i. Since the convergence of Π_k^i is guaranteed by the stabilizability of (A, Q) and the detectability of (C^i, A), the existence of $\bar{\mathcal{T}}_{z^i \to z^j}$ is proved. Finally, we note that $\bar{\mathcal{T}}_{y^i \to y^j} \geq 0$ and $\bar{\mathcal{T}}_{z^i \to z^j} \geq 0$ follow from that $\Psi_k^i \geq \Phi_k^i \geq 0$ and $\Pi_k^i \geq \Upsilon_k^i \geq 0$. ■

The above theorem provides a generic convergence result for transfer entropy of linear dynamical systems. In the first part of Theorem 1, we show that the transfer entropy measure for the measurement processes converges for stable systems; the underlying intuition is that the state process $\{x_k\}$ exponentially converges to a stationary process, which enforces the convergence of the transfer entropy. One interesting observation, however, is that the convergence of the transfer entropy measure can be maintained beyond the stability of the system. To see this, we can consider the simple case that x_k is scalar-valued and $A > 1$ (unstable); in this case, although the covariance P_k of x_k goes to infinity as $k \to \infty$, we still have that for $X \in \mathbb{R}$,

$$\lim_{X\to\infty} \tilde{g}(X) = \lim_{X\to\infty} (X^{-1} + C^\top R^{-1} C)^{-1} = (C^\top R^{-1} C)^{-1}$$

(and a similar result for $\lim_{X\to\infty} \tilde{g}_i(X)$). This limit guarantees the existence of $\lim_{k\to\infty} \mathcal{T}_{y^i \to y^j}(k)$. For the transfer entropy measure defined for innovation processes, weaker conditions on detectability and stabilizability are sufficient to guarantee convergence.

For anomaly detection, the convergence of the nominal values of the countermeasures is important, as the causality countermeasures should remain constant during routine operations. The values of the causality countermeasures should change, however, when attack signals are inserted into the system. In the following, the asymptotic properties of the transfer entropy measures with respect to τ for fixed values of k , μ, and l are further analyzed. These results provide guidelines on the choice of τ in attack detection.

Theorem 2: Assume the values of k, μ, and l are fixed. Then,

1) if A is stable, it holds that
$$\lim_{\tau\to\infty} \mathcal{T}_{y^i \to y^j}(k+\tau) = \lim_{\tau\to\infty} \mathcal{T}_{y^j \to y^i}(k+\tau) = 0;$$
2) if (A, Q) is stabilizable and (C^i, A) and (C^j, A) are detectable, it holds that
$$\lim_{\tau\to\infty} \mathcal{T}_{z^i \to z^j}(k+\tau) = \lim_{\tau\to\infty} \mathcal{T}_{z^j \to z^i}(k+\tau) = 0.$$

Proof: From (17) and (19), we observe that if A is stable,

$$\lim_{\tau\to\infty} \bar{P}_{k+\tau} = \lim_{\tau\to\infty} \hat{P}_{k+\tau} = \check{P}$$

holds for all finite and fixed values of k, μ, and l, where $\check{P}$ is the unique stabilizing solution to the Lyapunov equation $P = h(P)$. From (18) and (20), this further implies

$$\lim_{\tau\to\infty} \Phi_{k+\tau}^i = \lim_{\tau\to\infty} \Psi_{k+\tau}^i = C^i \check{P} (C^i)^\top + R^i.$$

The first conclusion follows from the fact that $\mathcal{T}_{y^j \to y^i}(k+\tau) = \frac{1}{2} \log \frac{\det \Psi_{k+\tau}^i}{\det \Phi_{k+\tau}^i}$ [see (9)].

To prove the second conclusion, we start with an analysis of Γ_k defined in (27). We decompose Γ_k as

$$\Gamma_k = \begin{bmatrix} \operatorname{Cov}(z_{k+\tau}^i) & \\ \operatorname{Cov}(z_{k+\tau}^i, [z_{k-\mu+1:k}^i, z_{k-l+1:k}^j]^\top)^\top & \\ & \operatorname{Cov}(z_{k+\tau}^i, [z_{k-\mu+1:k}^i, z_{k-l+1:k}^j]^\top) \\ & \operatorname{Cov}([z_{k-\mu+1:k}^i, z_{k-l+1:k}^j]^\top) \end{bmatrix}.$$

From (33) and (36), we have

$$\lim_{\tau\to\infty} \operatorname{Cov}(z_{k+\tau}^i, [z_{k-\mu+1:k}^i, z_{k-l+1:k}^j]^\top) = 0$$

provided (A, Q) is stabilizable and (C^i, A) is detectable. This implies

$$\lim_{\tau\to\infty} \Gamma_k = \begin{bmatrix} \lim_{\tau\to\infty} \operatorname{Cov}(z_{k+\tau}^i) & 0 \\ 0 & \times \end{bmatrix}$$

where $\times$ denotes a matrix that does not affect our analysis. From (26) and the definition of Γ_k, we have

$$\lim_{\tau\to\infty}\Pi^i_{k+\tau} = \lim_{\tau\to\infty}\Upsilon^i_{k+\tau}.$$

Based on equation (23), we further have

$$\lim_{\tau\to\infty}\mathcal{T}_{z^j\to z^i}(k+\tau) = 0.$$

The derivations for $\lim_{\tau\to\infty}\mathcal{T}_{z^i\to z^j}(k+\tau)=0$ can be obtained following the same line of argument given the stabilizability of (A,Q) and the detectability of (C^j,A). ■

The above results indicate that the causality relationship between two measurement/innovation processes can be blurred as the value of τ approaches infinity. This observation is critical, as in order to apply the transfer entropy measures in attack detection, the causal and noncausal relationships are required to be distinguishable in terms of the transfer entropy values, so that changes in these values can be observed. Therefore, a general guideline is that the value of τ should be kept relatively small.

Convergence of the transfer entropy measure with respect to μ and l may be obtained following similar ideas; in this paper, however, we do not investigate these properties, as they do not affect the application of transfer entropy to attack detection. In [36], the authors point out that the values of μ and l also play a critical role in numerically identifying the causal relationships between process variables, and need to be carefully tuned according to certain rules. These rules can be used in this paper as well.

Finally, we note that the transfer entropy analysis in this paper is built on the overall state-space model of the considered system; that is, some of the states may represent the states of a dynamic feedback controller. In this regard, the proposed analysis applies equally to the case when the system has output feedback. The numerical calculation of the transfer entropy is data driven, and thus is not affected by feedback structures.

IV. Transfer Entropy for Anomaly Detection

In this section, we analyze the effectiveness of the causality countermeasures against different attack policies. Specifically, we consider four attack scenarios: DoS, replay, innovation-based deception, and data injection attacks. The first three scenarios of attacks can be represented in a unified way of the form $\eta^j_k = y^j_k + D_k\theta_k$, where the term $D_k\theta_k$ depends on the particular attack policy considered; we note that the term $D_k\theta_k$ depends on the particular attack policy considered; in particular, in [16], the authors showed that different attacks could lead to different underlying system structures (cf., [16, Remark 2, Fig. 1]). In the literature, many attack detection approaches and performance analysis results are developed by analyzing the features of the attacks (namely, the properties of the $D_k\theta_k$ term) considered, while in our work the information of the $D_k\theta_k$ term is not utilized in calculating the transfer entropy countermeasures. This point is important for attack detection, since the malicious agents would not inform the system what type of attack policies will be used and consequently the $D_k\theta_k$ term is in general unknown.

When attacks on sensor measurements are concerned, we denote the sensor being attacked as sensor j, and η^j_k to denote the received contaminated measurement, and use ζ^j_k to denote the received contaminated innovation. For notational brevity, we define an indicator variable γ^j_k for sensor j such that $\gamma^j_k = 1$ means sensor j is attacked at time instant k and $\gamma^j_k = 0$ otherwise. When the attacks act on the process equation, we use γ_k to denote the indicator variable. The transfer entropy countermeasures in attack detection, however, does not require knowledge of γ^j_k or γ_k. The basic assumption in applying the transfer entropy countermeasures for attack detection is that the process is operating in a normal state before it is attacked, so that the values of the transfer entropy between the monitored variables during the no-attack stage can be computed, and changes in the causality measures can be observed when the attacks come into effect.

A. Principle and Limitations

Since transfer entropy measures the directed causation relationship from one random process to the other, the underlying principle of using transfer entropy countermeasures for anomaly detection is whether the existence of anomalies change the causality relationship between the random processes. Given two random processes y^i and y^j, one way to detect the existence of anomalies on y^j (based on the transfer entropy measures defined for y^i and y^j) is that $\mathcal{T}_{y^i\to y^j} > 0$ holds for the nominal anomaly-free case; while the appearance of anomalies change this causation relationship so that the anomalies can be detected. A design technique to ensure this type of cause-effect pairs is presented in Section V-B.

Transfer entropy measures cannot be applied when the anomalies do not alter the cause-effect relationships in the targeted variables (e.g., zero-dynamics attacks [19]); nor can they be used for anomaly detection for y^j when $\mathcal{T}_{y^j\to y^i} = 0$ holds for the unattacked case, unless the anomalies introduce an additional causation relationship such that $\mathcal{T}_{y^j\to y^i} > 0$. The other issue in applying transfer entropy measure is the effect of noise, which is discussed in Section VI-H. These are the basic limitations in utilizing transfer entropy for anomaly detection. The rest of this section is devoted to the application of transfer entropy countermeasures to the four types of attacks considered in this paper.

B. DoS Attacks

DoS attacks deny the successful transmission of data between nodes (e.g., sensors, actuators, and controllers) in control systems. In networked sampled-data control systems, zero-order-hold modules are used to generate continuous-time signals between the update times of the discrete-time signals. When a measurement y^j_k of sensor j at time instant k is not received by a remote controller or estimator, the previous received measurement of sensor j is often used instead. In this way, the contaminated measurement process η^j_k can be defined as

$$\eta^j_k = \begin{cases} y^j_k, & \text{if } \gamma^j_k = 0 \\ \eta^j_{k-1}, & \text{otherwise} \end{cases}. \tag{42}$$

From the above equation, we observe the switching structure. This change in system dynamics will directly lead to a change of transfer entropy between the sensor measurement processes when the sensor communication is under DoS attack.

C. Replay Attacks

Replay attacks prevent the system nodes from knowing the true data and normally have two phases. In the first phase, the adversary records the process data for a certain period of time and replays the recorded data repeatedly during the second phase so that destruction on the system can be performed in a stealthy way. For the causality countermeasure, the successful detection of this type of stealthy attack is intuitive, as it is difficult for the replayed data to maintain the same causal relationship with data obtained from other resources (e.g., another sensor). To be specific, the repeatedly replayed data can be viewed as a periodic deterministic signal independent of the true data. Assuming the measurements of sensor j are corrupted by a replay attack, we thus have

$$\mathcal{T}_{y^i\to\eta^j}(k+\tau)=\mathcal{T}_{\eta^j\to y^i}(k+\tau)=0. \tag{43}$$

However, given the fact that some causal relationship exists between sensor i and sensor j, $\mathcal{T}_{y^i\to y^j}(k+\tau)\neq 0$ or $\mathcal{T}_{y^j\to y^i}(k+\tau)\neq 0$ hold, indicating that the replay attacks can be detected by the transfer entropy countermeasure.

D. Innovation-Based Deception Attacks

When the measurements are preprocessed on the sensor side such that innovation sequences are sent to the remote estimator/controller, the adversary might try to downgrade the performance of the system by performing attacks on the innovation process. In this section, we consider two possible policies and investigate the effect of these attacks on the transfer entropy measure.

The first policy replaces the original innovation sequences $\{z_k^j\}$ with a sequence of realizations of a Gaussian white noise process $\{\zeta_k^j\}$ with covariance $\Xi_k^j:=C^j\hat{P}_k^{j,j}(C^j)^\top+R^j$. Although ζ_k^j is i.i.d. zero-mean Gaussian and has the same covariance as z_k^j, the transfer entropy can capture that the innovation processes corresponding to different sensors are correlated [see (33)–(36)]. Therefore, the replacement of z_k^j with ζ_k^j can be detected based on the transfer entropy countermeasure. By definition, we have

$$\begin{aligned}\mathcal{T}_{z^i\to\zeta^j}(k+\tau)=&\int f\left(\zeta_{k+\tau}^j,\zeta_{k-\mu+1:k}^j,z_{k-l+1:k}^i\right)\\&\times\log\frac{f(\zeta_{k+\tau}^j|\zeta_{k-\mu+1:k}^j,z_{k-l+1:k}^i)}{f(\zeta_{k+\tau}^j|\zeta_{k-\mu+1:k}^j)}\\&\times\mathrm{d}\zeta_{k+\tau}^j\mathrm{d}\zeta_{k-\mu+1:k}^j\mathrm{d}z_{k-l+1:k}^i.\\=&\int f(\zeta_{k+\tau}^j,\zeta_{k-\mu+1:k}^j)\\&\times\log\frac{f(\zeta_{k+\tau}^j|\zeta_{k-\mu+1:k}^j)}{f(\zeta_{k+\tau}^j|\zeta_{k-\mu+1:k}^j)}\mathrm{d}\zeta_{k+\tau}^j\mathrm{d}\zeta_{k-\mu+1:k}^j\\=&\,0.\end{aligned}$$

A more generic attack policy that treats the above policy as a special case is to perform an affine operation on the innovation process such that

$$\zeta_k^j=T_k^jz_k^j+b_k^j. \tag{44}$$

To maintain the stealthiness of the attack, b_k^j is chosen to be an i.i.d. zero-mean Gaussian process, and its covariance Σ_k^j is chosen together with the linear map T_k^j such that

$$\Sigma_k^j:=\Xi_k^j-T_k^j\Xi_k^j(T_k^j)^\top. \tag{45}$$

Now, we analyze the effect of this attack policy on the transfer entropy. Denote the starting time of the attack as ς, so that we have

$$\zeta_k^j=\begin{cases}z_k^j, & \text{if } k<\varsigma\\ T_k^jz_k^j+b_k^j, & \text{if } k\geq\varsigma\end{cases}.$$

To simplify our analysis, we use the notation

$$\bar{\Gamma}_k:=\begin{bmatrix}\bar{\Gamma}_k^{1,1} & \bar{\Gamma}_k^{1,2}\\ (\bar{\Gamma}_k^{1,2})^\top & \bar{\Gamma}_k^{2,2}\end{bmatrix} \tag{46}$$

to denote the covariance matrix of the joint distribution of $\zeta_{k+\tau}^j$, $\zeta_{k-\mu+1:k}^j$ and $z_{k-l+1:k}^i$, where $\bar{\Gamma}_k^{1,1}:=\mathrm{Cov}([\zeta_{k+\tau}^j,\zeta_{k-\mu+1:k}^j]^\top)$ and $\bar{\Gamma}_k^{2,2}:=\mathrm{Cov}([z_{k-l+1:k}^i]^\top)$, and $\bar{\Gamma}_k^{1,2}:=\mathrm{E}([\zeta_{k+\tau}^j,\zeta_{k-\mu+1:k}^j][z_{k-l+1:k}^i]^\top)$. It is easy to verify that the $\bar{\Gamma}_k^{1,1}$ and $\bar{\Gamma}_k^{2,2}$ matrices are block-diagonal matrices, and in particular, we have

$$\bar{\Gamma}_k^{1,1}=\mathrm{Cov}([z_{k+\tau}^j,z_{k-\mu+1:k}^j]^\top).$$

However, we also observe

$$\begin{aligned}\mathrm{E}[\zeta_k^j(z_t^i)^\top]&=\mathrm{E}[T_k^jz_k^j(z_t^i)^\top+b_k^j(z_t^i)^\top]\\&=T_k^j\mathrm{E}[z_k^j(z_t^i)^\top]\end{aligned} \tag{47}$$

which means that $\bar{\Gamma}_k^{1,2}\neq\mathrm{E}([z_{k+\tau}^j,z_{k-\mu+1:k}^j]^\top[z_{k-l+1:k}^i])$ in general. From Lemma 1, the conclusion here is that the attack will lead to a change in the causality measure, but the amount of change is determined by the choice of T_k^j and Σ_k^j. In particular, the case that $T_k^j=-I$ is the most critical, which has been proved in the sense that it is the worst case choice such that the corresponding state estimation error is maximized [30]; in fact, this attack is difficult to be detected, as is summarized in the following result.

Theorem 3: Consider the system (1)–(2) and the innovation-based deception attack (44) with $T_k^j=-I$ and $b_k^j=0$. Assume the attack starts from time instant ς. For $k>\varsigma+\mu$, it holds that

$$\mathcal{T}_{z^i\to\zeta^j}(k+\tau)=\mathcal{T}_{z^i\to z^j}(k+\tau).$$

Proof: Since $T_k^j=-I$, (47) becomes

$$\mathrm{E}[\zeta_k^j(z_t^i)^\top]=-\mathrm{E}\left[z_k^j(z_t^i)^\top\right]. \tag{48}$$

Write

$$\tilde{\Gamma}_k:=\begin{bmatrix}\tilde{\Gamma}_k^{1,1} & \tilde{\Gamma}_k^{1,2}\\ (\tilde{\Gamma}_k^{1,2})^\top & \tilde{\Gamma}_k^{2,2}\end{bmatrix} \tag{49}$$

to denote the covariance matrix of the joint distribution of $z_{k+\tau}^j$, $z_{k-\mu+1:k}^j$ and $z_{k-l+1:k}^i$, where $\tilde{\Gamma}_k^{1,1}:=\mathrm{Cov}([z_{k+\tau}^j,z_{k-\mu+1:k}^j]^\top)$ and $\tilde{\Gamma}_k^{2,2}:=\mathrm{Cov}([z_{k-l+1:k}^i]^\top)$, and $\tilde{\Gamma}_k^{1,2}:=\mathrm{E}([z_{k+\tau}^j,z_{k-\mu+1:k}^j][z_{k-l+1:k}^i]^\top)$. The relationship in

(48) indicates that for $k \geq \varsigma + \mu - 1$, we have

$$\bar{\Gamma}_k := \begin{bmatrix} \tilde{\Gamma}_k^{1,1} & -\tilde{\Gamma}_k^{1,2} \\ -(\tilde{\Gamma}_k^{1,2})^\top & \tilde{\Gamma}_k^{2,2} \end{bmatrix}. \tag{50}$$

Correspondingly, we define $\tilde{\Lambda}_k = (\bar{\Gamma}_k)^{-1}$ and

$$\tilde{\Lambda}_k := \begin{bmatrix} \tilde{\Lambda}_k^{1,1} & \tilde{\Lambda}_k^{1,2} \\ (\tilde{\Lambda}_k^{1,2})^\top & \tilde{\Lambda}_k^{2,2} \end{bmatrix} \tag{51}$$

such that $\tilde{\Lambda}_k^{1,1}$ has a same size as $\tilde{\Gamma}_k^{1,1}$. To aid our analysis, we need an alternative decomposition of $\bar{\Gamma}_k$:

$$\bar{\Gamma}_k := \begin{bmatrix} \hat{\Gamma}_k^{1,1} & \hat{\Gamma}_k^{1,2} \\ (\hat{\Gamma}_k^{1,2})^\top & \hat{\Gamma}_k^{2,2} \end{bmatrix} \tag{52}$$

where $\hat{\Gamma}_k^{1,1} := \mathrm{Cov}(\zeta_{k+\tau}^j)$, $\hat{\Gamma}_k^{2,2} := \mathrm{Cov}([\zeta_{k-\mu+1:k}^j, z_{k-l+1:k}^i]^\top)$, and $\hat{\Gamma}_k^{1,2} := \mathrm{E}([\zeta_{k+\tau}^j][\zeta_{k-\mu+1:k}^j, z_{k-l+1:k}^i]^\top)$. We represent the corresponding decomposition of $\tilde{\Lambda}_k$ as

$$\tilde{\Lambda}_k := \begin{bmatrix} \hat{\Lambda}_k^{1,1} & \hat{\Lambda}_k^{1,2} \\ (\hat{\Lambda}_k^{1,2})^\top & \hat{\Lambda}_k^{2,2} \end{bmatrix} \tag{53}$$

such that $\hat{\Lambda}_k^{1,1}$ has a same size as $\hat{\Gamma}_k^{1,1}$. From Lemma 1, we have

$$\begin{aligned} &\mathrm{Cov}\left(\zeta_{k+\tau}^j | \zeta_{k-\mu+1:k}^j, z_{k-l+1:k}^i\right) \\ &= \hat{\Gamma}_k^{1,1} - \hat{\Gamma}_k^{1,2}(\hat{\Gamma}_k^{2,2})^{-1}(\hat{\Gamma}_k^{1,2})^\top. \end{aligned} \tag{54}$$

By definition, $\hat{\Lambda}_k^{1,1}$ is a subblock of $\tilde{\Lambda}_k^{1,1}$. On the other hand, from the matrix inversion lemma, we have

$$\hat{\Lambda}_k^{1,1} = \left[\hat{\Gamma}_k^{1,1} - \hat{\Gamma}_k^{1,2}\left(\hat{\Gamma}_k^{2,2}\right)^{-1}\left(\hat{\Gamma}_k^{1,2}\right)^\top\right]^{-1}. \tag{55}$$

From (50) and (51), we have

$$\tilde{\Lambda}_k^{1,1} = \left[\tilde{\Gamma}_k^{1,1} - \tilde{\Gamma}_k^{1,2}(\tilde{\Gamma}_k^{2,2})^{-1}(\tilde{\Gamma}_k^{1,2})^\top\right]^{-1}. \tag{56}$$

Combining (49), (55), and (56) as well as the fact that $\hat{\Lambda}_k^{1,1}$ is a subblock of $\tilde{\Lambda}_k^{1,1}$, we conclude that for $k \geq \varsigma + \mu - 1$, it holds that

$$\begin{aligned} &\mathrm{Cov}\left(\zeta_{k+\tau}^j | \zeta_{k-\mu+1:k}^j, z_{k-l+1:k}^i\right) \\ &= \mathrm{Cov}\left(z_{k+\tau}^j | z_{k-\mu+1:k}^j, z_{k-l+1:k}^i\right). \end{aligned} \tag{57}$$

Moreover, by the definition of ζ_k^j in (44), we have

$$\mathrm{Cov}\left(\zeta_{k+\tau}^j | \zeta_{k-\mu+1:k}^j\right) = \mathrm{Cov}\left(z_{k+\tau}^j | z_{k-\mu+1:k}^j\right) \tag{58}$$

which implies

$$\mathcal{T}_{z^i \to \zeta^j}(k+\tau) = \mathcal{T}_{z^i \to z^j}(k+\tau) \tag{59}$$

for $k \geq \varsigma + \mu - 1$. ■

The above result indicates that no discrepancy in the transfer entropy measure can be observed after a short time period from the starting time of the attack. Fortunately, we still have a finite-time window to detect the attack. The rationale is that when the sequence of $\{\zeta_k^j\}$ utilized to calculate the transfer entropy measure $\mathcal{T}_{z^i \to \zeta^j}$ contains some of the uncontaminated innovation segments z_t^j for $t < \varsigma$, (48) together with a similar argument as that in the above proof imply that changes in the transfer entropy measure can be observed at the initial stage of the attack. The window length should be no larger than $\tau + \mu$. The data-driven evaluation of the transfer entropy measure, however, might help enlarge this window length, due to the utilization of historical innovation data.

E. Data Injection Attacks

Finally, we investigate the effect of input data injection attacks. From the system (4), except for some special cases (e.g., the zero-dynamic attacks discussed in [19]), we note that the additional unknown inputs will lead to changes of the virtual equivalent process noise input statistics, which consequently leads to changes of the transfer entropy values.

V. Discussions on Implementations

In this section, discussions on the practical application of the considered transfer entropy countermeasure are presented.

A. Detector Design

For engineering applications, it is necessary to build detectors such that the existence of anomalies can be automatically detected. This can be achieved based on the transfer entropy countermeasure as well. To do this, given two signals $\{\alpha_k\}$ and $\{\beta_k\}$, it suffices to obtain the nominal average transfer entropy reading $\mathcal{T}_{\alpha\to\beta}^{\mathrm{norm}}$ empirically based on normal process operation data and consider the decision rule

$$\gamma_{\alpha\to\beta}(k) = \begin{cases} 1, & |\mathcal{T}_{\alpha\to\beta}(k) - \mathcal{T}_{\alpha\to\beta}^{\mathrm{norm}}| \geq \delta \cdot \mathcal{T}_{\alpha\to\beta}^{\mathrm{norm}} \\ 0, & \text{otherwise} \end{cases} \tag{60}$$

where δ is a user-specified tuning knob to adjust the risk of missed and false alarms. Since the transfer entropy measure is not a random process, the concept of missed alarm rates and false alarm rates in signal detection and estimation theory [46] cannot be directly applied. On the other hand, as the transfer entropy is evaluated in a data-driven fashion, when a large number of data points are utilized to calculate the transfer entropy at each time instant, the choice of a relatively small value of δ would be sufficient to detect the existence of attacks. A consequence, however, is that a relatively large detection delay would be incurred due to the large number of historical data used. In Section VI, to provide an informative picture of the behavior of transfer entropy with respect to the attacks, the transfer entropy readings rather than the outputs of the detector for a fixed value of δ are provided.

B. Design of Artificial Causal Sensor Pairs

The underlying principle of using the transfer entropy countermeasure for anomaly detection is to detect the changes of transfer entropy caused by the anomaly. To be specific, let $\{\beta_k\}$ denote the process subject to potential attacks, and let $\{\alpha_k\}$ denote another process. One simple way to detect whether β_k is attacked is to choose $\{\alpha_k\}$ such that $\mathcal{T}_{\alpha\to\beta} \neq 0$; when β_k is contaminated by an attack, it is very likely that this causal relationship would be altered (see Section VI for examples) and

thus can be detected by evaluating $\mathcal{T}_{\alpha\to\beta}$. For large-scale control systems, however, it is not always easy to find a sensor pair $\{y_k^i, y_k^j\}$ such that $\mathcal{T}_{y^i\to y^j}\neq 0$ holds, due to the complex cause-effect relationships for large systems, and the existence of disturbances and noise. One question is whether systematic techniques can be adopted to these scenarios. We now show one possible technique. Let

$$\check{x}_{k+1} = \check{A}\check{x}_k + \check{B}u_k + \check{w}_k \tag{61}$$

$$\check{y}_k = \check{C}\check{x}_k + \check{v}_k \tag{62}$$

where u_k denotes the external input, and $\check{x}_k$ may contain controller states utilized to achieve closed-loop performance specifications based on output feedback. Let b denote one row of $\check{B}$, and define $\bar{x}_{k+1} := bu_k$. In addition, let $\hat{x}_{k+1} = \check{x}_k$. Letting $\hat{v}_k = \check{v}_{k-1}$ and introducing an additional scalar measurement noise process $\bar{v}_k$, we construct the following augmented system:

$$\underbrace{\begin{bmatrix}\hat{x}_{k+1}\\ \bar{x}_{k+1}\\ \check{x}_{k+1}\end{bmatrix}}_{x_{k+1}} = \underbrace{\left[\begin{array}{c|c|c} 0 & 0 & I\\ \hline 0 & 0 & 0\\ \hline 0 & 0 & \check{A}\end{array}\right]}_{A}\underbrace{\begin{bmatrix}\hat{x}_k\\ \bar{x}_k\\ \check{x}_k\end{bmatrix}}_{x_k} + \underbrace{\begin{bmatrix}0\\ b\\ \check{B}\end{bmatrix}u_k + \begin{bmatrix}0\\ 0\\ I\end{bmatrix}\check{w}_k}_{w_k} \tag{63}$$

$$\underbrace{\begin{bmatrix}\hat{y}_k\\ \bar{y}_k\end{bmatrix}}_{y_k} = \underbrace{\left[\begin{array}{c|c|c} \check{C} & 0 & 0\\ \hline 0 & 1 & 0\end{array}\right]}_{C}\underbrace{\begin{bmatrix}\hat{x}_k\\ \bar{x}_k\\ \check{x}_k\end{bmatrix}}_{x_k} + \underbrace{\begin{bmatrix}\hat{v}_k\\ \bar{v}_k\end{bmatrix}}_{v_k}. \tag{64}$$

In (63), u_k can either be the original external input signal or an artificial random excitation signal; we note that $\bar{y}_k$ is an artificial measurement of bu_k (since u_k is available) with $\bar{v}_k$ being an artificial measurement noise process, and $\hat{y}_k = \check{y}_{k-1}$ corresponds to the delayed measurement of $\check{x}_k$. The delay in (63) (i.e., $\hat{x}_{k+1} = \check{x}_k$) is intentionally included to enhance the causality from $\bar{y}_k$ to $\hat{y}_k$ (or equivalently, from bu_k to $\check{y}_k$). If the system is controllable and observable,[1] there exists an element $\hat{y}_k^j$ in $\hat{y}_k$ such that $\bar{y}_k \to \hat{y}_k^j$ holds. In this way, a pair of sensor measurements with guaranteed causality can be obtained, which is also of particular help in applying the transfer entropy countermeasure to the case when only one sensor is available. Since $\bar{y}_k$ does not need to be measured by an actual sensor, the measurements are naturally secured, which further enhances the reliability and applicability of the transfer entropy countermeasures. From (63)–(64), we observe that the analysis in Section IV applies provided the noise terms w_k and v_k satisfy the assumptions of the corresponding terms in (1)–(2). The discussions here will be illustrated in Section VI through the Tennessee-Eastman process [47], [48]. In addition, we note that the idea in this artificial causal relationship construction technique is similar to that of the physical watermarking approach utilized in [49] for detecting replay attacks.

Finally, we note that the idea of the above procedure also provides a method to identify a virtual process $\bar{y}_k$ for a chosen sensor y_k^j (which is an element of $\check{y}_k$), so that $\bar{y}_k \to y_k^j$ holds and the transfer entropy countermeasure can be utilized to detect the existence of attacks for the chosen sensor y^j. To see this, it suffices to note that if the state-space model of the subsystem that represents the dynamics from a specific input u_k^i (which is a component of u_k) to the chosen sensor output y_k^j is controllable and observable; then, the transfer function from the input u_k^i to the sensor output y_k^j is nonzero, which indicates a causal relationship from u_k^i and the specified sensor output y_k^j and helps construct a signal $\bar{y}_k = u_k^i$ with the required causality property.

[1]Note that this condition is trivial as a controllable and observable realization of the system in (63)–(64) can be always obtained.

C. Transfer Entropy Between Two Sensor Groups

Transfer entropy quantifies the information flow from one random process to another. Two random processes considered in the definition are neither required to be scalar-valued nor to have the same dimension. In this sense, the transfer entropy countermeasure can not only be used for anomaly detection between two scalar-valued sensors, but also be applied to two sensor groups. The data-driven transfer entropy evaluation algorithm proposed in [36] focuses on scalar-valued sensors, which is why we limit the discussions to scalar-valued sensors in Section VI. This further puts forward the requirement of developing computational efficient transfer entropy algorithms for vector-valued random processes. We plan to explore this problem in a separate work. Finally, note that the theoretical developments in this paper apply to both scalar and vector cases, as no assumption is made on the dimension of the sensors in the derivations.

VI. Numerical Example

In this section, we evaluate the effectiveness of the transfer entropy measures through a numerical example and make comparisons to alternative approaches.

A. System Setup

We consider a state-space model in the form of (61)–(62) extracted from the simplified Tennessee-Eastman challenge problem [47], [48], which is composed of eight states, four inputs, and ten sensors. The matrices $\check{A}$, $\check{B}$, and $\check{C}$ are obtained based on their continuous-time counterparts in [48, Table 3] with sampling period 0.2 h and are omitted here due to space limitation. For illustration purpose, we assume u_k to be a zero-mean white Gaussian process with covariance $\hat{Q} = \mathrm{diag}\{1.46, \ldots, 1.46\}$; based on the nominal values of the input variables in [48, Table 1], this choice can be regarded as a small perturbation added to the original nominal control inputs. We assume the covariance of $\check{w}_k$ to be $\check{Q} = \mathrm{diag}\{0.05, \ldots, 0.05\}$, and the covariance of v_k in (64) to be $\mathrm{diag}\{0.002, \ldots, 0.002\}$; these values are first set to be small here to aid the later sensitivity analysis. To illustrate the proposed results, we choose b in (63) as the first row of $\check{B}$; the two sensors used for transfer entropy evaluation are the artificial sensor measuring bu_k together with the fifth sensor in $\hat{y}_k$, which shall be named as sensor i and sensor j, respectively, for national consistence with the discussions in Sections II–III. The goal is to verify whether the causality countermeasure can be used to detect the cyber-attacks on sensor j. The values of transfer entropy countermeasures are numerically evaluated based on historical data according to the algorithm introduced in [36] with $\tau = 1$, $\mu = 2$, and $l = 2$. At each time instant, the length of historical data points utilized to calculate the transfer entropy

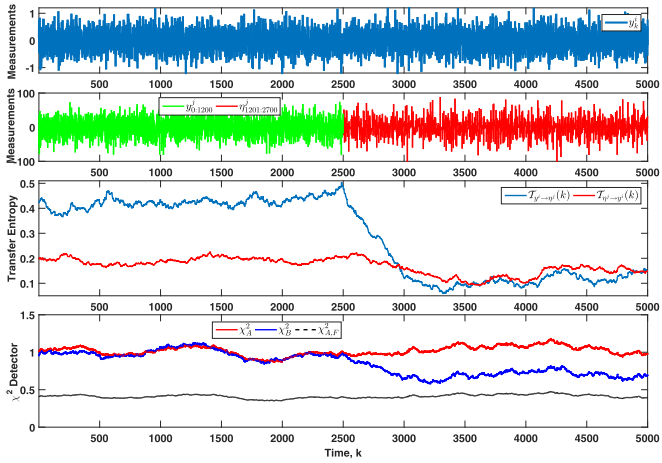


Fig. 2. Anomaly detection for DoS attack.

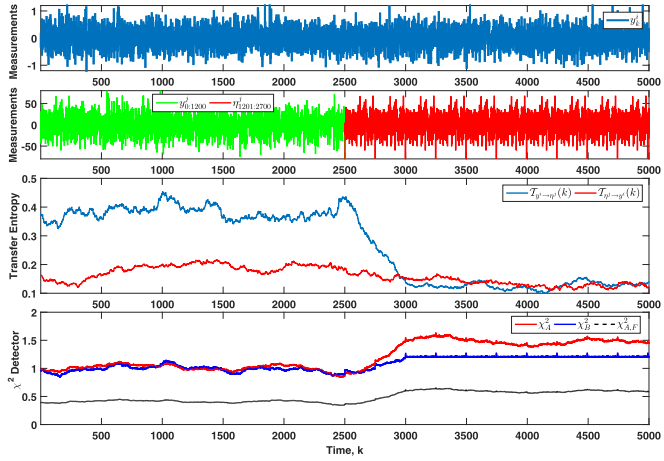


Fig. 3. Replay attack (record period = 250).

is $L = 500$. For all the scenarios, we assume the attacks are inserted into the system at $k = 2500$. We assume sensor j to be the sensor being attacked, although this information is neither needed nor used to calculate the transfer entropy values.

For comparison purpose, two χ^2 detectors are considered, which are implemented as in [9]. The first detector (which we name as "χ^2_A detector") is based on the information from both sensor i and sensor j. Write $y_k := [(y^i_k)^\top (y^j_k)^\top]^\top$. The χ^2_A detector first calculates the innovation process $\{z_{k,A}\}$ based on $\{y_k\}$ and the standard Kalman filter equations, and then evaluates

$$\frac{1}{L}\sum_{t=k-L+1}^{k} z_{k,A}^\top \mathcal{P}_A z_{k,A} \underset{\mathrm{H}_1}{\overset{\mathrm{H}_0}{\lessgtr}} \delta_A \tag{65}$$

where $\mathcal{P}_A$ denotes the steady-state covariance of $z_{k,A}$, δ_A denotes a prespecified threshold, and H_1 denotes a triggered alarm. Note that when the local innovation values z^i (or z^j) calculated only based on the information sensor i (or sensor j) are transmitted [see (21)], the corresponding y^i_k and y^j_k values are first reconstructed based on the Kalman filter recursions to implement the χ^2_A detector.

The second detector "χ^2_B detector" is based on the information from sensor j and is implemented in a similar way as (65):

$$\frac{1}{L}\sum_{t=k-L+1}^{k} z_{k,B}^\top \mathcal{P}_B z_{k,B} \underset{\mathrm{H}_1}{\overset{\mathrm{H}_0}{\lessgtr}} \delta_B. \tag{66}$$

From the above discussion, we note that these χ^2 detectors are model-based rather than data-driven as the system model is required in the Kalman filter calculations, and that an alarm will not be triggered unless the magnitude of the innovation sequence becomes sufficiently large. To provide a fair comparison of the countermeasures, we provide the values of the terms on the left-hand side of the inequalities in (65) and (66) (which we shall term as the *readings* of the χ^2 detectors), rather than those of the binary-valued decision variables.

B. Nominal Transfer Entropy Levels

As indicated by the transfer entropy readings in Figs. 2–6 for $k \le 2500$, $\mathcal{T}_{y^i \to y^j} > \mathcal{T}_{y^j \to y^i}$ holds, which is consistent with the

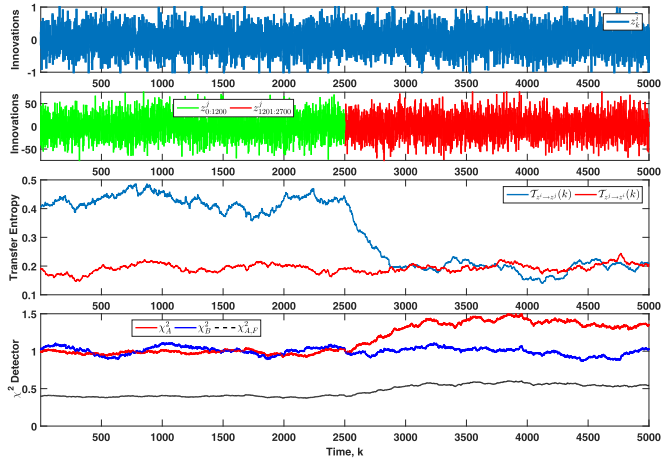


Fig. 4. Innovation-based false data attack, scenario I.

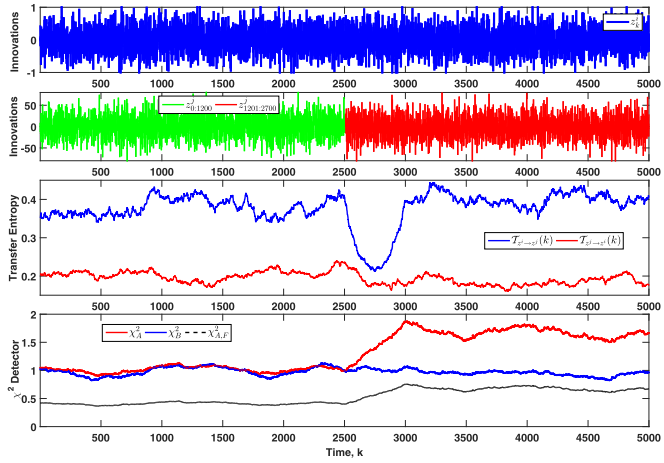


Fig. 5. Innovation-based false data attack, scenario II.

fact that y^i and y^j have an input–output relationship. Since the transfer entropy measures are numerically evaluated in a data-driven fashion, the value of $\mathcal{T}_{y^j \to y^i}$ does not exactly equal 0, and the level of $\mathcal{T}_{y^j \to y^i}$ serves as an approximate indication of what transfer entropy values to expect when a causation relationship does not exist from one random variable to the other. This level

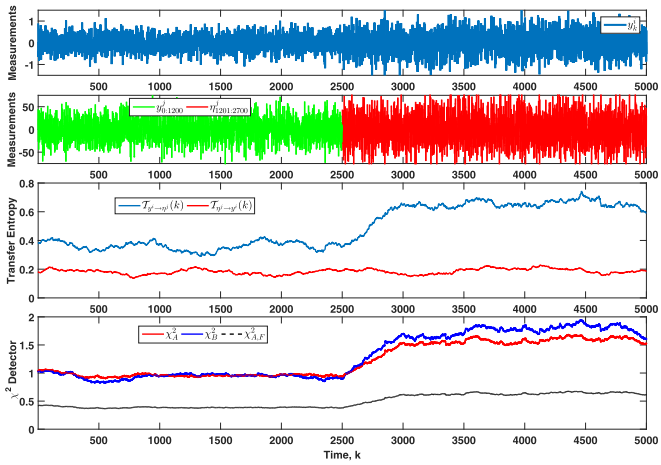


Fig. 6. Input data injection attacks.

is thus a baseline for analyzing the sensitivity of the transfer entropy countermeasure with respect to the anomalies.

C. DoS Attack

To illustrate the impact of DoS attacks on the transfer entropy measures, the zero-order hold protocol is implemented for the system considered by randomly generating γ_k^j with a DoS rate of 60%. The obtained measurement data and transfer entropy sequences are provided in Fig. 2, where a change in transfer entropy can be observed after the attacks are inserted on sensor j. For this case, the reading of χ_A^2 detector does not have an obvious response to the change of data caused by the DoS attack; decrease of the χ_B^2 reading is observed, although this change cannot raise an alarm for attacks due to that only a sufficient increase in the readings can cause a detection alarm according to (66).

D. Replay Attack

Now we focus on the effect of replay attacks. Consider the scenario that the adversary first records the data of sensor j for 250 samples and, then, replay these recorded data to perform an attack on the same sensor. The responses of the transfer entropy measure and the χ^2 detectors are provided in Fig. 3. From this figure, we observe that an abrupt change can be observed in the transfer entropy countermeasure. Increased readings of the two χ^2 detectors in respond to the replay attacks are observed as well, and in particular, the reading increase of the χ_A^2 detector is more obvious than that of the χ_B^2 detector, potentially due to the utilization of the healthy data from sensor i.

E. Innovation-Based Deception Attack

Two scenarios are considered here to illustrate the effect of innovation-based deception attacks; according to Section IV-D, an i.i.d. random process $b_k^j \sim \mathcal{N}(0, \Sigma_k^j)$ is generated to construct the fake innovation sequence $\{\zeta_k^j\}$, where Σ_k^j is calculated according to (45) and the specific choice of T_k^j. In the first scenario, starting from $k = 2500$, a fake white noise sequence $\{\zeta_k^j\}$ with the same covariance as that of the original innovation sequence $\{z_k^j\}$ is used to replace the original innovation sequence, which corresponds to the case $T_k^j = 0$. The results are provided in Fig. 4, where an obvious change in the transfer entropy measure again can be observed for $k \geq 2500$; reading changes in the χ_A^2 detector are also observed, although no obvious change is observed in the readings of the χ_B^2 detector.

The second scenario focuses on attacks of the form $\xi_k^j = T_k^j z_k^j + b_k^j$, with $T_k^j = -1$; the responses of transfer entropy and χ^2 detectors are shown in Fig. 5. For this case, an undershoot in the transfer entropy response happens at the attack; after that, the response returns to approximately the same level as that before the occurrence of the attacks, which is consistent with our theoretical analysis. An obvious reading change in the χ_A^2 detector is also observed, although no obvious change is observed in the readings of the χ_B^2 detector.

F. Data Injection Attack

In this section, we numerically evaluate the effect of data injection attacks. To do this, an input attack signal $a_k \in \mathbb{R}^4$ is chosen as a zero-mean Gaussian process with variance $\text{diag}\{1.5, \ldots, 1.5\}$ for $k \geq 2500$, and is mixed into the input u_k in (63). The responses of the transfer entropy and the χ^2 detectors are provided in Fig. 6, where step changes in the transfer entropy curve and the readings of the χ_A^2 detector are observed at $k = 2500$.

G. Comparison With χ^2 Detectors

From the above discussions, we observe that the proposed transfer entropy countermeasures are capable of sensing the existence of attacks for all the considered scenarios. We also observe that the χ_A^2 detector can respond to the attacks for most scenarios (with exception for the DoS attack in Fig. 2). We further compare further the χ_A^2 detector and the transfer entropy countermeasures in this section.

First, we note that for certain cases, the χ_A^2 detector can behave better than the transfer entropy countermeasure; an example can be found in Fig. 5, where a transient undershoot response is observed for the transfer entropy countermeasure while the χ_A^2 detector reacts with a consistent step change in its readings, which can help the detection. Second, since the χ^2 detectors are built on innovation sequences, which need to be calculated based on the model of the system, inevitable model mismatches also affect the performance of the χ^2 detectors. To evaluate this aspect, we consider the simple case that the $\tilde{Q}$ matrix (namely, the covariance of $\breve{w}_k$) is not accurately estimated. Suppose its estimated value is 2.5 times of its real value. The corresponding detector readings are plotted in Figs. 2–6 denoted as $\chi_{A,F}^2$. Note that the readings are much smaller than the nominal values; this can cause critical issues for attack detection, since the χ^2 detector may not be able to capture the attacks when the readings are small [see the definition in (65)]. Hence, χ^2 detectors can be severely affected by model uncertainties. The transfer entropy countermeasures are directly calculated based on the sensor measurements, and thus are less affected by model mismatch.

H. Sensitivity Against System and Measurement Noises

In this section, we analyze the effect of system and measurement noises on the transfer entropy measures. To do this, we focus on the replay attack and consider three noise parameter variations. First, we consider the effect of system noise by

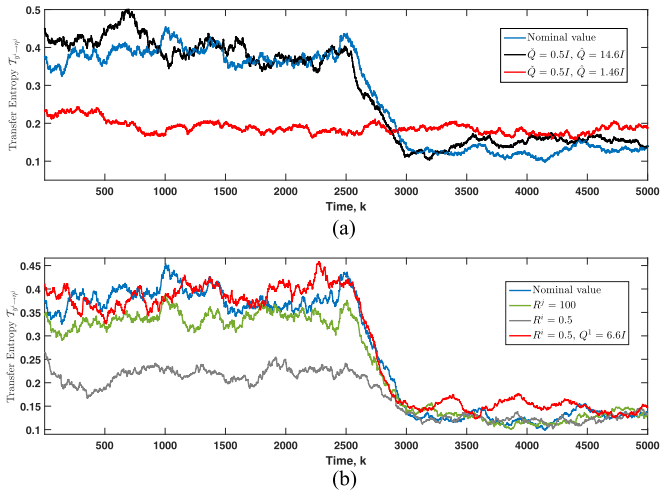


Fig. 7. Sensitivity analysis with respect to noises.

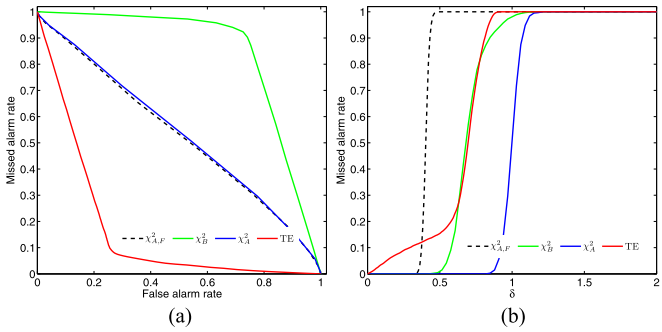


Fig. 8. ROC curves for DoS attack.

increasing the covariance of $\check{w}_k$ by ten times. The corresponding responses are shown in red in Fig. 7(a), where changes in the transfer entropy response can no longer be observed due to the enlarged system noise. One remedy to overcome this issue is to increase the signal-to-noise ratio; to show this, the transfer entropy curve obtained by enlarging $\hat{Q}$ [namely, the covariance of u_k in (63)] ten times of its nominal value is provided in the same figure, where the ability of attack detection is resumed. We further evaluate the effect of measurement noise; to do this, we compare the transfer entropy response curves obtained by separately increasing the measurement noise covariances R^j to 100 and R^i to 0.5, respectively [see Fig. 7(b)]. We observe that the transfer entropy is not sensitive to the increase of R^j, which corresponds to the measurement noise covariance of the "cause" variable, while it is sensitive to the increase of R^i, namely, the measurement noise covariance of the "effect" variable. Again, we note that the capability of attack detection can be recovered by increasing the signal-to-noise ratio, as indicated by the red curve in Fig. 7(b).

I. Receiver Operating Characteristic (ROC) Analysis

In this section, the ROC curves of the considered detectors are numerically evaluated by taking different values of δ (namely, thresholds of the detectors) in (60), (65), and (66) and repeating the simulations 20 times (for each δ) for the attack scenarios considered in Figs. 2–4 and 6. The results are shown in panel (a)

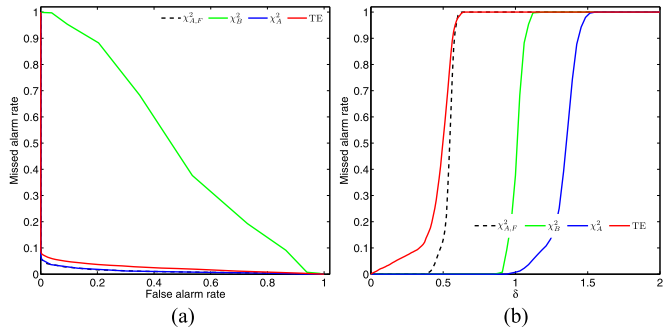


Fig. 9. ROC curves for replay attack.

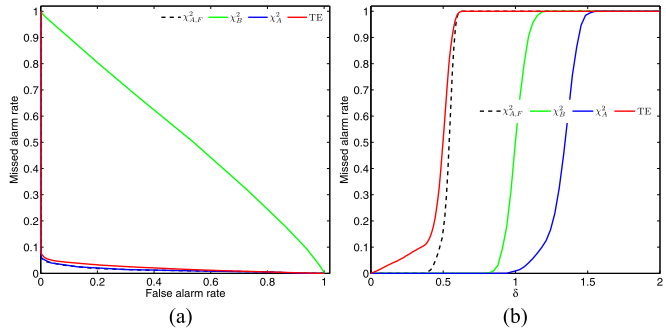


Fig. 10. ROC curves for innovation-based attack (scenario I).

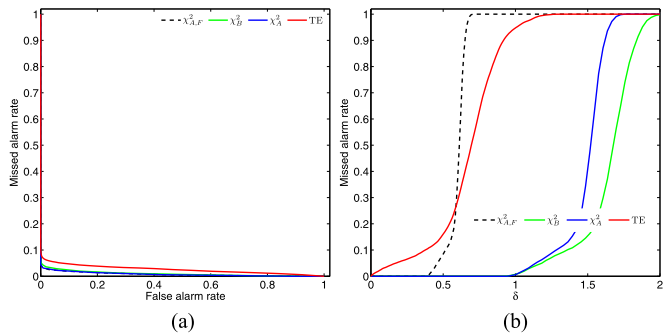


Fig. 11. ROC curves for input data injection attack.

of Figs. 8–11. We observe that in terms of ROC, the transfer entropy based detector behaves obviously better than the χ^2 detectors for the considered DoS attack scenario, while for the rest three scenarios, the ROC curves of transfer entropy based detector, the χ^2_A detector, and the $\chi^2_{A,F}$ detector are close to each other with the performance of χ^2 detectors being slightly better than the transfer entropy based detector. To take a closer look, we also plot the relationship between missed alarm rate and the threshold parameter δ in panel (b) of Figs. 8–11; from these plots, choosing $\delta \in [0.2, 0.4]$ for the transfer entropy based detector seems to be a helpful rule-of-thumb for all these scenarios. For the χ^2 detector, however, the favorable choice of δ varies with the available system model; to see this, note that in Fig. 9, given the exact system model (namely, χ^2_A detector), taking $\delta \in [1.1, 1.3]$ would yield an ROC point close to origin, while given a model with enlarged $\check{Q}$ matrix ($\chi^2_{A,F}$ detector), a favorable choice for δ would be around 0.5. This is consistent with our discussions on robustness in Section VI-G.

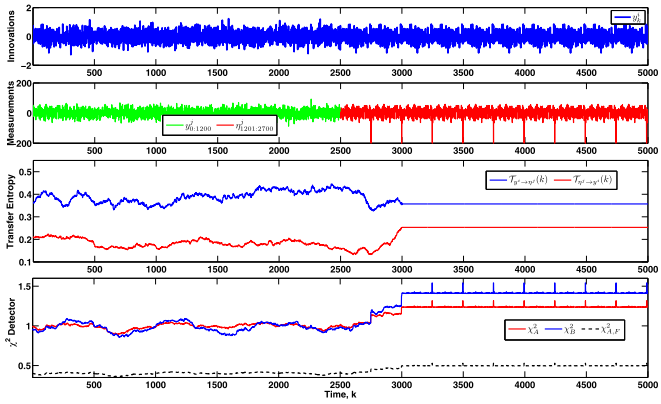


Fig. 12. Replay attacks on both sensors.

We also note that the proposed detector in (60) does not apply to all attack scenarios. For instance, a successful detection of the innovation-based false data attacks (scenario II) in Fig. 5 would be based on the undershoot pattern in transfer entropy readings caused by the attack (and thus the ROC curve for this scenario is not considered); another example is the scenario of replay attacks on both sensor channels (see Section VI-J). Therefore, it seems necessary to achieve attack detection for different scenarios using detector banks, in which each detector focuses on particular attack scenarios, while the transfer entropy seems to be capable of serving as an efficient source for detector bank design.

J. Attacks on Both Sensors

In this section, we briefly discuss how the transfer entropy countermeasures respond to attacks that affect both sensor channels. To do this, we consider a scenario that both sensors y^i and y^j are subject to replay attacks with recording period 250; the results are shown in Fig. 12, where we observe that a special pattern is created by the attack, namely, a constant value in transfer entropy. The reason is due to the data-driven evaluation of transfer entropy, as the periodically replayed data would lead to the same conditional distribution estimate at each time instant after a period of the attacks. The readings of the χ^2 detectors show some special patterns too.

In addition, one related question is how would the performance of the transfer entropy based detector change when attacks on other sensors happen. Since the principle of anomaly detection using transfer entropy lies in the change of causal relationship when anomaly occurs, given a fixed pair of sensors for which transfer entropy is evaluated, the performance of the transfer entropy based detector for the fixed sensor pair only changes when the other sensor signals that have cause-effect relationship with the fixed pair of sensors are attacked.

VII. Conclusion

In this paper, transfer entropy countermeasures for attack detection have been introduced and analyzed, and the effectiveness of these measures toward different attack scenarios has been evaluated utilizing theoretical analysis and numerical experiments. So far, the attempt of utilizing causality countermeasures in attack detection is encouraging. To simplify the analysis, we focused our attention on attack detection based on the information from two sensors with scalar-valued measurements. To be comprehensive, it would be better to monitor the causality relationships based on all (or at least a large number of) available measurement signals. The results developed can theoretically be applied to deal with this case; in fact, systems of larger scale are more appealing, since in a large system the sensor measurements are more likely to exhibit causality relationships. To numerically calculate the transfer entropy values for this case, however, the development of computationally efficient algorithms to evaluate the transfer entropy measures for sensor pairs and sensor groups (i.e., when the transfer entropy from one sensor group to another is considered) is necessary, as the computational burden may be heavily increased when the transfer entropy values for a set of sensor and actuator pairs are calculated for a large-scale system; one possible approach is to formulate the transfer entropy evaluation problem into an adaptive estimation problem, so that the techniques on distributed optimization and adaptive learning can be utilized. These problems point out directions of our future work.

Appendix A
Data-Driven Evaluation of Transfer Entropy

In this appendix, we discuss the numerical calculation of the transfer entropy countermeasure. The algorithm used to compute the transfer entropy measures in our work is not new and was originally proposed in [36].

We focus on the evaluation of the transfer entropy defined for measurement processes at time instant $k+\tau$, that is, $\mathcal{T}_{y^i\to y^j}(k+\tau)$. The most recent historical data sequence of length N, namely, $y^i_{k-N+1+\tau:k+\tau}$ and $y^j_{k-N+1+\tau:k+\tau}$ are used to calculate $\mathcal{T}_{y^i\to y^j}(k+\tau)$. The algorithm is composed of two steps:

Step 1: Evaluation of the conditional distributions. Based on Bayes' rule, the conditional distributions $f(y^i_{k+\tau}|y^i_{k-\mu+1:k}, y^j_{k-l+1:k})$ and $f(y^i_{k+\tau}|y^i_{k-\mu+1:k})$ are expressed in terms of joint distributions, which are obtained using a kernel estimation method [50]. In particular, for q-dimensional multivariate data, the Fukunaga method [50] is utilized to estimate the joint probability density function (pdf). Letting x denote a q-dimensional random process and $X_1, X_2, \ldots, X_L$ be a set of realizations of x, the kernel estimation of the joint pdf is

$$\hat{f}(x) = \frac{(\det \mathbf{S})^{-1/2}}{L(\sqrt{2\pi}\varpi)^q} \sum_{i=1}^{L} \exp\left[-\frac{\varpi^{-2}}{2}(x-X_i)^\top \mathbf{S}^{-1}(x-X_i)\right]$$

where ϖ is chosen as $1.06L^{-1/(4+q)}$, and $\mathbf{S}$ is the covariance matrix of the sampled data $X_{1:L}$. The computation complexity of the kernel estimation method is $O(L^2q^2)$.

Step 2: Calculation of the transfer entropy. From the definition in (5), we have

$$\mathcal{T}_{y^j\to y^i}(k+\tau) = \mathrm{E}\left[\log \frac{f(y^i_{k+\tau}|y^i_{k-\mu+1:k}, y^j_{k-l+1:k})}{f(y^i_{k+\tau}|y^i_{k-\mu+1:k})}\right]$$

and, therefore, the transfer entropy can be numerically evaluated as

$$\mathcal{T}_{y^j \to y^i}(k+\tau) \doteq \frac{1}{N-\tau-\max\{\mu-1,l,2\}} \times \sum_{i=N-\tau-\max\{\mu-1,l,2\}}^{N-\tau} \log \frac{f(y^i_{k+\tau}|y^i_{k-\mu+1:k}, y^j_{k-l+1:k})}{f(y^i_{k+\tau}|y^i_{k-\mu+1:k})}$$

where the conditional distributions are estimated based on the kernel estimation method in Step 1. For this step, approximately N summations are required.

To calculate the transfer entropy, four parameters need to be determined: N, τ, μ, and l. When the measurement processes are at steady state, choosing a larger N will improve the consistency of the transfer entropy readings in the sense that the transfer entropy calculated at different time instants will stay at almost the same level. Also, a larger N will help improve the chance of capturing certain attacks (see the discussions in Section IV-D and Fig. 5). On the other hand, increase of N potentially increases the detection delay, which is often critical in attack detection. In [36], the determination of τ, μ, and l is discussed in detail, as the choice of these parameters are critical in the context of causality analysis. In our work, the transfer entropy is used as a measure of "change of causality" to detect the existence of attacks, so the guidelines in [36] should apply also in our case. We refer the readers to [36, Sec. II.D.4] for detailed discussions.

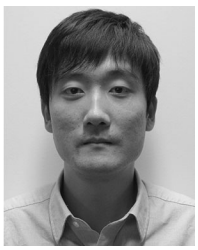
Acknowledgment

D. Shi would like to thank Dr. P. Duan for discussions on the implementation issues of transfer entropy computations. The authors would like to thank the Associate Editor and the anonymous reviewers for their suggestions that have improved the quality of the work.

References

[1] J. Slay and M. Miller, "Lessons learned from the Maroochy water breach," Critical Infrastructure Protection, vol. 253, pp. 73–82, 2007.
[2] J. Farwell and R. Rohozinski, "StuxNet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.
[3] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Distributed internet-based load altering attacks against smart power grids," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 667–674, Dec. 2011.
[4] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Security*, 2009, pp. 21–32.
[5] Y. Mo and B. Sinopoli, "False data injection attacks in cyber physical systems," in *Proc. 1st Workshop Sec. Control Syst.*, 2010.
[6] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proc. 49th IEEE Conf. Decis. Control*, 2010, pp. 5967–5972.
[7] Y. Liu, M. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Security*, vol. 14, 2011, Art. no. 13.
[8] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. 47th Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2009, pp. 911–918.
[9] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396–1407, Jul. 2014.
[10] F. Miao, M. Pajic, and G. Pappas, "Stochastic game approach for replay attack detection," in *Proc. IEEE 52nd Annu. Conf. Decis. Control*, Dec. 2013, pp. 1854–1859.
[11] S. Amin, A. Cardenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Hybrid Systems: Computation and Control* (Lecture Notes in Computer Science), vol. 5469, R. Majumdar and P. Tabuada, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 31–45.
[12] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal DoS attack policy against remote state estimation," in *Proc. IEEE 52nd Annu. Conf. Decis. Control*, Dec. 2013, pp. 5444–5449.
[13] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 3023–3028, Nov. 2015.
[14] G. Befekadu, V. Gupta, and P. Antsaklis, "Risk-sensitive control under Markov modulated denial-of-service (dos) attack strategies," *IEEE Trans. Autom. Control*, vol. 60, no. 12, pp. 3299–3304, Dec. 2015.
[15] C. De Persis and P. Tesi, "Input-to-state stabilizing control under denial-of-service," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 2930–2944, Nov. 2015.
[16] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.
[17] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.
[18] M. Pajic *et al.*, "Robustness of attack-resilient state estimators," in *Proc. Int. Conf. Cyber-Phys. Syst.*, Apr. 2014, pp. 163–174.
[19] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.
[20] D. Shi, R. J. Elliott, and T. Chen, "On finite-state stochastic modeling and secure estimation of cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 65–80, Jan. 2017.
[21] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. Pappas, "The wireless control network: Monitoring for malicious behavior," in *Proc. 49th IEEE Conf. Decis. Control*, Dec. 2010, pp. 5979–5984.
[22] A. Teixeira, S. Amin, H. Sandberg, K. Johansson, and S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *Proc. 49th IEEE Conf. Decis. Control*, Dec. 2010, pp. 5991–5998.
[23] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design," in *Proc. 50th IEEE Conf. Decis. Control Eur. Control Conf.*, Dec. 2011, pp. 2195–2201.
[24] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2011, pp. 337–344.
[25] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling against linear quadratic Gaussian control," in *Proc. Amer. Control Conf.*, Jun. 2014, pp. 3996–4001.
[26] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 1145–1151, Apr. 2015.
[27] Y. Chen, S. Kar, and J. M. F. Moura, "Cyber-physical systems: Dynamic sensor attacks and strong observability," in *Proc. 2015 IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015, pp. 1752–1756.
[28] D. Shi, T. Chen, and M. Darouach, "Event-based state estimation of linear dynamic systems with unknown exogenous inputs," *Automatica*, vol. 69, pp. 275–288, 2016.
[29] H. Sandberg, S. Amin, and K. H. Johansson, Eds., "Cyberphysical security in networked control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 8–12, Feb. 2015.
[30] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 4–13, Mar. 2017.
[31] A. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601–611, 1976.
[32] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000.
[33] P. Verdes, "Assessing causality from multivariate time series," *Phys. Rev. E—Stat. Nonlinear Soft Matter Phys.*, vol. 72, no. 2, 2005, Art. no. 026222.
[34] B. Gourévitch, R. Le Bouquin-Jeannès, and G. Faucon, "Linear and nonlinear causality between signals: Methods, examples and neurophysiological applications," *Biol. Cybern.*, vol. 95, no. 4, pp. 349–369, 2006.
[35] K. Hlaváčková-Schindler, M. Palušš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Phys. Rep.*, vol. 441, no. 1, pp. 1–46, 2007.

[36] P. Duan, F. Yang, T. Chen, and S. Shah, "Direct causality detection via the transfer entropy approach," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 6, pp. 2052–2066, Nov. 2013.
[37] L. Faes, D. Marinazzo, F. Jurysta, and G. Nollo, "Linear and non-linear brain-heart and brain-brain interactions during sleep," *Physiol. Meas.*, vol. 36, no. 4, pp. 683–698, 2015.
[38] W. Yu and F. Yang, "Detection of causality between process variables based on industrial alarm data using transfer entropy," *Entropy*, vol. 17, no. 8, pp. 5868–5887, 2015.
[39] S. Weerakkody, B. Sinopoli, S. Kar, and A. Datta, "Information flow for security in control systems," in *Proc. 2016 IEEE 55th Conf. Decis. Control*, Dec. 2016, pp. 5065–5072.
[40] D. E. Denning and P. J. Denning, "Certification of programs for secure information flow," *Commun. ACM*, vol. 20, no. 7, pp. 504–513, 1977.
[41] P. Duan, F. Yang, S. Shah, and T. Chen, "Transfer zero-entropy and its application for capturing cause and effect relationship between variables," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 3, pp. 855–867, May 2015.
[42] V. Marques, C. Munaro, and S. Shah, "Detection of causal relationships based on residual analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1525–1534, Oct. 2015.
[43] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.
[44] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[45] G. De Nicolao, "On the time-varying Riccati difference equation of optimal filtering," *SIAM J. Control Optim.*, vol. 30, no. 6, pp. 1251–1269, 1992.
[46] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Berlin, Germany: Springer-Verlag, 2008.
[47] N. L. Ricker, "Tennessee Eastman Challenge Archive," 2015. [Online]. Available: http://depts.washington.edu/control/LARRY/TE/download.html. Accessed on: Jan. 23, 2015.
[48] N. L. Ricker, "Model predictive control of a continuous, nonlinear, two-phase reactor," *J. Process Control*, vol. 3, no. 2, pp. 109–123, 1993.
[49] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst.*, vol. 35, no. 1, pp. 93–109, Feb. 2015.
[50] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York, NY, USA: Chapman & Hall, 1986.



Dawei Shi received the B.Eng. degree in electrical engineering and its automation from the Beijing Institute of Technology, Beijing, China, in 2008, the Ph.D. degree in control systems from the University of Alberta, Edmonton, AB, Canada, in 2014.

In December 2014, he was appointed as an Associate Professor at the School of Automation, Beijing Institute of Technology. In February 2017, he joined Harvard John A. Paulson School of Engineering and Applied Sciences, USA, as a Postdoctoral Fellow in bioengineering. His research interests include event-based control and estimation, robust model predictive control and tuning, and wireless sensor networks.

Dr. Shi is a reviewer for a number of international journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, and *Systems & Control Letters*. In 2009, he received the Best Student Paper Award in IEEE International Conference on Automation and Logistics.

Ziyang Guo was born in Henan, China, in 1992. She received the B.Eng. (Hons.) degree from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2014. She is currently working toward the Ph.D. degree in electronic and computer engineering at the Hong Kong University of Science and Technology, Hong Kong.

Her research interests include cyber-physical system security, state estimation, and wireless sensor network.

Karl Henrik Johansson (F'13) received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University, Lund, Swedenin, in 1992 and 1997, respectively.

He is currently the Director of the Stockholm Strategic Research Area ICT The Next Generation and a Professor at the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests include networked control systems, cyber-physical systems, and applications in transportation, energy, and automation.

Prof. Johansson is a member of the IEEE Control Systems Society Board of Governors and the European Control Association Council. He has received several best paper awards and other distinctions, including a ten-year Wallenberg Scholar Grant, a Senior Researcher Position with the Swedish Research Council, the Future Research Leader Award from the Swedish Foundation for Strategic Research, and the Triennial Young Author Prize from IFAC. He is an IEEE Distinguished Lecturer.

Ling Shi received the B.S. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Hong Kong, in 2002, and the Ph.D. degree in control and dynamical systems from California Institute of Technology, Pasadena, CA, USA, in 2008.

He is currently an Associate Professor in the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include cyber-physical systems security, networked control systems, sensor scheduling, and event-based state estimation.

Dr. Shi has been serving as a Subject Editor of the *International Journal of Robust and Nonlinear Control* from March 2015, and an Associate Editor of the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS from July 2016, and an Associate Editor of the *IEEE Control Systems Letters* from Feb 2017. He was an Associate Editor for a special issue on Secure Control of Cyber Physical Systems in the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS in 2015–2017.