Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-21 (2015) 746–753

Sequential monitoring of SCADA systems against cyber/physical attacks [⋆]

Van Long DO [*] **Lionel FILLATRE** [**] **Igor NIKIFOROV** [*]

[*] *University of Technology of Troyes, CNRS, ICD/LM2S, UMR 6281, 10004 Troyes Cedex, France (e-mail: van_long.do@utt.fr and igor.nikiforov@utt.fr)*

[**] *University Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France (e-mail: lionel.fillatre@i3s.unice.fr)*

Abstract: The sequential monitoring of SCADA systems against cyber/physical attacks is considered in this paper. The SCADA systems are described by the discrete-time state space models in the presence of random noises. The cyber/physical attacks are modeled as additive signals of short duration impacted both the state evolution and the sensor measurement equations. The detection of attacks is formulated as the problem of sequential transient change detection in stochastic-dynamical systems. The steady-state Kalman filter and the fixed-size parity space are utilized for generating the sequence of residuals. The unified statistical model is developed to describe the residual generation by both methods. Based on this statistical model, the Variable Threshold Window Limited CUmulative SUM (VTWL CUSUM) algorithm is designed to detect the transient changes. Taking into consideration the detection criterion, which aims at minimizing the worst-case probability of missed detection subject to a given value on the worst-case probability of false alarm, the thresholds are tuned for optimizing the VTWL CUSUM algorithm. It is shown that the optimal choice of thresholds leads to the simple Finite Moving Average (FMA) detection rule. The proposed algorithms are applied to detect the covert attack on a simple SCADA water distribution network.

Keywords: SCADA systems, cyber attacks, transient change detection, Window Limited CUSUM, Finite Moving Average.

1. INTRODUCTION AND RELATED WORKS

Supervisory Control and Data Acquisition (SCADA) systems have been used to control and monitor a large number of critical infrastructures such as electric power grids, transportation systems, communication networks, gas pipelines and water distribution networks. Due to their geographically dispersed characteristics, the SCADA systems become more and more susceptible to cyber/physical attacks, not only on the physical infrastructures but also on the communication network and the control center. Some examples of recent cyber incidents include the Maroochy water breach, see Slay and Miller (2007), the U.S. electricity grid penetration, see Gorman (2009), or the Stuxnet virus, see Brunner et al. (2010).

Two approaches have been considered to study the security of SCADA systems against cyber/physical attacks: information security approach and secure control theory one, in Kwon et al. (2013). The information security methods concentrate mainly on the authentication, access control or message integrity. On the other hand, the secure control theory approach exploits the analytical redundancy of the SCADA systems to detect the attacks.

In recent years, a great deal of effort has been paid for studying the vulnerabilities of networked control systems. More precisely, these weak points are exploited to design stealthy attacks which can partially or completely bypass traditional anomaly detectors. These undetectable attacks can be classified into the replay attack, see Mo and Sinopoli (2009), the false data injection attack, see Mo and Sinopoli (2010), the zero-dynamics attack, see Teixeira et al. (2012), or the covert attack, see Smith (2011).

The problem of attack detection and isolation has been also considered. The detection of cyber attacks on process control systems has been formulated in Cárdenas et al. (2011) as the fault diagnosis problem. The non-parametric CUSUM algorithm is utilized to detect the attacks. Moreover, the security of water irrigation networks against cyber attacks has been considered in Amin et al. (2012a,b), where a bank of unknown input observers is designed to detect and isolate the attacks. Finally, a comprehensive framework has been introduced in Pasqualetti et al. (2013) to study the attack detection and identification problem. The deterministic state space model is utilized to describe SCADA systems. The cyber attacks are modeled as additive signals to both state evolution and sensor measurement equations. Centralized and distributed algorithms are proposed to detect and identify the attacks.

It has been shown that the attack detection and isolation is closely related to the fault detection and isolation (FDI) problem. This problem is generally solved by using the analytical redundancy approach, which consists of two

[⋆] This work received financial support from the French National Research Agency (ANR).

2405-8963 © 2015, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2015.09.616

steps: residual generation and residual evaluation. The residuals are first generated by using traditional methods (i.e., the Kalman filter or the parity space approaches) and they are then evaluated by utilizing change detection techniques (see, among others, Basseville and Nikiforov (1993); Chen and Patton (1999)).

2. CONTRIBUTION AND ORGANISATION

Usually, a cyber/physical attack comes in unexpected time and unexpected situation. The attacker may prefer to perform his malevolent action within a short period due to the resource limit. Hence, the changes produced by the malicious attack in the SCADA system arrive suddenly and their duration is usually short. From the other hand, for safety-critical applications, it is required to detect the attacks with the detection delay upper bounded by a certain prescribed value. For these reasons, it is more adequate to formulate the detection of attacks as the sequential detection of a short signal (also called a transient change) in a stochastic-dynamical system.

The change detection problem, including the traditional abrupt change detection and the transient change detection, consists in calculating the stopping time T at which the changes produced at change-point k_0 are detected. In the traditional abrupt change detection, the post-change period is assumed to be infinitely long (see for details the recent survey of Lai (2001) and also recent books Poor and Hadjiliadis (2009); Tartakovsky et al. (2014)). Typically, the criterion of optimality aims at minimizing the average detection delay for a given average run length to a false alarm. Unfortunately, such an optimality criterion is not adequate for the transient change detection problem which deals with the changes of short duration. The delayed (or latent) detection of changes after their disappearance is considered as missed.

The mathematical formulation of the transient change detection problem for the independent observation model has been considered in Guépié et al. (2012a,b); Guépié (2013). The optimality criterion involves the minimization of the worst-case probability of missed detection for a given value on the worst-case probability of false alarm within any time window of chosen length. A sub-optimal algorithm with respect to (w.r.t.) this criterion has been introduced for the case of independent normal observations.

This paper, pursuing the work started in Guépié et al. (2012a,b); Guépié (2013), is dedicated to the detection of cyber/physical attacks on SCADA systems described by the discrete-time state space model in the presence of unknown system states and random noises.

The contribution of this paper is threefold. First, we introduce a unified statistical model of the residuals generated by either the steady-state Kalman filter approach or the fixed-size parity space approach applicable to the system with transient changes. Second, we design a sub-optimal algorithm (for both the Kalman filter-based approach and the parity space-based approach) to detect transient changes in discrete-time state space models. Finally, we apply the proposed algorithms to the detection of cyber/physical attacks on a simple SCADA system.

The originality of this paper w.r.t. Guépié et al. (2012a,b); Guépié (2013) is a more sophisticated model of observations. The design and the study of the transient change detectors in these previous papers depend heavily on the concept of associated random variables, see details and results in Lehmann (1966); Esary et al. (1967). Unfortunately, it is impossible to use again this concept due to the fact that the observations are generated by the state space model in the presence of unknown system states (nuisances) and random noises. For this reason, another method of the design and the study of the detector is used now.

The structure of the paper is organized as following. The problem statement and the unified statistical model are given in section 3 and section 4, respectively. Based on the unified statistical model, the VTWL CUSUM and FMA tests are designed in section 5. The application of the proposed algorithms to the detection of attack on the water network is presented in section 6. Some concluding remarks are drawn in section 7.

3. PROBLEM STATEMENT

In this paper, we utilize the following discrete-time state space model to describe SCADA systems and cyber/physical attacks:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k + Fd_k + B_a a_k + w_k \\ y_k &= Cx_k + Du_k + Gd_k + D_a a_k + v_k \end{cases} ; \ x_0 = \overline{x}_0, \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the vector of system states with unknown initial condition $\overline{x}_0$, $u_k \in \mathbb{R}^m$ is the vector of control signals, $d_k \in \mathbb{R}^q$ is the vector of disturbances, $y_k \in \mathbb{R}^p$ is the vector of measurements, $a_k \in \mathbb{R}^s$ is the vector of attack signals, $w_k \in \mathbb{R}^n$ is the vector of process noises and $v_k \in \mathbb{R}^p$ is the vector of sensor noises; the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $F \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $G \in \mathbb{R}^{p\times q}$, $B_a \in \mathbb{R}^{n\times s}$ and $D_a \in \mathbb{R}^{p\times s}$ are assumed to be known.

For simplicity, let us suppose that the control signals u_k and the disturbances d_k are completely known. The control signals are the outputs of controllers so they are generally known to system operators. For safety-critical applications such as electric power grids, gas pipelines or water distribution networks, the disturbances d_k correspond to customer demands which are often estimated by special-designed software with an acceptable level of errors. Generally, these estimation errors are unbiased, so they can be integrated into the process noises w_k and/or the sensor noises v_k.

The process noises $w_k \sim \mathcal{N}(0, Q)$ and the sensor noises $v_k \sim \mathcal{N}(0, R)$ are assumed to be independent identically distributed (i.i.d.) zero-mean Gaussian random vectors, where the covariance matrices $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{p\times p}$ are exactly known and R is positive definite.

Let us suppose that the attacker performs his malicious action during a short period $\tau_a = [k_0, k_0 + L - 1]$, where k_0 is the unknown attack instant and L is the known change period. The attack vector a_k is then described as follows:

$$a_k = \begin{cases} 0 & \text{if } k < k_0 \\ \theta_{k-k_0+1} & \text{if } k_0 \leq k < k_0 + L \\ 0 & \text{if } k \geq k_0 + L \end{cases}, \tag{2}$$

where $\theta_1, \theta_2, \cdots, \theta_L \in \mathbb{R}^s$ are the attack profiles, which are assumed to be completely known.

Remark 1. It has been shown in Mo and Sinopoli (2009); Mo et al. (2010); Smith (2011); Teixeira et al. (2012) that the attack vector a_k can be designed to disrupt the systems while bypassing traditional anomaly detectors. The security analysis process is required to render such attacks detectable. For example, more secure sensors may be equipped in vulnerable points of the systems, making these stealthy attacks detectable (see, for example, Mo and Sinopoli (2010), Teixeira et al. (2012) or Kwon et al. (2013)). For these reasons, we consider the detection of only detectable attacks.

Remark 2. In this paper, the vectors $\theta_1, \theta_2, \cdots, \theta_L$ are assumed to be exactly known. These vectors are defined by the dynamics of equipment and by the type of attack. For some water distribution systems, the dynamics of equipment (pumps, compressors, etc.) is *a priori* known or can be pre-calculated. It also can be assumed that each attack scenario leads to a particular attack signature.

Let us consider a simple SCADA water distribution network presented in Fig. 1 in section 6, where a pump is utilized for supplying water to a reservoir. It is assumed that the water distribution system is equipped with a constant speed pump. Such a pump operates in two modes: "on" and "off". It is assumed that the attacker performs his malicious attack for switching the pump "off" while it is functioning (see Zetter (2011) for a real attack on a water utility). In this case, the vectors $\theta_1, \theta_2, \cdots, \theta_L$ are completely specified.

In this paper, we use the transient detection criterion which was first introduced in Guépié et al. (2012b); Guépié (2013). We wish to minimize the following worst-case probability of missed detection

$$\inf_{T \in C_\alpha} \left\{ \overline{\mathbb{P}}_{md}(T; L) = \sup_{k_0 \geq L} \mathbb{P}_{k_0}(T - k_0 + 1 > L | T \geq k_0) \right\}, \tag{3}$$

among all stopping times $T \in C_\alpha$ satisfying

$$C_\alpha = \left\{ T : \overline{\mathbb{P}}_{fa}(T; m_\alpha) = \sup_{l \geq L} \mathbb{P}_0 \{ l \leq T < l + m_\alpha \} \leq \alpha \right\}, \tag{4}$$

where $\overline{\mathbb{P}}_{md}$ denotes the worst-case probability of missed detection and $\overline{\mathbb{P}}_{fa}$ stands for the worst-case probability of false alarm within any time window of length m_α.

4. UNIFIED STATISTICAL MODELS

In this section, the unified statistical model of the residuals generated by both the steady-state Kalman filter approach and the fixed-size parity space approach is developed.

4.1 Steady-state Kalman filter

Let us suppose that the steady-state Kalman filter is used for generating the sequence of residuals. The steady-state Kalman gain is calculated as

$$K = PC^T \left(CPC^T + R \right)^{-1}, \tag{5}$$

where P denotes the steady-state covariance matrix of the state estimation error, which can be found by solving the following discrete-time algebraic Riccati equation:

$$P = APA^T - APC^T \left(CPC^T + R \right)^{-1} CPA^T + Q. \tag{6}$$

The operation of the steady-state Kalman filter is then described by the following equation:

$$\begin{cases} \hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + Bu_k + Fd_k + AK\left(y_k - \hat{y}_{k|k-1} \right) \\ \hat{y}_{k|k-1} &= C\hat{x}_{k|k-1} + Du_k + Gd_k \end{cases}, \tag{7}$$

where $\hat{x}_{k|k-1} \in \mathbb{R}^n$ is state estimate with initial condition $\hat{x}_{0|-1} = \overline{x}_0$, and $\hat{y}_{k|k-1} \in \mathbb{R}^p$ is the output estimate. Let $r_k = y_k - \hat{y}_{k|k-1} \in \mathbb{R}^p$ be the residuals (or the innovations). The innovations $\{r_k\}_{k \geq 0}$ have been shown to be independent Gaussian random variables with the covariance matrix $CPC^T + R$. Let $\varrho_0, \varrho_1, \cdots, \varrho_k \in \mathbb{R}^p$ be i.i.d. zero-mean Gaussian random vectors satisfying $\varrho_k \sim \mathcal{N}\left(0, CPC^T + R\right)$.

The model of the innovations is then expressed by

$$r_k = \begin{cases} \varrho_k & \text{if } k < k_0 \\ \psi_{k-k_0+1} + \varrho_k & \text{if } k_0 \leq k < k_0 + L \\ \tilde{\psi}_k + \varrho_k & \text{if } k \geq k_0 + L \end{cases}, \tag{8}$$

where the transient change profiles (or innovation signatures) $\psi_1, \psi_2, \cdots, \psi_L \in \mathbb{R}^p$ are calculated by

$$\begin{cases} \epsilon_{k+1} &= (A - AKC)\,\epsilon_k + (B_a - AKD_a)\,\theta_k \\ \psi_k &= C\epsilon_k + D_a\theta_k \end{cases}; \ \epsilon_1 = 0, \tag{9}$$

and the post-change profiles $\tilde{\psi}_k$ (i.e., for $k \geq k_0 + L$) are of no interest. The interested readers can find additional details on the calculation of innovation signatures in Basseville and Nikiforov (1993); Lai and Shan (1999).

Let $r_{k-L+1}^k = \left[r_{k-L+1}^T, \cdots, r_k^T \right]^T \in \mathbb{R}^{Lp}$ be the vector of residuals, $\varrho_{k-L+1}^k = \left[\varrho_{k-L+1}^T, \cdots, \varrho_k^T \right] \in \mathbb{R}^{Lp}$ be the vector of random noises, and $\psi_{k-L+1}^k(k_0) \in \mathbb{R}^{Lp}$ be the vector of transient signals. The vector $\psi_{k-L+1}^k(k_0)$ depends on the relative position of the change-point k_0 within the window $[k - L + 1, k]$ by the following relation:

$$\psi_{k-L+1}^k(k_0) = \begin{cases} [0] & \text{if } k < k_0 \\ \begin{bmatrix} 0 \\ \psi_1 \\ \vdots \\ \psi_{k-k_0+1} \end{bmatrix} & \text{if } k_0 \leq k < k_0 + L \\ \tilde{\psi}_{k-L+1}^k(k_0) & \text{if } k \geq L \end{cases}, \tag{10}$$

where the vector of post-change profiles $\tilde{\psi}_{k-L+1}^k(k_0) \in \mathbb{R}^{Lp}$ are of no interest. Putting together with (8)–(10), the statistical model of the residual vector r_{k-L+1}^k generated by the steady-state Kalman filter is described as

$$r_{k-L+1}^k = \psi_{k-L+1}^k(k_0) + \varrho_{k-L+1}^k, \tag{11}$$

where the random noises $\varrho_{k-L+1}^k \sim \mathcal{N}(0, \Sigma_\varrho)$, where $\Sigma_\varrho = \operatorname{diag}\left(CPC^T + R\right) \in \mathbb{R}^{Lp \times Lp}$ is the diagonal matrix formed of L blocks of $CPC^T + R$.

4.2 Fixed-size parity space

By eliminating the control signals u_k and the disturbances d_k (since they are assumed to be known), we obtain the following simplified observation model:

$$\underbrace{\begin{bmatrix} z_{k-L+1}\\ z_{k-L+2}\\ \vdots \\ z_k\end{bmatrix}}_{z^k_{k-L+1}} = \underbrace{\begin{bmatrix} C\\ CA\\ \vdots\\ CA^{L-1}\end{bmatrix}}_{\mathcal{C}} x_{k-L+1} + \underbrace{\begin{bmatrix} v_{k-L+1}\\ v_{k-L+2}\\ \vdots\\ v_k\end{bmatrix}}_{v^k_{k-L+1}} + \underbrace{\begin{bmatrix} D_a & 0 & \cdots & 0\\ CB_a & D_a & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ CA^{L-2}B_a & CA^{L-3}B_a & \cdots & D_a\end{bmatrix}}_{\mathcal{M}} \underbrace{\begin{bmatrix} a_{k-L+1}\\ a_{k-L+2}\\ \vdots\\ a_k\end{bmatrix}}_{\theta^k_{k-L+1}(k_0)} + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0\\ C & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ CA^{L-2} & CA^{L-3} & \cdots & 0\end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} w_{k-L+1}\\ w_{k-L+2}\\ \vdots\\ w_k\end{bmatrix}}_{w^k_{k-L+1}}, \tag{12}$$

where $z^k_{k-L+1} \in \mathbb{R}^{Lp}$ is the vector of simplified observations, $w^k_{k-L+1} \in \mathbb{R}^{Ln}$ is the vector of process noises, $v^k_{k-L+1} \in \mathbb{R}^{Lp}$ is the vector of sensor noises, $\theta^k_{k-L+1}(k_0) \in \mathbb{R}^{Ls}$ is the vector of transient signals; the matrices $\mathcal{C} \in \mathbb{R}^{Lp\times n}$, $\mathcal{H} \in \mathbb{R}^{Lp\times Ln}$ and $\mathcal{M} \in \mathbb{R}^{Lp\times Ls}$. The process noises $w^k_{k-L+1} \sim \mathcal{N}(0, \mathcal{Q})$ and the sensor noises $v^k_{k-L+1} \sim \mathcal{N}(0, \mathcal{R})$, where $\mathcal{Q} = \operatorname{diag}(Q) \in \mathbb{R}^{Ln\times Ln}$ and $\mathcal{R} = \operatorname{diag}(R) \in \mathbb{R}^{Lp\times Lp}$. The vector of attack profiles $\theta^k_{k-L+1}(k_0)$, depending on the relative position of the change-point k_0 within the window $[k-L+1, k]$, is described as

$$\theta^k_{k-L+1}(k_0) = \begin{cases} [0] & \text{if } k < k_0\\ \begin{bmatrix} [0]\\ \theta_1\\ \vdots\\ \theta_{k-k_0+1}\end{bmatrix} & \text{if } k_0 \le k < k_0 + L\\ \left[\tilde{\theta}^k_{k-L+1}(k_0)\right] & \text{if } k \ge k_0 + L\end{cases}, \tag{13}$$

where the post-change profiles $\tilde{\theta}^k_{k-L+1}(k_0) \in \mathbb{R}^{Ls}$ are of no interest. The observation model is then described as

$$z^k_{k-L+1} = \mathcal{C}x_{k-L+1} + \mathcal{M}\theta^k_{k-L+1}(k_0) + \mathcal{H}w^k_{k-L+1} + v^k_{k-L+1}. \tag{14}$$

The rejection of the nuisance parameter has been discussed in Fouladirad and Nikiforov (2005) by applying the invariant theory. The main idea is as follows. The observation vector z^k_{k-L+1} is projected onto the orthogonal complement space $R(\mathcal{C})^{\perp}$ of the column space $R(\mathcal{C})$ of matrix $\mathcal{C}$, which is assumed to be full column rank. The residual vector is calculated as $r^k_{k-L+1} = \mathcal{W}z^k_{k-L+1}$, where the columns of the matrix $\mathcal{W}^T \in \mathbb{R}^{Lp\times(Lp-n)}$ are composed of the eigenvectors of the projection matrix $\mathcal{P}_{\mathcal{C}} = \mathcal{I} - \mathcal{C}\left(\mathcal{C}^T\mathcal{C}\right)^{-1}\mathcal{C}^T$ corresponding to eigenvalue 1. The matrix $\mathcal{W}$ satisfies the following conditions: $\mathcal{W}\mathcal{C} = 0$, $\mathcal{W}^T\mathcal{W} = \mathcal{P}_{\mathcal{C}}$ and $\mathcal{W}\mathcal{W}^T = \mathcal{I}$. It is clear that the residual vector r^k_{k-L+1} is independent from the nuisance vector x_{k-L+1}.

Let $\varphi^k_{k-L+1}(k_0) = \mathcal{W}\mathcal{M}\theta^k_{k-L+1}(k_0)$ be the vector of transient signals and $\varsigma^k_{k-L+1} = \mathcal{W}\left(\mathcal{H}w^k_{k-L+1} + v^k_{k-L+1}\right)$ be the vector of random noises. The statistical model of the residual vector r^k_{k-L+1} generated by the fixed-size parity space approach is described as

$$r^k_{k-L+1} = \varphi^k_{k-L+1}(k_0) + \varsigma^k_{k-L+1}, \tag{15}$$

where the random noises $\varsigma^k_{k-L+1} \sim \mathcal{N}(0, \Sigma_\varsigma)$, where the covariance matrix $\Sigma_\varsigma = \mathcal{W}\left(\mathcal{H}\mathcal{Q}\mathcal{H}^T + \mathcal{R}\right)\mathcal{W}^T$.

Remark 3. Taking into account the information about the noise covariances, Gustafsson (2002) suggested to replace the unknown system state x_{k-L+1} by its least-square estimate $\hat{x}_{k-L+1}$. The residual generator by the least-square estimation method is described as

$$r^k_{k-L+1} = \mathcal{W}_{LS}z^k_{k-L+1}, \tag{16}$$

where the rows of the rejection matrix $\mathcal{W}_{LS} \in \mathbb{R}^{(Lp-n)\times Lp}$ form the basis for the row space of the projection matrix

$$\mathcal{P}_{LS} = \mathcal{I} - \mathcal{C}\left(\mathcal{C}^T\mathcal{S}^{-1}\mathcal{C}\right)^{-1}\mathcal{C}^T\mathcal{S}^{-1},$$

where $\mathcal{S} = \mathcal{H}\mathcal{Q}\mathcal{H}^T + \mathcal{R}$ and the projection matrix $\mathcal{P}_{LS}$ is of rank $Lp - n$, idempotent but not necessarily symmetric. It can be seen from (16) that the least-square estimation method offers the residuals with the same structure as the fixed-size parity space approach does. The main difference between the two methods is the calculation of the rejection matrices $\mathcal{W}$ and $\mathcal{W}_{LS}$. The rejection matrix $\mathcal{W}$ calculated by the fixed-size parity space approach is independent of the noise covariances while the computation of $\mathcal{W}_{LS}$ requires the information about noise covariance matrices R and Q. It has been discussed in Gustafsson (2002) that the least-square estimation method offers the residuals with minimum covariance. However, the residuals with minimum covariance do not guarantee the statistical performance of a detection procedure since small noise covariance is often associated with small value of the change magnitude (due to the projection).

An analogous problem of optimal fault detection has been addressed within the statistical framework in Fouladirad et al. (2008). A linear model with nuisance parameters and a general covariance matrix (not necessarily diagonal) has been considered in the context of the unknown but non-random nuisance parameters. Two different invariant tests have been designed in such a case. The first invariant statistics was based on the knowledge of the observation matrix and the noise covariance matrix and the second one was based on the observation matrix only. It was shown that these methods are equivalent. The comparison between residual-generation methods will be performed by means of the Kullback-Leibler distance in the next subsection and some results of numerical simulation are available in section 6.

4.3 Unified statistical model

It follows from (11) and (15) that the utilization of the steady-state Kalman filter and the fixed-size parity space leads to the following unified statistical model:

$$r^k_{k-L+1} = \phi^k_{k-L+1}(k_0) + \xi^k_{k-L+1}, \tag{17}$$

where r^k_{k-L+1} is the vector of residuals, $\phi^k_{k-L+1}(k_0)$ is the vector of transient signals and $\xi^k_{k-L+1} \sim \mathcal{N}(0, \Sigma)$ is the vector of random noises.

Remark 4. If the steady-state Kalman filter approach is used, the transient profiles $\phi_{k-L+1}^{k}(k_0) = \psi_{k-L+1}^{k}(k_0)$ and the random noises $\xi_{k-L+1}^{k} = \varrho_{k-L+1}^{k}$ (i.e., $\Sigma = \Sigma_\varrho$). On the other hand, if the fixed-size parity space is used, the transient profiles $\phi_{k-L+1}^{k}(k_0) = \varphi_{k-L+1}^{k}(k_0)$ and the random noises $\xi_{k-L+1}^{k} = \varsigma_{k-L+1}^{k}$ (i.e., $\Sigma = \Sigma_\varsigma$).

Let $\mathcal{P}_{k_0}$ (resp. $\mathcal{P}_0 \triangleq \mathcal{P}_\infty$) be the probability measure when the residuals $r_1^L, r_2^{L+1}, \cdots, r_{k-L+1}^{k}, \cdots$ follow the statistical model (17), $\mathbb{E}_{k_0}$ (resp. $\mathbb{E}_0 \triangleq \mathbb{E}_\infty$) denote the corresponding mathematical expectations, and p_{k_0} (resp. p_0) stand for the probability density function.

In this paper, the Kullback-Leibler (K-L) distance is utilized for comparing between residual-generation methods. It is assumed, for the sake of simplicity, that $k = L$ and $k_0 = 1$. Let us stack the vectors $\psi_1, \psi_2, \ldots, \psi_L$ (resp. $\varphi_1, \varphi_2, \ldots, \varphi_L$) into the concatenated vector $\psi_1^L(1)$ (resp. $\varphi_1^L(1)$). For the Gaussian noises, the K-L distances are calculated as (see (Basseville and Nikiforov, 1993, p. 122))

$$\rho_{KF} = \frac{1}{2}\left[\psi_1^L(1)\right]^T \left[\Sigma_\varrho^{-1}\right]\left[\psi_1^L(1)\right], \qquad (18)$$

$$\rho_{PS} = \frac{1}{2}\left[\varphi_1^L(1)\right]^T \left[\Sigma_\varsigma^{-1}\right]\left[\varphi_1^L(1)\right], \qquad (19)$$

where ρ_{KF} and ρ_{PS} are the K-L distances generated by the steady-state Kalman filter and the fixed-size parity space approaches, respectively.

Remark 5. It is well-known that the residuals with higher K-L distance give better statistical performances than the residuals with lower K-L distance (see Basseville and Nikiforov (1993)). The numerical comparison between the steady-state Kalman filter approach and the fixed-size parity space approach will be given by simulation result in section 6.

5. DETECTION ALGORITHMS

Based on unified statistical model (17), we design in this section the VTWL CUSUM algorithm to detect the transient changes. It will be shown that the optimal choice of thresholds leads to the simple FMA detection rule.

5.1 VTWL CUSUM algorithm

For each time instant $k \geq L$, the VTWL CUSUM algorithm utilizes the last L measurements $y_{k-L+1}, \cdots, y_k$ for decision making. The stopping time T of the VTWL CUSUM algorithm is calculated as follows:

$$T = \inf\left\{k \geq L : \max_{k-L+1 \leq i \leq k}\left(S_i^k - h_{k-i+1}\right) \geq 0\right\}, \qquad (20)$$

where $h_1, h_2, \cdots, h_L$ are chosen thresholds and S_i^k, for each index i within the time window $[k - L + 1, k]$, is the log-likelihood ratio (LLR) between the probability measure $\mathcal{P}_i$ and the pre-change probability measure $\mathcal{P}_0$. The LLR S_i^k is calculated in the Gaussian case as

$$S_i^k = \left[\phi_{k-L+1}^{k}(i)\right]^T \left[\Sigma^{-1}\right]\left[r_{k-L+1}^{k} - \frac{1}{2}\phi_{k-L+1}^{k}(i)\right]. \qquad (21)$$

The operation of the VTWL CUSUM algorithm is as follows. The LLR S_i^k, for each time index i from $k-L+1$ to k, is first calculated by (21), where the vector of residuals r_{k-L+1}^{k}, the vector of transient signals $\phi_{k-L+1}^{k}(i)$, and the covariance matrix Σ of the random noises ξ_{k-L+1}^{k} can be calculated by either the steady-state Kalman filter approach or the fixed-size parity space approach. Then, the LLR S_i^k is compared to each threshold h_{k-i+1} and the alarm time T is raised if one of the LLRs is greater than or equal to its corresponding threshold. Especially, the thresholds $h_1, h_2, \cdots, h_L$ are considered as the tuning parameters for optimizing the VTWL CUSUM algorithm.

5.2 FMA detection rule

The optimization of the VTWL CUSUM algorithm has been considered in Guépié et al. (2012b); Guépié (2013) for the independent normal observations. The main idea is as follows. Firstly, the upper bound $\tilde{\mathbb{P}}_{md}$ on the worst-case probability of missed detection $\overline{\mathbb{P}}_{md}$ and the upper bound $\tilde{\mathbb{P}}_{fa}$ on the worst-case probability of false alarm $\overline{\mathbb{P}}_{fa}$ are calculated instead of their exact values. Secondly, the thresholds $h_1, h_2, \cdots, h_L$ are tuned for minimizing the upper bound $\tilde{\mathbb{P}}_{md}$ for a given value of the upper bound $\tilde{\mathbb{P}}_{fa}$. It has been shown that the optimal choice of thresholds corresponds to $h_1, \cdots, h_{L-1} \to +\infty$ and h_L is chosen for assuring an acceptable level of false alarms.

Fortunately, similar results are still valid in a more general case of stochastic-dynamical systems. The main outcome is given as the following

Theorem 1. Consider the VTWL CUSUM test defined in (20)–(21). Then,

1) The optimal choice of the thresholds $h_1, h_2, \cdots, h_L$ leads to the following optimization problem:

$$\begin{cases} \inf\limits_{h_1, h_2, \cdots, h_L} & \tilde{\mathbb{P}}_{md}(h_L) \\ \text{subject to} & \overline{\mathbb{P}}_{fa}(T; m_\alpha; h_1, h_2, \cdots, h_L) \leq \alpha \end{cases}, \qquad (22)$$

where α is the acceptable level for the worst-case probability of false alarm within any time window of length m_α. The optimization problem (22) has the unique solution $(h_1^*, h_2^*, \cdots, h_L^*)$ for a given $\alpha \in (0, 1)$, where $h_1^*, h_2^*, \cdots, h_{L-1}^* \to +\infty$ and h_L^* is calculated from the following equation:

$$\mathbb{P}_0\left(\bigcap_{k=L}^{L+m_\alpha-1}\left\{S_{k-L+1}^{k} < h_L^*\right\}\right) = 1 - \alpha. \qquad (23)$$

2) The optimized VTWL CUSUM test is equivalent to the following FMA test:

$$\widetilde{T}\left(\tilde{h}_L\right) = \inf\left\{k \geq L : \left[\phi_1^L(1)\right]^T \Sigma^{-1} r_{k-L+1}^{k} \geq \tilde{h}_L\right\}, \qquad (24)$$

where the threshold $\tilde{h}_L$ of the FMA test (24) is related to the optimal threshold h_L^* of the optimized VTWL CUSUM test (20) by

$$\tilde{h}_L = h_L^* + \frac{1}{2}\left[\phi_1^L(1)\right]^T \left[\Sigma^{-1}\right]\left[\phi_1^L(1)\right]. \qquad (25)$$

Remark 6. It follows from the comparison between (20) and (24) that the FMA test significantly reduces the computational burden of the VTWL CUSUM test. Moreover, theoretical results of Theorem 1 will be illustrated by numerical examples to show that the FMA test performs better than the WL CUSUM test (i.e., the VTWL CUSUM test with equal thresholds $h_1 = \cdots = h_L$), which is an asymptotically optimal test for the classical non-Bayesian change detection, introduced in Willsky and Jones (1976) and studied in Lai (1995, 1998), in the next section.

6. APPLICATION AND NUMERICAL EXAMPLES

In this section, the proposed and other conventional algorithms are utilized for detecting cyber/physical attacks on a simple SCADA water distribution network.

6.1 SCADA water distribution network

Let us consider a simple SCADA water distribution network as shown in Fig. 1. The water network is comprised of a treatment plant W_1, a reservoir R_1, a pump P_1, 3 junctions N_2, N_3 and N_4, 4 pipelines C_{01}, C_{12}, C_{23}, and C_{24} and 2 consumers d_1 and d_2. Two pressure sensors S_1 and S_2 are equipped for measuring pressure heads h_1 at the reservoir and h_2 at the node N_2, respectively.

The linearized model of the water network (in Fig. 1) can be described in the discrete-time state space model (1), where $x_k \in \mathbb{R}$ is the pressure head h_1 at the reservoir, $u_k \in \mathbb{R}$ is the control signals sent from the control center to the local controller for regulating the flow rate Q_{01} through the pump, $d_k \in \mathbb{R}^2$ is the disturbances corresponding to the consummation of customers at nodes N_3 and N_4, $y_k \in \mathbb{R}^2$ is the measurements of sensors S_1 and S_2. The process noises $w_k \sim \mathcal{N}(0, Q)$ and the sensor noises $v_k \sim \mathcal{N}(0, R)$; the matrices $A \in \mathbb{R}^{1\times1}$, $B \in \mathbb{R}^{1\times1}$, $F \in \mathbb{R}^{1\times2}$, $C \in \mathbb{R}^{2\times1}$, $D \in \mathbb{R}^{2\times1}$, $G \in \mathbb{R}^{2\times2}$, $Q \in \mathbb{R}^{1\times1}$, and $R \in \mathbb{R}^{2\times2}$ (corresponding to $n = 1$, $m = 1$, $p = 2$, $q = 2$).

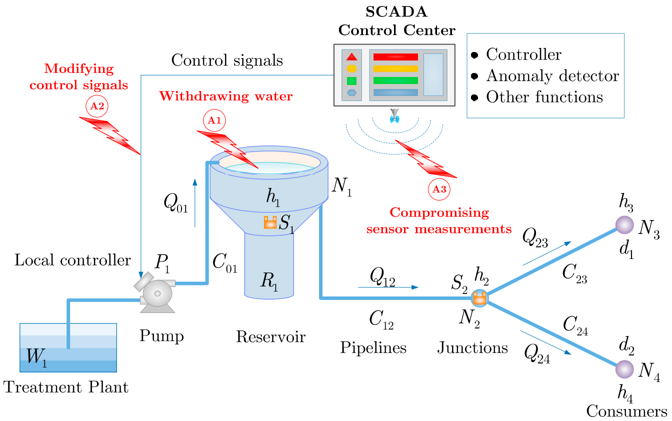


Fig. 1. The simple SCADA water distribution network in our case study.

For the demonstration purpose, let us consider an attack scenario where the attacker performs coordinated attacks by stealing water from the reservoir with a constant flow rate Q_0, turning off the pump P_1 and compromising the measurements of sensors S_1 and S_2 during the attack period $\tau_a = [k_0, k_0 + L - 1]$, where k_0 is the unknown attack instant and L is the known attack duration. This attack scenario is motivated by a real attack on city water utility where the pump was burned out after being turned on and off, as reported in Zetter (2011). As a result, the attack vector $a_k \in \mathbb{R}^4$ is designed by the adversary and the matrices $B_a \in \mathbb{R}^{1\times4}$ and $D_a \in \mathbb{R}^{2\times4}$ are decided by system operators (corresponding to $r = 2$ and $s = r + p = 4$).

The simulation parameters are chosen as follows. The sample time $T_S = 100$s and the initial pressure head $\overline{x}_0 = 100$m. The system matrices $A = 1$, $B = 0.5$, $F = [\,-0.5\ \ -0.5\,]$, $C = \begin{bmatrix}1\\1\end{bmatrix}$, $D = \begin{bmatrix}0\\0\end{bmatrix}$, $G = \begin{bmatrix}0 & 0\\-10 & -10\end{bmatrix}$. The attack matrices $B_a = [\,0.5\ 0.5\ 0\ 0\,]$ and $D_a = \begin{bmatrix}0&0&0&0\\0&0&0&1\end{bmatrix}$. The covariance matrices $Q = 0.02$ (and $Q = 0.2$) and $R = \begin{bmatrix}1&0\\0&1\end{bmatrix}$. The stolen flow rate is $Q_0 = 0.2\,\mathrm{m}^3/\mathrm{s}$. The attack duration $L = 8$ observations, corresponding to a period of 13.3min. The false alarm rate is measured by the time window of length $m_\alpha = 3L = 24$ observations, being equivalent to a duration of 40min. The attack vector $a_k \in \mathbb{R}^4$ is designed by the covert attack strategy, which was first introduced in Smith (2011), as follows:

$$a_k = \begin{cases} [0] & \text{if } k < k_0 \\ \begin{bmatrix} -0.2 \\ -1 \\ 0.6\,(k - k_0) \\ 0.6\,(k - k_0) \end{bmatrix} & \text{if } k_0 \le k < k_0 + L, \\ [0] & \text{if } k \ge k_0 + L \end{cases} \tag{26}$$

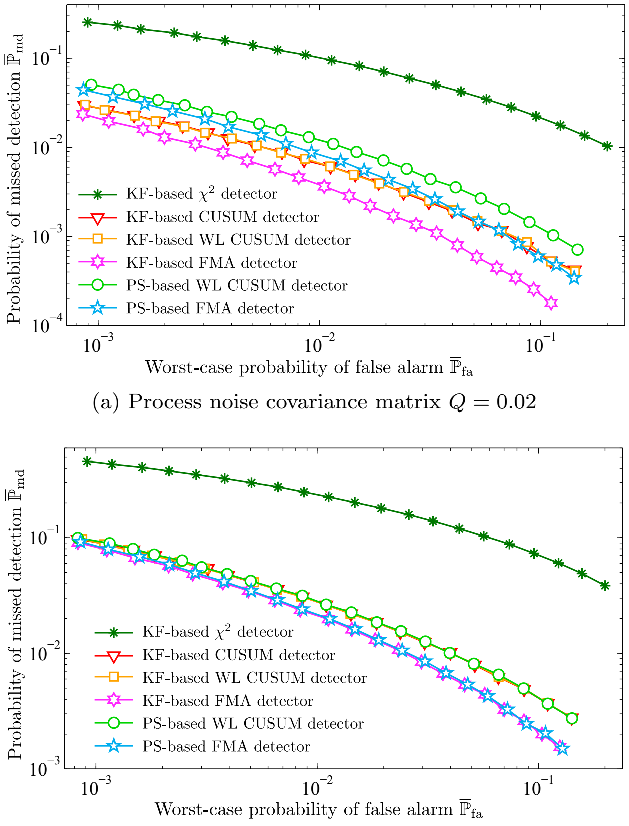
where $[0]$ is the null vector.

Remark 7. The attack vector $a_k \in \mathbb{R}^4$ contains all information describing the model of the attack. The first element reflects the physical attack to withdraw water from the reservoir with the flow rate of $Q_0 = 0.2\,\mathrm{m}^3/\mathrm{s}$. The second element reflects the cyber attack on the control signals for turning off the pump. The modification of the sensor measurements by the covert attack strategy Smith (2011) is described by the two last elements.

Remark 8. The idea of the covert attack is to coordinate the state attack vector and the sensor attack vector for realizing his malicious target while bypassing traditional anomaly detectors. The sensor attack vector is designed in such a way that it can compensate for the negative impact (i.e., the change in the distribution of the residuals) caused by the state attack vector. It has been shown that the covert attack is completely stealthy to traditional anomaly detectors if the attacker is able to compromise all sensors. From the defender's point of view, this stealthy attack needs to be rendered detectable before applying any detection algorithm. In other words, reliable measurements containing some information about the attack must be transmitted successfully to the detection algorithm. For the demonstration purpose, we propose in this numerical example a simple countermeasure for rendering the covert attack detectable. This method consists in protecting sensor S_1 so that its measurements can not be modified by the attacker. This sensor protection scheme is reflected in the matrix D_a, where $D_a(1, 3) = 0$ means that sensor S_1 is secure and $D_a(2, 4) = 1$ signifies that sensor S_2 is vulnerable.

6.2 Numerical examples

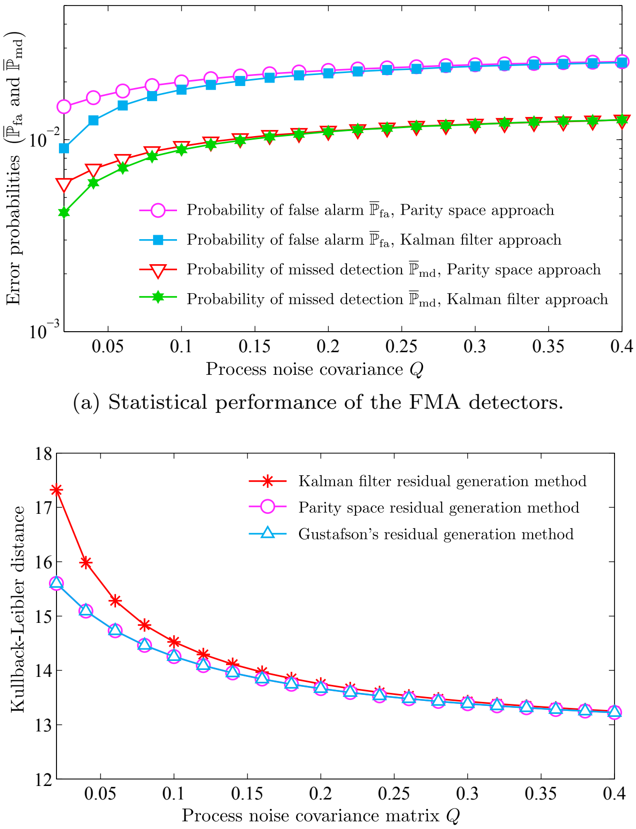
The statistical performances of some detection rules by 10^6 Monte Carlo simulation are given in Fig. 2. The thresholds are selected such that the worst-case probability $\overline{\mathbb{P}}_{fa}$ of all algorithms are almost the same for comparing the probability of missed detection $\overline{\mathbb{P}}_{md}$. For example, with process noise variance $Q = 0.2$ (in Fig. 2b), the thresholds $h_1 = \cdots = h_L$ of the WL CUSUM algorithm are chosen



(a) Process noise covariance matrix $Q = 0.02$

(b) Process noise covariance matrix $Q = 0.2$

Fig. 2. Comparison between the steady-state Kalman filter-based detectors (χ^2, CUSUM, WL CUSUM and FMA) and the fixed-size parity space-based detectors (WL CUSUM and FMA).



(a) Statistical performance of the FMA detectors.

(b) Kullback-Leibler distance.

Fig. 3. Comparison between the steady-state Kalman filter approach and the fixed-size parity space approach for different values of process noise covariance matrix Q.

from 3.0 to 7.9 and the thresholds $\tilde{h}_L$ of the FMA test are selected from 15.9 to 24.2, for both steady-state Kalman filter and fixed-size parity space approaches.

Following remarks can be drawn from the simulation results. Firstly, the proposed algorithms perform much better than the traditional non-parametric χ^2 detector. Secondly, given an acceptable level on the probability of false alarm, the probability of missed detection of the FMA tests is smaller than that of the WL CUSUM tests. In other words, the simple FMA tests outperform the WL CUSUM tests (for both the steady-state Kalman filter approach and the fixed-size parity space approach). Finally, the statistical performance of the Kalman filter-based algorithms is better than that of the parity space-based tests, particularly with small process noises.

The comparison between the steady-state Kalman filter approach and the fixed-size parity space approach is shown in Fig. 3. It can be seen from Fig. 3a that the steady-state Kalman filter approach outperforms the fixed-size parity space approach, especially when the process noise is small. This phenomenon is presented in Fig. 3b, where the K-L distances of the residuals generated by two approaches are drawn as functions of the process noise variance. The steady-state Kalman filter generates the residuals with higher K-L distance than the residuals generated by the fixed-sized parity space. The difference becomes significant if the process noise is small. In contrast, when the process noise is large, the difference is negligible. This phenomenon is explained by the fact that the approximation of the Bayesian approach, i.e., a steady-state Kalman filter, by the minimax approach, i.e., by the fixed-size parity space filter, produces a significant error only if the process noise is small and, hence, the *a priori* information plays an important role.

Moreover, the K-L distance of the residuals generated by the least-square estimation method proposed by Gustafsson (2002) coincides with the K-L distance of the residuals generated by the fixed-size parity space. It can be concluded that the least-square estimation method is as efficient as the parity space approach in generating the residuals for fault/attack detection and identification.

7. CONCLUSION

The sequential monitoring of SCADA systems against cyber/physical attacks has been considered in this paper. The main contributions of this paper are to develop the unified statistical model of the residuals generated by either the steady-state Kalman filter or the fixed-size parity space, to design the VTWL CUSUM algorithm adapted to the unified statistical model and to optimize the thresholds w.r.t. the transient change detection criterion. It is concluded from the simulation results that the steady-state Kalman filter approach generates the residuals with higher K-L distance, thus providing better statistical performance than the fixed-size parity space approach, especially when the process noise is small. In addition, the proposed FMA tests perform much better than the traditional χ^2, CUSUM, WL CUSUM tests (for both the steady-state Kalman filter and the fixed-size parity space approaches) with respect to the transient change detection criterion. The attack profiles are assumed to be completely known to simplify the theoretical analysis. The future work will be focused on the scenarios, where the attack profiles are (partially or completely) unknown. The sensibility analysis of the proposed FMA test w.r.t. operational parameters will be also considered.

ACKNOWLEDGEMENTS

This work was supported by the French National Research Agency (ANR)-through ANR CSOSG Program (Project ANR-11-SECU-0005).

REFERENCES

Amin, S., Litrico, X., Sastry, S., and Bayen, A. (2012a). Cyber security of water scada systems:(i) analysis and experimentation of stealthy deception attacks. *IEEE Transactions on Control Systems Technology*, 20.

Amin, S., Litrico, X., Sastry, S., and Bayen, A. (2012b). Cyber security of water scada systems:(ii) attack detection using an enhanced hydrodynamic model. *IEEE Transactions on Control Systems Technology*, 20.

Basseville, M. and Nikiforov, I. (1993). *Detection of abrupt changes: theory and application.* Prentice Hall Englewood Cliffs.

Brunner, M., Hofinger, H., Krauss, C., Roblee, C., Schoo, P., and Todt, S. (2010). Infiltrating critical infrastructures with next-generation attacks. *Fraunhofer Institute for Secure Information Technology (SIT), Munich.*

Cárdenas, A.A., Amin, S., Lin, Z.S., Huang, Y.L., Huang, C.Y., and Sastry, S. (2011). Attacks against process control systems: risk assessment, detection, and response. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, 355–366.

Chen, J. and Patton, R.J. (1999). *Robust model-based fault diagnosis for dynamic systems.* Kluwer academic publishers.

Esary, J.D., Proschan, F., and Walkup, D.W. (1967). Association of random variables, with applications. *Ann. Math. Statist.*, 38(5), 1466–1474.

Fouladirad, M., Freitag, L., and Nikiforov, I. (2008). Optimal fault detection with nuisance parameters and a general covariance matrix. *International Journal of Adaptive Control and Signal Processing*, 22(5), 431–439.

Fouladirad, M. and Nikiforov, I. (2005). Optimal statistical fault detection with nuisance parameters. *Automatica*, 41(7), 1157–1171.

Gorman, S. (2009). Electricity grid in us penetrated by spies. *Wall Street Journal*, 8.

Guépié, B.K. (2013). *Détection séquentielle de signaux transitoires : application à la surveillance d'un réseau d'eau potable.* Ph.D. thesis, Université de Technologie de Troyes.

Guépié, B.K., Fillatre, L., and Nikiforov, I. (2012a). Sequential detection of transient changes. *Sequential Analysis*, 31(4), 528–547.

Guépié, B.K., Fillatre, L., and Nikiforov, I.V. (2012b). Sequential monitoring of water distribution network. In *16th IFAC Symposium on System Identification, SYSID 2012, July 11-13*, 392–397. Brussels, Belgium.

Gustafsson, F. (2002). Stochastic fault diagnosability in parity spaces. In *Proceedings of the 15th IFAC World Congress*, 736–736.

Kwon, C., Liu, W., and Hwang, I. (2013). Security analysis for cyber-physical systems against stealthy deception attacks. In *American Control Conference (ACC), 2013*, 3344–3349.

Lai, T.L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 613–658.

Lai, T.L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *Information Theory, IEEE Transactions on*, 44(7), 2917–2929.

Lai, T.L. (2001). Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, 11(2), 303–350.

Lai, T.L. and Shan, J.Z. (1999). Efficient recursive algorithms for detection of abrupt changes in signals and control systems. *Automatic Control, IEEE Transactions on*, 44(5), 952–966.

Lehmann, E.L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5), 1137–1153.

Mo, Y., Garone, E., Casavola, A., and Sinopoli, B. (2010). False data injection attacks against state estimation in wireless sensor networks. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, 5967–5972.

Mo, Y. and Sinopoli, B. (2009). Secure control against replay attacks. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, 911–918.

Mo, Y. and Sinopoli, B. (2010). False data injection attacks in control systems. In *Preprints of the 1st Workshop on Secure Control Systems.*

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

Poor, H. and Hadjiliadis, O. (2009). *Quickest Detection.* Cambridge, University Press.

Slay, J. and Miller, M. (2007). Lessons learned from the maroochy water breach. *Critical Infrastructure Protection*, 73–82.

Smith, R.S. (2011). A decoupled feedback structure for covertly appropriating networked control systems. *Proc. IFAC World Congress*, 90–95.

Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014). *Sequential Analysis : Hypothesis Testing and Changepoint Detection.* CRC Press, Taylor & Francis Group.

Teixeira, A., Shames, I., Sandberg, H., and Johansson, K.H. (2012). Revealing stealthy attacks in control systems. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, 1806–1813.

Willsky, A. and Jones, H. (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *Automatic Control, IEEE Transactions on*, 21(1), 108–112.

Zetter, K. (2011). Attack on city water station destroys pump. `online`.