Journal of Process Control 116 (2022) 64–79

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Active multiplicative cyberattack detection utilizing controller switching for process systems

Shilpa Narasimhan, Nael H. El-Farra, Matthew J. Ellis [*]

Department of Chemical Engineering, University of California, Davis, One Shields Ave., Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 13 January 2022
Received in revised form 28 May 2022
Accepted 30 May 2022
Available online 15 June 2022

Keywords:
Multiplicative cyberattacks
Active attack detection
Controller switching
Process control systems

ABSTRACT

Multiplicative cyberattacks manipulating data over the process control system (PCS) communication links are cyberattacks that malicious agents may carry out against PCSs. These attacks are modeled by multiplying the data communicated over the link by a factor, and may be designed to be stealthy without extensive knowledge of process dynamics. The current work characterizes the relationship between the control system parameters, the closed-loop stability, and the detectability of a multiplicative sensor–controller communication link attack with respect to a class of residual-based detection schemes. The analysis reveals that control system parameters may be selected to aid in attack detection. Specifically, control system parameters, called attack-sensitive parameters, may be selected so that the closed-loop process is stable under attack-free operation and is destabilized by a cyberattack, rendering the attack detectable. With the attack-sensitive parameters, however, the attack-free closed-loop process performance may be worse than that with parameters selected based on standard design criteria. To address the potential trade-off between attack-free closed-loop performance and attack detectability, a novel active attack detection methodology utilizing control system parameter switching is developed. The control system switches between the nominal parameters (selected based on standard design criteria) and the attack-sensitive parameters to improve attack detection capabilities while avoiding substantial degradation in the attack-free closed-loop performance. The active detection methodology is applied to an illustrative chemical process example and shown to enhance the attack detection capabilities of two representative residual-based detection schemes.

1. Introduction

Complex cyberattacks on industrial control systems (ICSs) have demonstrated the proficiency of modern-day cyberattackers in side-stepping traditional information technology-based cybersecurity approaches for ICSs [1–4]. This has motivated research work on controller-based approaches for enhancing the cyberattack resilience of ICSs [5–7]. Cyberattack resilience of an ICS may be defined as the ability to detect, identify, and recover from a cyberattack [8–10]. Several approaches for designing detection, identification, and mitigation schemes have been proposed (see, for example, the reviews [11–13]). Recent research has focused on incorporating cyberattack resilience as part of the ICS design to develop inherently cyberattack resilient controller designs. For example, a gain adjustable observer design with a dynamic event-triggered communication scheme was proposed that compensates for aperiodic communication-jamming denial of service attacks [14]. A linear matrix inequality-based framework was derived for control system parameter selection. A secure polynomial control scheme based on secret sharing and a multi-party computation approach for evaluating the polynomial feedback control schemes was proposed [15]. The control scheme eliminates and limits direct unencrypted communication between different ICS components, making it difficult for an attacker to understand the process data based on the data flow in the ICS network.

False-data injection cyberattacks targeting process control systems (PCSs) may compromise the integrity of the control law, the data communicated over the controller–actuator communication link, or the data communicated over the sensor–controller communication link. The objective of an attack could be to cause instability or cause adverse economic, environmental, or human life impacts. Carrying out such attacks while avoiding detection by process monitoring systems may be a goal of a cyberattacker. Due to the complex dynamics of chemical processes, the design of false-data injection attacks achieving these objectives may require that the attacker possess some process knowledge [16]. Several cyberattacks targeting communication links of

[*] Corresponding author.
E-mail address: mjellis@ucdavis.edu (M.J. Ellis).

https://doi.org/10.1016/j.jprocont.2022.05.014

PCSs have been considered in the literature [17–21]. Two common models of false-data injection attacks targeting the sensor–controller link include additive attacks [22], and multiplicative attacks [23]. Additive sensor–controller link attacks inject false data by adding a factor to the sensor measurement data communicated over the link, leading to the controller receiving the sensor measurement plus a factor added to it. Multiplicative sensor–controller link attacks inject false data by multiplying the data in the communication link by a factor.

Stealthy attacks are designed to evade detection by falsifying data in the communication links, making the data difficult to distinguish from the data of the attack-free process. As a result, stealthy attacks are especially challenging to detect [24]. This realization has motivated detection scheme designs aimed at enabling the detection of stealthy attacks [22,23,25–31]. Additive sensor–controller link attacks may need careful design to remain stealthy, requiring process information (see Remark 5 in [32]). Multiplicative sensor–controller link cyberattacks are unique because they may be designed to be stealthy without requiring intimate knowledge of process dynamics (see Remark 6 in [32]), and are the focus of the present work.

Attack detection methods can be broadly divided into two categories, including passive and active attack detection schemes. Passive attack detection schemes monitor a process for anomalies based on regular operational data without employing external intervention or applying a perturbation. These schemes have been extensively explored [22,25–27]. For example, one passive scheme differentiates the behavior of an attacked process from its attack-free behavior by characterizing the skewness in the detection metric distribution [25]. Another approach uses a two-tier controller-detector architecture, with a neural network-based detection scheme to monitor for some attacks [33]. Other approaches use standard residual-based detection schemes such as the cumulative sum (CUSUM) or χ^2 detection schemes to identify anomalous behavior [22,26,27]. False-data injection cyberattacks targeting phasor measurement units have been considered where conditions were derived for undetectable additive and multiplicative attacks with a standard detection scheme (e.g., χ^2 detection scheme) [27]. An enhanced detection scheme was proposed. Closed-loop systems, where falsified output measurements are used in the controller, were not considered. The use of the CUSUM and χ^2 detection schemes for monitoring closed-loop systems under additive false-data injection sensor-control link attacks was considered in [22,26].

Passive approaches for attack detection may not always be successful in differentiating the anomalous behavior in the attacked process from its attack-free behavior (e.g., Section III, [28]). As an alternative, an active detection method may potentially enhance the detection capabilities. Active attack detection methods involve external intervention to induce an attack-detecting perturbation in the closed-loop process [23,28–31]. Two active detection methods were presented in [28]. The first approach utilizes a watermarking scheme, i.e., a secret noisy input is added to the computed control input to the process, and an attack is detected if the distribution of the detection metric deviates from the distribution expected for an attack-free process. The second approach uses a moving target scheme, i.e., the original system is augmented with an authenticating subsystem with time-varying dynamics and additional sensors to estimate the subsystem state. In this approach, the attack detection scheme is based on the difference between the (potentially) falsified output and the expected output. An approach that uses a combination of watermarking and a moving target scheme has also been explored [29]. In [30], the detectability of stealthy attacks exciting the zero-dynamics was characterized as a function of the observability of the attacked process. Leveraging this characterization, detection schemes that use redundant sensors and actuators were proposed to enable the detection of a zero-dynamics exciting attack. A few active detection approaches for the detection of multiplicative cyberattacks have been proposed [23,31]. Specifically, a watermarking approach that adds a constant to the sensor measurement before communicating the resulting value to the controller was proposed for multiplicative attacks targeting sensor–controller and controller–actuator links [23]. The controller subtracts the constant before computing its control action. Another watermarking scheme was presented in [31], utilizing an additive secret signal with a known distribution to the control input to detect several attacks such as multiplicative cyberattacks. The sensor data reported by all sensors are subject to two tests developed based on a statistical hypothesis testing criterion.

The connection between the control system design and multiplicative sensor–controller link attack detectability has been considered in [32]. Considering this connection, a controller screening methodology was proposed to identify and discard control system parameters that mask an attack from a class of residual-based detection schemes. This screening methodology provides a framework for the inclusion of cybersecurity considerations within the standard control design criteria (e.g., closed-loop stability, performance considerations, and robustness to uncertainty [34–36]). To the best of the authors' knowledge, a rigorous characterization of the relationship between closed-loop stability, control system parameter selection, and attack detectability has not been carried out. Motivated by this, the theoretical results presented in this work rigorously characterize the relationship between closed-loop stability, control system parameters, and attack detectability for a residual-based detection scheme. The results are used to identify a set of control system parameters (called "attack-sensitive" parameters) under which a multiplicative sensor–controller link attack can destabilize the closed-loop system. The selection of attack-sensitive control system parameters can enhance the ability to detect attacks, but can also degrade the performance of the attack-free closed-loop system. A novel active attack detection methodology employing control system parameter switching is developed to balance this trade-off. The controller switches between the nominal control system parameters, chosen based on standard control design criteria, and the attack-sensitive parameters with the proposed detection method. The main contributions of the paper include (1) a rigorous analysis of the relationship between closed-loop stability, the control design parameters, and the attack detectability for a residual-based detection scheme, and (2) the development of an active detection methodology that utilizes control system parameter switching to enhance the detection capabilities of the detection scheme.

The remainder of the paper is organized as follows: In Section 2, the notation, process model, model of the multiplicative sensor–controller link cyberattack, and control system design are presented. In Section 3, the residual-based detection scheme and the proposed active detection method via control parameter switching are presented. In Section 4, the application of the active detection method to enhance the attack detection capabilities of a residual-based detection scheme is demonstrated using a chemical process example consisting of a continuous stirred tank reactor (CSTR). A linearized CSTR model is considered to verify the theoretical results. Additionally, an extension of the theoretical results presented in this paper is investigated by applying the proposed approach to the nonlinear CSTR model. The enhanced detection capabilities are demonstrated considering the proposed residual-based detection and CUSUM detection schemes.

2. Preliminaries

2.1. Notation and definitions

For a vector $x \in \mathbb{R}^n$, its Euclidean norm is denoted by $\|x\|$, and its infinity norm is denoted by $\|x\|_\infty$. The closed Euclidean ball and infinity ball centered at the origin with radius $R > 0$ are denoted by $B^n(R) := \{x \in \mathbb{R}^n \mid \|x\| \leq R\}$ and $B^n_\infty := \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq R\}$. For a compact set $D \subset \mathbb{R}^n$, R_D denotes the minimal radius of the Euclidean ball enclosing the set, i.e., $R_D := \max_{x \in D} \|x\|$. For a set $D \subset \mathbb{R}^n$, the linear transformation of the set is denoted by $AD := \{Ax \mid x \in D\}$. Given two nonempty sets $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^n$, their Minkowski sum is defined as $X \oplus Y = \{x+y \mid x \in X, y \in Y\}$. For matrices, $\mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_n)$ represents an $n \times n$ diagonal matrix with diagonal elements $\beta_1, \beta_2, \ldots, \beta_n$, I represents the identity matrix of appropriate dimensions, and $\lambda_i(A)$ is the ith eigenvalue of the matrix A. Sequences are denoted with boldface letters, i.e., $\mathbf{d} := \{d(0), d(1), d(2), \ldots\}$ where $d(t) \in \mathbb{R}^n$ for all $t \geq 0$. For the discrete-time linear system: $z(t+1) = Az(t) + v(t)$, where $z(t) \in \mathbb{R}^n$, $v(t) \in V$ for all $t \geq 0$, and V is a compact set, a set $D_z \subset \mathbb{R}^n$ is said to be robust positively invariant if $z(t) \in D_z$ implies that $z(t+1) \in D_z$ for any $v(t) \in V$. A set $M_z \subset \mathbb{R}^n$ is said to be a minimum robust positively invariant set if M_z is contained within every closed robust positively invariant set [37]. For simplicity of presentation, the minimum robust positively invariant set will be referred to as the minimum invariant set in this paper.

2.2. Class of processes and control system design

In this work, processes modeled by discrete-time linear time-invariant systems and subject to bounded process disturbances and bounded measurement noise are considered:

$$x(t+1) = Ax(t) + Bu(t) + Gw(t) \tag{1}$$

where $x(t) \in \mathbb{R}^{n_x}$ is the process state vector, $u(t) \in \mathbb{R}^{n_u}$ is the manipulated input vector, $w(t) \in W \subset \mathbb{R}^{n_w}$ is the bounded process disturbance vector, and the set W is assumed to be a (compact) polytope containing the origin. Without loss of generality, the initial time is taken to be zero. The matrices A, B, and G are of appropriate dimensions. The value of the measured output received by the controller may be corrupted by a multiplicative sensor–controller link attack. The measured output is modeled by:

$$y(t) = \Lambda(Cx(t) + v(t)) \tag{2}$$

where $y(t) \in \mathbb{R}^{n_y}$ is the potentially falsified output vector received by the controller, $v(t) \in V \subset \mathbb{R}^{n_y}$ is the measurement noise vector, the set V is assumed to be a (compact) polytope containing the origin, and Λ is the matrix modeling multiplicative sensor–controller link attack on the process. The matrix C is of appropriate dimensions. The matrix Λ is referred to as the attack magnitude where $\Lambda \neq I$ indicates the presence of an attack on the process and $\Lambda = I$ indicates the absence of an attack.

The matrix pair (A, B) is assumed to be controllable, and the matrix pair (A, C) is assumed to be observable. A Luenberger observer is used to estimate the process states and is given by:

$$\begin{aligned} \hat{x}(t+1) &= A\hat{x}(t) + Bu(t) + L(y(t) - \hat{y}(t)) \\ \hat{y}(t) &= C\hat{x}(t) \end{aligned} \tag{3}$$

where $\hat{x}(t) \in \mathbb{R}^{n_x}$ is the estimated state vector, $\hat{y}(t) \in \mathbb{R}^{n_y}$ is the estimated output vector, and $L \in \mathbb{R}^{n_x \times n_y}$ is the observer gain selected so the eigenvalues of $A - LC$ are within the unit circle. Without loss of generality, the desired operating steady-state for the process is assumed to be the origin. To steer the process state to the origin, a linear feedback control law is used:

$$u(t) = -K\hat{x}(t) \tag{4}$$

where $K \in \mathbb{R}^{n_u \times n_x}$ is the controller gain, selected such that the eigenvalues of $A - BK$ are within the unit circle.

The estimation error is defined as $e(t) = x(t) - \hat{x}(t)$, and the estimation error dynamics are given by:

$$e(t+1) = L(I - \Lambda)Cx(t) + (A - LC)e(t) + Gw(t) - L\Lambda v(t) \tag{5}$$

To analyze the stability of the overall closed-loop process consisting of the process in Eqs. (1)–(2) with the feedback control law in Eq. (4) using the estimated state from the observer in Eq. (3), an augmented state vector $\xi(t) = [x^T(t)\ e^T(t)]^T$ is defined. The augmented state dynamics are given by:

$$\xi(t+1) = \underbrace{\begin{bmatrix} (A - BK) & BK \\ L(I - \Lambda)C & (A - LC) \end{bmatrix}}_{=:A_\xi(\Lambda,K,L)} \xi(t) + \underbrace{\begin{bmatrix} G & 0_{n_x \times n_y} \\ G & -L\Lambda \end{bmatrix}}_{=:B_\xi(\Lambda,K,L)} d(t) \tag{6}$$

where $d(t) := \begin{bmatrix} w^T(t) & v^T(t) \end{bmatrix}^T \in F$ is the augmented disturbance and measurement noise vector, and $F := \left\{ \begin{bmatrix} w \\ v \end{bmatrix} \mid w \in W, v \in V \right\}$ is the set of disturbances. Here, $A_\xi(\Lambda, K, L)$ and $B_\xi(\Lambda, L)$ are the system matrices for the augmented state dynamics. In the remainder, the admissible set of disturbance and measurement noise sequences is denoted by $\mathcal{F} := \{\mathbf{d} \mid d(t) \in F,\ \forall\ t \geq 0\}$.

Given that chemical processes are typically operated at steady-state for long periods, all analyses in the present work focus on the process operating at its steady-state, i.e., after the augmented state of the closed-loop process has converged to its terminal set, which is the minimum invariant set. The minimum invariant set for the augmented system in Eq. (6) when $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| < 1$ may be expressed as the infinite Minkowski sum [37]:

$$D_\xi(\Lambda, K, L) = \bigoplus_{i=0}^{\infty} A^i_\xi(\Lambda, K, L) B_\xi(\Lambda, L)\, F \tag{7}$$

Based on Eq. (7), the minimum invariant set of the augmented closed-loop system is dependent on the attack matrix Λ, the controller gain K, and the observer gain L. For simplicity, the process operated at steady-state refers to the system of Eq. (6) after the augmented state has converged to the minimum invariant set, i.e., $\xi(t) \in D_\xi(\Lambda, K, L)$ implying that $\xi(t+1) \in D_\xi(\Lambda, K, L)$ for any $d(t) \in F$. For the remainder, the term closed-loop process refers to the process described by Eqs. (1)–(2) under the feedback law given by Eq. (4) using the estimates of states generated by the observer in Eq. (3).

3. Active multiplicative attack detection utilizing controller switching

In this section, the residual-based detection scheme considered is introduced, and a detectability-based classification of attacks is presented. Theoretical results characterizing the relationship between closed-loop stability, control system parameter selection, and the detectability of an attack with respect to the residual-based detection scheme considered are presented. The results of this analysis are used to develop an active attack detection methodology using occasional control system parameter switching.

3.1. Residual-based detection scheme and attack detectability

For the closed-loop process, the residual vector ($r(t)$) is defined as the difference between the output ($y(t)$) and its estimate generated by the observer ($\hat{y}(t)$), i.e.,

$$r(t) := y(t) - \hat{y}(t)$$

Writing the residual in terms of the augmented state ($\xi(t)$) and the disturbance vector ($d(t)$) yields:

$$r(t) = \underbrace{\begin{bmatrix}(\Lambda - I)C & C\end{bmatrix}}_{=:A_r(\Lambda)} \xi(t) + \underbrace{\begin{bmatrix}0_{n_y \times n_w} & \Lambda\end{bmatrix}}_{=:B_r(\Lambda)} d(t) \tag{8}$$

When $A_\xi(\Lambda, K, L)$ has eigenvalues that lie within the unit circle and F is compact, $D_\xi(\Lambda, K, L)$ is forward invariant [37] and compact (Sec. 4 in [38]), and the residual is ultimately bounded within a terminal set. From Eq. (8), the residual terminal set is given by:

$$D_r(\Lambda, K, L) = A_r(\Lambda) D_\xi(\Lambda, K, L) \oplus B_r(\Lambda) F \tag{9}$$

For every $\xi(t) \in D_\xi(\Lambda, K, L)$ and $\mathbf{d} \in \mathcal{F}$, all possible realizations of the residual will be contained within its terminal residual set, i.e., $r(t) \in D_r(\Lambda, K, L)$. Based on Eq. (8), in the absence of an attack ($\Lambda = I$), the residual is dependent on the estimation error (Eq. (5)) and the disturbance ($d(t)$). However, in the presence of a multiplicative sensor–controller link attack ($\Lambda \neq I$), the residual is also coupled to the process state. In addition to its dependence on the disturbance set F, the minimum invariant set is dependent on both the controller gain (K) and the observer gain (L). This is true for both the attack-free and the attacked process. However, the dependency of the terminal residual set on the controller and observer gains varies for the attack-free and the attacked processes. Specifically, the attack-free terminal residual set is dependent on the observer gain only, whereas the attacked terminal residual set is dependent on the both the controller gain and the observer gain. Nonetheless, to maintain uniformity of notation, $D_r(I, K, L)$ is used to represent the attack-free terminal residual set even though the terminal residual set is independent of K when $\Lambda = I$.

Residual-based anomaly detection schemes are model-based detection schemes that are commonly used for process monitoring [39–43]. These detection schemes monitor the process without using external intervention. Consequently, they are passive detection schemes. Two types of residual-based detection schemes commonly employed for cyberattack detection are the χ^2 and CUSUM detection schemes [22,26]. Both schemes are scalar detection schemes in the sense that their output values are scalar values. To monitor changes in the residual behavior over time, the schemes may be formulated using the 2-norm of the residual vector as the input driving the detector output (see, for example, [32] for further discussion on this point). To tune the detector to raise zero false alarms when the process is operating at steady-state, the tuning must account for the fact that the maximum achievable value of the 2-norm of the residual is equal to the radius of the ball enclosing the residual terminal set, i.e., $\|r(t)\| \leq R_{D_r}(I, K, L)$ where is $R_{D_r}(I, K, L)$ is the minimum radius of the 2-norm ball enclosing the residual terminal set ($D_r(I, K, L) \subseteq B^{n_y}(R_{D_r}(I, K, L))$) [32]. A limitation of such detection schemes is that they do not account for the shape of the terminal residual set of the attack-free closed-loop process $D_r(I, K, L)$. For example, if the residual of the attacked closed-loop process is such that it is outside the terminal residual set of the attack-free closed-loop process but bounded within the 2-norm ball enclosing the terminal residual set of the attack-free closed-loop process ($r(t) \in B^{n_y}(R_{D_r}(I, K, L)) \setminus D_r(I, K, L)$), the 2-norm residual-based detection schemes will not detect the attack. To overcome this limitation, a set membership-based detection scheme is considered in this work. Specifically, the detection scheme given by:

$$z(t) = \begin{cases} 0, & r(t) \in D_r(I, K, L) \\ 1, & r(t) \notin D_r(I, K, L) \end{cases} \tag{10}$$

is considered, where $z(t)$ represents the output of the detection scheme. An output of $z(t) = 0$ indicates normal process operation (no attack detection), and $z(t) = 1$ indicates that an attack is detected. Since the set membership-based detection scheme does not use external intervention to monitor the process, it is considered a passive detection scheme.

Cyberattacks may be classified based on the ability of the detection scheme in Eq. (10) to detect the attack. For the closed-loop process operated at steady-state monitored by the detection scheme in Eq. (10), an attack is said to be detected at time t_d if $r(t_d) \notin D_r(I, K, L)$ with the output of the detection scheme $z(t_d) = 1$. An attack is defined as a detectable attack with respect to the detection scheme in Eq. (10) if the attack is detected in finite time for all $\xi(0) \in D_\xi(\Lambda, K, L)$ and $\mathbf{d} \in \mathcal{F}$. If the attack renders the closed-loop process unstable, then by convention, the set $D_\xi(\Lambda, K, L)$ is taken to be the Euclidean space $\mathbb{R}^{2n_x}$. An attack is defined as an undetectable attack with respect to the detection scheme in Eq. (10), if the residual of the attacked closed-loop process satisfies $r(t) \in D_r(I, K, L)$ for all $t \geq 0$ for all $\xi(0) \in D_\xi(\Lambda, K, L)$ and $\mathbf{d} \in \mathcal{F}$. Finally, an attack is defined as potentially detectable with respect to the detection scheme in Eq. (10), if the attack is neither detectable nor undetectable. The set of initial conditions considered is $D_\xi(\Lambda, K, L)$ because steady-state operation is considered. For some initial conditions in $D_\xi(\Lambda, K, L)$, the attack is detected immediately by the detection scheme in Eq. (10). However, this does not imply that the attack is detectable, as the attack needs to be detected in finite-time for all initial conditions in $D_\xi(\Lambda, K, L)$. While the definitions for attack detectability with respect to the detection scheme in Eq. (10) are valid for any attack, multiplicative sensor–controller link attacks are considered in the present work. Owing to the process disturbances and measurement noise, the augmented process states of the stable closed-loop process (Eq. (6)) are ultimately bounded within its minimum invariant set. Thus, the notion of closed-loop stability considered is ultimate boundedness of the augmented state of the closed-loop process. The closed-loop process in Eq. (6) is considered to be unstable if $\|\xi(t)\| \to \infty$ as $t \to \infty$.

To motivate the proposed active detection methodology, the relationship between closed-loop stability and detectability is analyzed first. Proposition 1 below establishes a relationship between the undetectability of a multiplicative attack and the terminal residual sets of the attack-free and attacked closed-loop process.

Proposition 1. *Consider the closed-loop process operated at steady-state with control system parameters (K, L) under a multiplicative sensor–controller link attack of magnitude Λ. If the attack is such that the closed-loop process remains stable, i.e., the eigenvalues of $A_\xi(\Lambda, K, L)$ lie within the unit circle, the multiplicative attack is undetectable with respect to the detection scheme in Eq. (10), if and only if $D_r(\Lambda, K, L) \subseteq D_r(I, K, L)$.*

Proof. Consider the closed-loop process operated at steady-state with control system parameters (K, L) under a multiplicative sensor–controller link attack of magnitude Λ. If the pair (K, L) are stabilizing under the attack, then the minimum invariant set of the process $D_\xi(\Lambda, K, L)$ is compact and forward invariant. Additionally, the augmented state of the attacked closed-loop process is bounded within its minimum invariant set for all time, i.e., $\xi(t) \in D_\xi(\Lambda, K, L)$ for all $t \geq 0$. As a result, the residuals of the attacked closed-loop process are also bounded within the terminal set of residuals, i.e., $r(t) \in D_r(\Lambda, K, L)$ for all $\xi(0) \in D_\xi(\Lambda, K, L)$ and $\mathbf{d} \in \mathcal{F}$. If the terminal residual set of the attacked process is a subset of or equal to the terminal residual set of the attack-free process ($D_r(\Lambda, K, L) \subseteq D_r(I, K, L)$), the residuals of the attacked process are contained within its attack-free terminal residual set, i.e., $r(t) \in D_r(\Lambda, K, L) \subseteq D_r(I, K, L)$ for all $t \geq 0$ and the attack is undetectable. Hence, the attack is undetectable if $D_r(\Lambda, K, L) \subseteq D_r(I, K, L)$.

To show that $D_r(\Lambda, K, L) \subseteq D_r(I, K, L)$ is also a necessary condition for undetectability, the proof proceeds by contradiction. Assume there is an undetectable multiplicative sensor–controller link attack of magnitude Λ on the closed-loop process with control system parameters (K, L) such that the attacked closed-loop process is stable and the terminal residual set of the attacked process is not a subset of or equal to the terminal residual set of the attack-free process, i.e., $D_r(\Lambda, K, L) \nsubseteq D_r(I, K, L)$. Based on the definition of undetectable attacks, the multiplicative sensor–controller link attack is such that for any augmented state initialized in the minimum invariant set of the attacked process $\xi(0) \in D_\xi(\Lambda, K, L)$ and all $\mathbf{d} \in \mathcal{F}$, the residuals of the attacked process are contained within the attack-free terminal residual set, i.e., $r(t) \in D_r(I, K, L)$ for all time $t \geq 0$. Since $D_r(\Lambda, K, L) \nsubseteq D_r(I, K, L)$, the set $D_r(\Lambda, K, L) \setminus D_r(I, K, L)$ is non-empty. Moreover, there exist $\xi(0) \in D_\xi(\Lambda, K, L)$ and $\mathbf{d} \in \mathcal{F}$ that result in $r(t) \in D_r(\Lambda, K, L) \setminus D_r(I, K, L)$ for some $t \geq 0$ implying that $r(t) \notin D_r(I, K, L)$ for some $t \geq 0$. This leads to a contradiction, completing the proof. □

From Proposition 1, the question may arise as to whether a conventional condition for instability, i.e., $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| > 1$, and/or $D_r(\Lambda, K, L) \nsubseteq D_r(I, K, L)$ are sufficient conditions for a detectable attack. However, these conditions alone are not sufficient conditions for a detectable attack, and can only be used to guarantee potential detectability of an attack, which is stated in the next proposition.

Proposition 2. *Consider the closed-loop process operated at steady-state with control system parameters (K, L) under a multiplicative sensor–controller link attack of magnitude Λ. If the attack is such that (1) the attacked closed-loop process is stable with the eigenvalues of $A_\xi(\Lambda, K, L)$ within the unit circle, and $D_r(\Lambda, K, L) \nsubseteq D_r(I, K, L)$, or (2) the attacked closed-loop process is such that $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| > 1$, then the attack is potentially detectable with respect to the detection scheme in Eq. (10).*

Proof. The proof is divided into two parts. Part 1 considers the case when the attacked closed-loop process is stable, but $D_r(\Lambda, K, L) \nsubseteq D_r(I, K, L)$. Part 2 considers the case when the attack renders the closed-loop process unstable in the conventional sense such that $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| > 1$.

Part 1: Consider that the attacked closed-loop process remains stable with the eigenvalues of $A_\xi(\Lambda, K, L)$ lying within the unit circle, and $D_r(\Lambda, K, L) \nsubseteq D_r(I, K, L)$. Since the origin is contained within the disturbance set (i.e., $0 \in F$), the origin is contained within the minimum invariant sets: $D_\xi(I, K, L)$ and $D_\xi(\Lambda, K, L)$ for the attack-free and attacked closed-loop process, respectively, from Eq. (7). If the disturbance is identically equal to 0 ($\mathbf{d} \equiv 0 \in \mathcal{F}$), the augmented state will be maintained at the origin for $\xi(0) = 0 \in D_\xi(\Lambda, K, L)$ implying that the residual of the attacked process is also maintained at the origin, which is within the attack-free terminal residual set, i.e., $r(t) = 0 \in D_r(I, K, L)$ for all $t \geq 0$. For such a realization of the disturbance and initial condition, the attack will go undetected for all $t \geq 0$. However, since $D_r(\Lambda, K, L) \nsubseteq D_r(I, K, L)$, $r(t) \in D_r(\Lambda, K, L) \setminus D_r(I, K, L)$ is possible for some $t \geq 0$, $\mathbf{d} \in \mathcal{F}$, and $\xi(0) \in D_r(\Lambda, K, L)$ following similar arguments as that used in the proof of Proposition 1. This implies that the attack is potentially detectable.

Part 2: Consider that the attacked closed-loop process is such that $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| > 1$. Similar logic as that used in Part 1 may be applied to show that the attack is potentially detectable. If $\xi(0) = 0$ and $\mathbf{d} \equiv 0 \in \mathcal{F}$, the attack is not detected. On the other hand, since $D_\xi(\Lambda, K, L) = \mathbb{R}^{2n_x}$ by convention when the closed-loop process is rendered unstable by the attack, there exist $\xi(0) \in \mathbb{R}^{2n_x}$ such that $r(0) \notin D_r(I, K, L)$ and the attack is detected at $t = 0$. Therefore, the attack is potentially detectable. □

If the closed-loop process under an attack is unstable such that $\|\xi(t)\| \to \infty$ as $t \to \infty$ the attack will be detected in finite time, if an additional observability condition is satisfied. This result is formally stated in Proposition 3.

Proposition 3. *Consider the closed-loop process with control system parameters (K, L) under a multiplicative attack of magnitude $\Lambda \neq I$. Let the control system parameters (K, L) stabilize the attack-free closed-loop process. If the attack renders the closed-loop process unstable in the sense that $\|\xi(t)\| \to \infty$ as $t \to \infty$ and the pair $(A_\xi(\Lambda, K, L), A_r(\Lambda))$ is observable, the attack is detected in finite time with respect to the detection scheme in Eq. (10).*

Proof. If the closed-loop process under attack is rendered unstable in the sense that $\|\xi(t)\| \to \infty$ as $t \to \infty$ and the pair $(A_\xi(\Lambda, K, L), A_r(\Lambda))$ is observable, the residuals are unbounded in the sense that $\|r(t)\| \to \infty$ as $t \to \infty$. This follows from Theorem 1 Appendix. Since the attack-free closed-loop process with control system parameters (K, L) is stable, its minimum invariant set $D_\xi(I, K, L)$ is a compact (closed and bounded) set. As a result, the attack-free terminal residual set is also a compact set (from Eq. (9)). There exists $R > 0$ such that $D_r(I, K, L) \subseteq B^{n_y}(R)$. Because the residuals of the attacked process are unbounded ($\|r(t)\| \to \infty$ as $t \to \infty$), for all $\epsilon > 0$ there exists $T > 0$ such that $\|r(t)\| > \epsilon$ for all $t > T$. Choosing $\epsilon > R$ shows there exists a finite time T_1 such that $\|r(T_1)\| > \epsilon > R$, which implies that $r(T_1) \notin D_r(I, K, L)$. Thus, the attack is detected in finite time and the attack is detectable. □

The only assumption made about the process disturbance and measurement noise is that they are bounded. Even if some eigenvalues of $A_\xi(\Lambda, K, L)$ are outside the unit circle, this assumption does not exclude potential realizations of the disturbance and measurement noise that results in the augmented state remaining bounded for all times. These are cases where the disturbance can effectively act as a stabilizing input in the sense that the state remains bounded for all time. In practice, disturbances are exogenous inputs and are not expected to stabilize a process. The condition $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| > 1$ is a necessary, but not sufficient, condition for the type of closed-loop instability considered in this work. Therefore, closed-loop instability cannot be verified solely by checking the eigenvalues of $A_\xi(\Lambda, K, L)$. Nevertheless, a multiplicative attack is said to be destabilizing if the eigenvalues of $A_\xi(\Lambda, K, L)$ are outside the unit circle ($\max_i |\lambda_i(A_\xi(\Lambda, K, L))| > 1$) to highlight that the attack is responsible for destabilization.

3.2. Active attack detection methodology

Traditional control system design approaches use closed-loop stability, performance, and robustness to uncertainty as criteria to determine the control system design [34–36]. Although attack detectability is linked to the control system design (Section 3.1), traditional design methods do not consider cyberattack detectability and may result in selecting control system parameters that mask the cyberattack in the sense that a cyberattack goes undetected with these parameters. From an attack detection

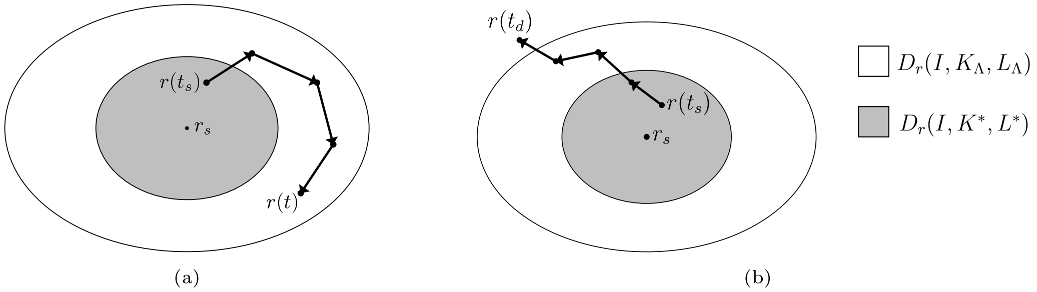


Fig. 1. (a) An example residual trajectory for the attack-free closed-loop process with the control system switch from nominal parameters to attack-sensitive parameters occurring at t_s. (b) An example residual trajectory for the attacked closed-loop process with the control parameter switch occurring at t_s where the attack is detected at t_d.

standpoint (Proposition 3), selecting control system parameters that are "sensitive" to cyberattacks, in the sense that the closed-loop process is rendered unstable by the attack, may be preferred. However, sustained operation with these control system parameters may not be desirable because the closed-loop performance may be worse than that achieved under parameters determined by traditional design approaches. To manage the trade-off between attack detection and closed-loop performance, we propose an active detection methodology that utilizes occasional switching from the nominal control system parameters, determined by traditional design approaches, to the so-called attack-sensitive parameters. Control system parameter switching is one form of active detection that may be considered, owing to the link between control system parameters and attack detectability established in Section 3.1.

The nominal parameters are denoted by (K^*, L^*), while the attack-sensitive parameters are denoted by (K_Λ, L_Λ). With the active detection methodology, the control system parameters switch from the nominal parameters to the attack-sensitive parameters at t_s. After the control system switches from the nominal parameters to attack-sensitive parameters, the process is operated over a period $T_c > 0$ with the attack-sensitive parameters. Under attack-free operations (Fig. 1(a)), the residual trajectory after the switch will evolve in the terminal residual set of the attack-free closed-loop process with attack-sensitive parameters $D_r(I, K_\Lambda, L_\Lambda)$. After the period T_c elapses, the control system switches back to the nominal parameters. In the presence of a multiplicative attack, the residual trajectory may evolve outside the terminal residual set of the attack-free closed-loop process with attack-sensitive parameters (Fig. 1(b)) resulting in the attack being detected.

Under the active detection methodology, the control system parameters vary over time. The detection scheme needs to account for this change because the residual terminal set under attack-free operation depends on the controller and observer gains. Therefore, the detection scheme is modified as follows:

$$z(t) = \begin{cases} 0, & r(t) \in D_r(I, K(t), L(t)) \\ 1, & r(t) \notin D_r(I, K(t), L(t)) \end{cases} \tag{11}$$

where $K(t)$ is the controller gain used at time step t, and $L(t)$ is the observer gain at time step t, $z(t) = 0$ indicates a lack of anomaly detection, and $z(t) = 1$ indicates anomalous operation is detected. For the closed-loop process with the nominal parameters, $(K(t), L(t)) = (K^*, L^*)$, and for the closed-loop process with the attack-sensitive parameters, $(K(t), L(t)) = (K_\Lambda, L_\Lambda)$.

The attack-sensitive parameters are chosen such that the attack-free closed-loop process operated with the attack-sensitive parameters is stable, and the process is destabilized by an attack. Particularly, the attack-sensitive parameters are chosen so that some eigenvalues of the augmented system matrix lie outside the unit circle (i.e., $\max_i |\lambda_i(A_\xi(\Lambda, K_\Lambda, L_\Lambda))| > 1$), and the matrix pair $(A_\xi(\Lambda, K_\Lambda, L_\Lambda), A_r(\Lambda))$ is observable. The attack-sensitive parameters are chosen to be sensitive to a range of attack magnitudes. Ideally, the attack-sensitive parameters may be chosen so that the range of attack magnitudes is as large as possible. The attack-sensitive parameters exploit the dependence of the terminal residual set on the control system parameters.

The switching instance t_s and the period T_c (called the cycle time) are the two design parameters for the proposed active detection methodology. The switching instance may be selected by a process operator based on operational considerations. For example, one way the switching instance may be selected is when the closed-loop performance degradation due to operation with attack-sensitive parameters is acceptable based on process economic considerations. The cycle time T_c may be selected to balance a potential trade-off between attack detection and closed-loop performance and safety considerations. Given that T_c is finite, the closed-loop augmented process state will remain bounded over the period when the attack-sensitive parameters are used in the control system (t_s to $t_s + T_c$). This is true even if the closed-loop process is subjected to a destabilizing multiplicative attack during this period. Furthermore, it is expected that the likelihood of detecting potentially detectable and detectable attacks scales with T_c. Rigorous evaluation of this expectation is beyond the scope of the present work. However, the attack-sensitive parameters could result in closed-loop performance deterioration for the attack-free process compared to performance under the nominal control system parameters. Operating with the attack-sensitive parameters for long periods may not be desirable from a closed-loop performance perspective. Furthermore, for destabilizing attacks on the process operated under either control system (i.e., with nominal parameters or with the attack-sensitive parameters), the bound on the process state scales with T_c. If there is a state-space set whereby the process is operated safely, then T_c should be selected to be small enough to ensure that the state is maintained within the safe set in the presence of a destabilizing attack. The closed-loop performance and safety considerations limit how long the cycle time should be.

The main benefit of the proposed approach is to enhance the attack detection capabilities. An attacker may select a multiplicative attack that destabilizes the closed-loop process under the nominal parameters. Such an attack may be detected by the detection scheme in Eq. (10). If the attack is undetectable with the nominal control system parameters, it will not be detected. Even if the attack is potentially detectable, it may go undetected by the detection scheme. The active detection methodology enables the detection of attacks that are designed to be potentially detectable or undetectable with respect to the closed-loop process under nominal parameters.

Fig. 2 illustrates the flowchart for the active attack detection methodology. The proposed active detection methodology is summarized by the algorithm below. The algorithm is initialized with $t = 0$. The parameters of the methodology are the switching instance t_s and the cycle time T_c.

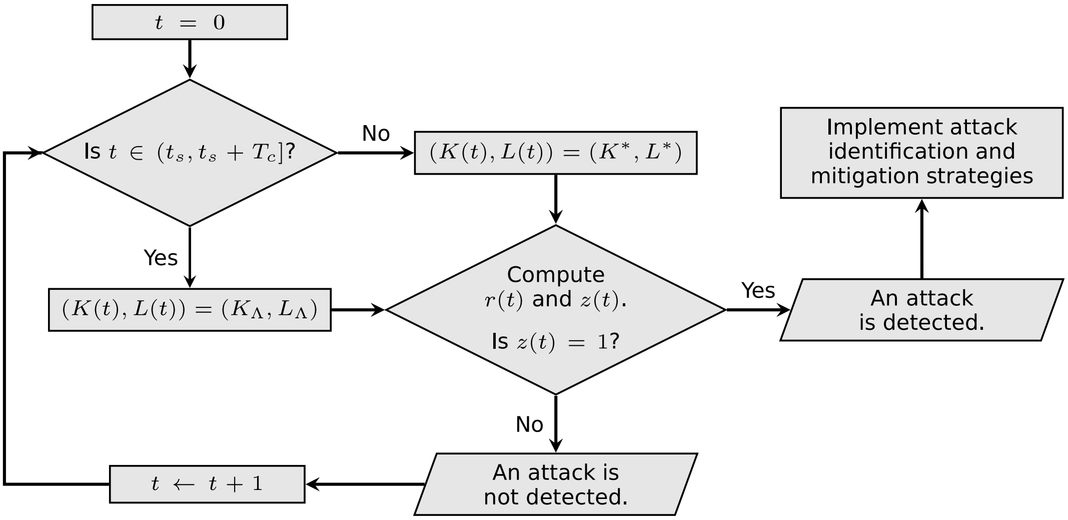

$t = 0$
Is $t \in (t_s, t_s + T_c]$?
No
$(K(t), L(t)) = (K^*, L^*)$
Implement attack identification and mitigation strategies
Yes
$(K(t), L(t)) = (K_\Lambda, L_\Lambda)$
Compute $r(t)$ and $z(t)$. Is $z(t) = 1$?
Yes
An attack is detected.
No
$t \leftarrow t + 1$
An attack is not detected.

Fig. 2. Flowchart for the active attack detection methodology.

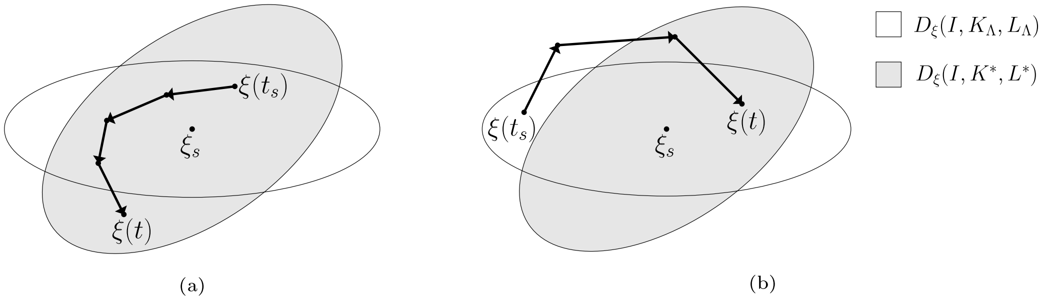


Fig. 3. (a) An example showing the evolution of the augmented state trajectory for the attack-free closed-loop process with a control parameter switch generating zero false alarms. (b) An example showing the evolution of the augmented state trajectory for the attack-free closed-loop process with a control system parameter switch that may generate false alarms.

1. If $t \in (t_s, t_s + T_c]$, set $(K(t), L(t)) = (K_\Lambda, L_\Lambda)$. Else, set $(K(t), L(t)) = (K^*, L^*)$.
2. Compute the residual $r(t)$ and the output of the detection scheme in Eq. (10).
3. If $z(t) = 1$, an attack is detected; implement attack identification and mitigation strategies. Else, an attack is not detected; go to Step 4.
4. Set $t \leftarrow t + 1$. Go to Step 1.

The methodology presented here illustrates a single switching cycle from the nominal to attack-sensitive parameters and back to the nominal parameters. To further enhance the detection capabilities, the methodology may be modified to include periodic switching from nominal to attack-sensitive parameters. Additionally, to detect a wider range of attack magnitudes, the methodology could be modified to include multiple control system switches from nominal parameters to other attack-sensitive parameters.

Under attack-free operation and when control system parameter switching takes place, the augmented state needs to be in the minimum invariant set of the attack-free process with the updated controller. In this way, the state moves from one minimum invariant set under one set of control system parameters to another. An example trajectory is shown in Fig. 3(a). In this case, the control system parameters switch from nominal to attack-sensitive parameters when the augmented state is within the intersection of the minimum invariant sets with the nominal and attack-sensitive parameters, i.e., $\xi(t_s) \in D_\xi(I, K^*, L^*) \cap D_\xi(I, K_\Lambda, L_\Lambda)$. After the switch, the augmented state moves from the minimum invariant set with nominal parameters to the minimum invariant set with attack-sensitive parameters. However, the augmented state is not measured. When the control parameters switch, the augmented state may be outside the minimum invariant set of the process under the updated controller. If the process is attack-free, the augmented state will converge to the minimum invariant set, but the residual during the transient period may take values outside the terminal residual set, triggering a false alarm. An example of this is shown in Fig. 3(b). For the attack-free process, the control system switches from nominal to attack-sensitive parameters at the time instance when the augmented state is outside the minimum invariant set associated with the attack-sensitive parameters, i.e., $\xi(t_s) \in D_\xi(I, K^*, L^*) \setminus D_\xi(I, K_\Lambda, L_\Lambda)$. As a result, the process dynamics is excited due to the switch, in the sense that the augmented state evolves briefly outside the minimum invariant set with attack-sensitive parameters. The residual during this transient period, may evolve outside the terminal residual set associated with the attack-sensitive parameters and trigger false alarms by the detection scheme in Eq. (11).

Under an attack, a control system parameter switch may result in the augmented state exhibiting a transient behavior that mimics the transient behavior of the attack-free process. An example is illustrated in Fig. 4. After a control system parameter switch to attack-sensitive parameters occurs at time t_s, the augmented state of the attacked closed-loop process evolves outside its attack-free minimum invariant set until time t_i. However, after time t_i, the augmented state evolves within the attack-free minimum invariant set until the time instance t_e. This may result

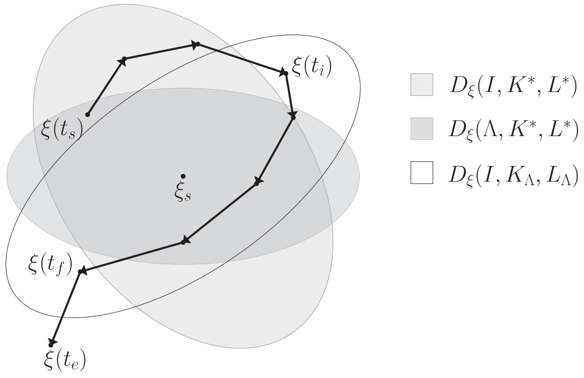


Fig. 4. An example showing the evolution of the augmented state trajectory for the attacked closed-loop process. After a control system parameter switch to attack-sensitive parameters, the augmented state mimics the transient behavior of an attack-free process briefly.

in the residuals of the process evolving briefly outside the attack-free terminal residual set, before converging to it. In this case, the alarms generated by the detection scheme monitoring the attacked process may be indistinguishable from the false alarm rate in the attack-free process.

Owing to the complications described above, false alarms are not desirable. To minimize false alarms, a modification to the detection scheme in Eq. (11) may be considered. In particular, the detection scheme may be modified to generate an alarm only if the residual remains outside the terminal set for a specified period. The period may be chosen to span a few sample times to account for the potential transient behavior in the attack-free process. Using a timer threshold whereby the detection logic must deem abnormal operating behavior over a period before raising an alarm is a common approach for minimizing nuisance alarms [44].

Remark 1. If an attack on the closed-loop process operated under either control mode, i.e., under nominal mode or under the attack-sensitive mode, is detected at any time $t_d \geq 0$, then attack identification and mitigation strategies could be employed to cope with the attack. These strategies are beyond the scope of this work and the subject of future work.

Remark 2. In addition to attack detection, the operating goals for a closed-loop process may be included as a constraint for selecting the attack-sensitive parameters. For example, it may be desired that the product concentration is within a certain range to ensure that the product is within specification. The attack-sensitive parameters may be selected to ensure that the potential values of the concentration in the corresponding minimum invariant set are within the acceptable range.

Remark 3. An attacker with prior knowledge of the active detection methodology may attempt to evade detection by using a destabilizing attack to target the process during the transient period after the control system switches from nominal parameters to attack-sensitive parameters. Randomly selecting the switching time (t_s) to minimize the possibility that the attacker knows when the controller switch occurs may be helpful in preventing the success of such attacks.

Remark 4. Detection of attacks that are potentially detectable under the nominal parameters is possible. In such cases, the attack identification and mitigation strategies could be activated following the detection of an attack while the closed-loop process is operated with the nominal parameters, and switching to the attack-sensitive parameters may not be needed.

Remark 5. Zero false alarms resulting from a parameter switch may be guaranteed under a special case when the minimum invariant set of the attack-free process under the updated parameters is a subset of the minimum invariant set of the process under the parameters used prior to the switch. For example, if the minimum invariant set for the attack-free process with nominal parameters is contained within the minimum invariant set for the process with attack-sensitive parameters (i.e., $D_\xi(I, K^*, L^*) \subset D_\xi(I, K_\Lambda, L_\Lambda)$), then, for the switch from nominal to attack-sensitive parameters, the augmented state of the attack-free process is contained within the minimum invariant set with attack-sensitive parameters, i.e., $\xi(t_s) \in D_\xi(I, K_\Lambda, L_\Lambda)$. As a result, there will be no transients, and zero false alarms can be guaranteed. However, zero false alarms cannot be guaranteed when a switch from the attack-sensitive parameters to the nominal parameters takes place because at the switching instance the augmented state may be outside the minimum invariant set associated with the nominal parameters, i.e., $\xi(t_s + T_c) \in D_\xi(I, K_\Lambda, L_\Lambda) \setminus D_\xi(I, K^*, L^*)$. As a result, the second switch may generate false alarms by the detection scheme, and the detection scheme may need some modification to minimize false alarms.

Remark 6. The detectability of an attack is defined based on the ability of the residual-based detection scheme to detect the attack in finite time. It is a system property and is not influenced by the control parameter switching instance t_s or the cycle time T_c. Under attack-free operation, both parameters (t_s and T_c) do not influence closed-loop stability. The switching instance t_s also does not influence closed-loop stability of the attacked process with either set of control system parameters, i.e., with nominal parameters or with attack-sensitive parameters.

Remark 7. Attacks on industrial control systems may take several forms. To characterize different types of attacks, the taxonomy of attacks on ICSs has been analyzed and presented in the literature [17–21]. In the present work, the active detection methodology is designed to enhance the detection capabilities of a residual-based passive detection scheme, monitoring the process for multiplicative sensor–controller link cyberattacks. Multiplicative sensor–controller link attacks multiply the data communicated over the sensor–controller communication channels by a factor. Under a multiplicative attack, the real-time process operational data communicated over the communication channels is masked by the attack. Replay attacks, another type of false-data injection attacks, communicate historic attack-free process operational data over the compromised controller communication channels and are fundamentally different from multiplicative attacks. Under a replay attack, the data communicated over the compromised controller communication channels has no correlation to the real-time process operational data. Characterization of the detection capability of the proposed active detection methodology to detect other types of cyberattacks is beyond the scope of the present work.

Remark 8. The proposed active detection methodology considers a single switch from nominal mode (during which the process is operated with nominal parameters) to the attack-sensitive mode (during which the process is operated with attack-sensitive parameters) and back to operating in the nominal mode thereafter. To further enhance the detection capabilities of the passive residual-based detection scheme with respect to a wider range of attack magnitudes, the proposed methodology may be modified to include control system switches between the nominal mode and multiple attack-sensitive modes. The closed-loop

Table 1
Parameters for the CSTR [45].

Volumetric flow rate (F)	$5.0\,\text{m}^3\,\text{h}^{-1}$
Reactor volume (V)	$1.0\,\text{m}^3$
Feed concentration of A (C_{A0})	$4.0\,\text{kmol}\,\text{m}^{-3}$
Activation energy (E)	$5.0 \times 10^4\,\text{kJ}\,\text{kmol}^{-1}$
Pre-exponential factor (k_0)	$8.46 \times 10^6\,\text{m}^3\,\text{h}^{-1}\,\text{kmol}^{-1}$
Gas constant (R)	$8.314\,\text{kJ}\,\text{kmol}^{-1}\,\text{K}$
Feed temperature (T_0)	$300\,\text{K}$
Density of reactor liquid hold-up (ρ)	$1000\,\text{kg}\,\text{m}^{-3}$
Heat of reaction (ΔH)	$-1.15 \times 10^4\,\text{kJ}\,\text{kmol}^{-1}$
Heat capacity (C_p)	$0.231\,\text{kJ}\,\text{kg}\,\text{K}^{-1}$
Steady-state heat rate added/removed from the reactor (Q_s)	$0\,\text{kJ}\,\text{h}^{-1}$
Steady-state reactant concentration (C_{As})	$1.22\,\text{kmol}\,\text{m}^3$
Steady-state temperature (T_s)	$438.2\,\text{K}$

process with the active detection methodology using multiple control system switches may be considered a switched system. Successive switching between different modes may compromise closed-loop stability of the process. Here, closed-loop stability for the switched system under bounded process disturbances and measurement noise means ultimate boundedness of the process state (and estimation error) in a small neighborhood of the origin. No guarantees can be made on the closed-loop stability of the attacked process without placing limitations on the attack magnitude. To guarantee the closed-loop stability of the attack-free process, two classical approaches may be employed (e.g., [34]). In the first approach, a common Lyapunov function may be used to find the control parameters for the nominal and attack-sensitive modes. The advantage of this approach is that the Lyapunov function value will decrease over time under any mode if the state is sufficiently far from the origin. However, this approach restricts the choice of control parameters. As an alternative approach, a Lyapunov function may be derived for each mode, i.e., the multiple Lyapunov function approach. In this case, the switching times must be carefully selected because the Lyapunov function value of the inactive modes may increase over time when another mode is active. Nonetheless, existing methods for determining the switching times could be employed (see, for example, [34]).

4. Application to a chemical process

A chemical process consisting of a CSTR is considered where a second-order, exothermic reaction of the form $A \rightarrow B$ occurs. The CSTR contents are assumed to be well-mixed, and the contents may be heated or cooled using, for example, a cooling jacket or submerged heat exchanger coil. A dynamic process model is obtained from mass and energy balances under standard modeling assumptions, and is given by the following system of ordinary differential equations:

$$\begin{aligned}\frac{dC_A}{dt} &= \frac{F}{V}(C_{A0} + \Delta C_{A_0} - C_A) - k_0 e^{\frac{-E}{RT}} C_A^2 \\ \frac{dT}{dt} &= \frac{F}{V}(T_0 + \Delta T_0 - T) - \frac{\Delta H k_0}{\rho C_p} e^{\frac{-E}{RT}} C_A^2 + \frac{Q}{\rho C_p V}\end{aligned} \tag{12}$$

where C_{A0} is the feedstock reactant concentration, T_0 is the feedstock temperature, C_A is the reactor reactant concentration, T is the reactor temperature, and Q is the heat added to or removed from the tank contents. The variables ΔC_{A_0} and ΔT_0 represent two bounded process disturbances, modeled as a deviation from the nominal feedstock reactant concentration and temperature. The process parameter definitions and their values are given in Table 1.

The control objective is to operate the process around its open-loop stable steady-state with $C_A = C_{As}$ and $T = T_s$ where the values are given in Table 1. The measured outputs are C_A and T, and the manipulated input is Q. Defining deviation variables, the state, input, process disturbance, and output are given by: $x = [x_1\ x_2]^T = [C_A - C_{As}\ T - T_s]^T$, $u = Q - Q_s$, $w = [\Delta C_{A0}\ \Delta T_0]^T$, and $y = [x_1 - x_{1s}\ x_2 - x_{2s}]^T$. The sensors measuring C_A and T are corrupted by bounded measurement noise.

To design a linear feedback control law for the CSTR, the nonlinear process model in Eq. (12) is linearized about its steady-state, yielding a continuous-time linear process model. The continuous-time linear model is discretized in time with a sampling interval of 10^{-2} h by assuming a zeroth-order hold of the inputs to obtain a discrete-time linear process model of the form in Eq. (1). The system matrices are given by:

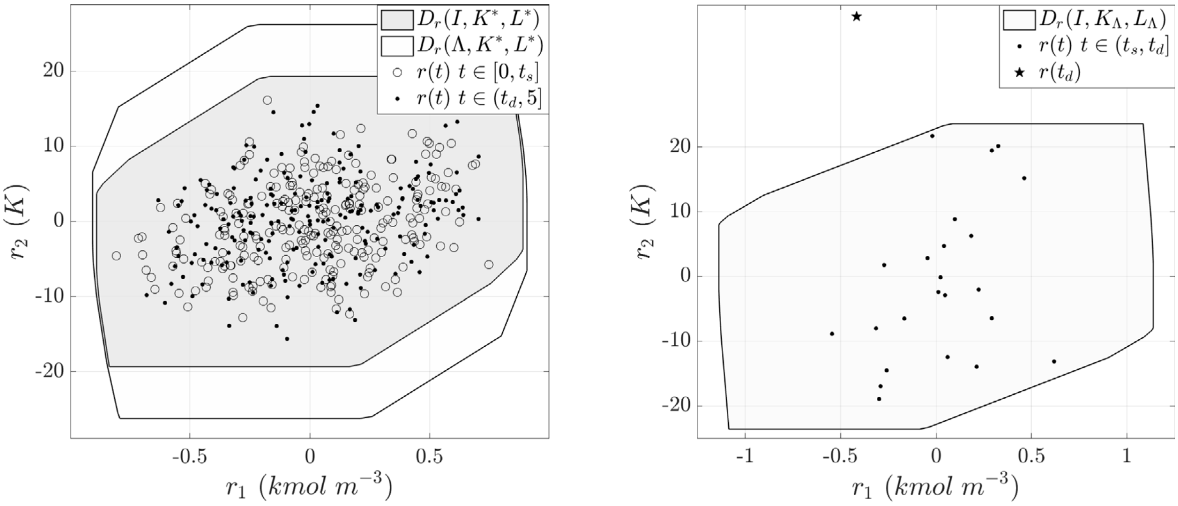
$$A = \begin{bmatrix} 0.7364 & -0.0041 \\ 10.6953 & 1.1560 \end{bmatrix}, \quad B = \begin{bmatrix} -9.0708 \times 10^{-8} \\ 4.6741 \times 10^{-5} \end{bmatrix}, \\ G = \begin{bmatrix} 0.0433 & -0.0001 \\ 0.2724 & 0.0540 \end{bmatrix} \tag{13}$$

The discrete-time linear model for the CSTR of the form in Eq. (1) with matrices in Eq. (13) is referred to as the linearized CSTR model.

In the simulations presented in subsequent sections, the Multi-Parametric Toolbox (MPT) 3.0 [46] is used for the calculation of the minimum invariant and residual terminal sets. Numerical approximations of the minimum invariant sets are computed based on the algorithm in [47] with an error bound of 5×10^{-5}. In comparing the numerical estimates of the terminal residual sets, the technique presented in [32] is used. In the remainder, time is represented in continuous time with a slight abuse of notation. In Section 4.1, the application of the active detection methodology for enhancing the detection capabilities of the residual-based detection scheme is demonstrated using the linearized CSTR model. In Section 4.2, the active detection methodology is applied to the nonlinear CSTR to evaluate the efficacy of the proposed approach when dealing with more complex process dynamics.

4.1. Application of the active detection methodology to the linearized CSTR

In this section, the CSTR is modeled using the linearized process model. The disturbance set (F) is described by an admissible process disturbance set (W) given by $\Delta C_{A0} \in [-0.5, 0.5]$ $\text{kmol}\,\text{m}^{-3}$ and $\Delta T_0 \in [-5, 5]$ K, and an admissible measurement noise set (V) described by $[-0.5, 0.5]$ $\text{kmol}\,\text{m}^{-3}$ and $[-5, 5]$ K for the concentration and temperature sensors, respectively. The control actions are computed using a linear control law of the form in Eq. (4) using estimates generated by a Luenberger observer of the form in Eq. (3). Pole placement is used to determine the controller and observer gains. The nominal parameters (K^*, L^*) are chosen to stabilize the attack-free process with the controller gain computed with poles placed at $[0.2\ {-0.1}]$, and the observer gain with the poles placed at $[0.2\ 0.3]$. For the attack-sensitive parameters (K_Λ, L_Λ), the controller gain is computed with poles placed at $[-0.33\ \ -0.3]$ and the observer gain is



(a)

(b)

Fig. 5. The residual values of the attacked linearized CSTR process with (a) with nominal parameters before and after a switch to the attack-sensitive parameters and (b) with attack-sensitive parameters where the attack is detected at time $t_d = 2.74$ h.

computed with poles placed at $[-0.2\ -0.3]$. In the absence of an attack, the closed-loop process with attack-sensitive parameters is stable in the sense that $\max_i |\lambda_i(A_\xi(I, K_\Lambda, L_\Lambda))| = 0.33 < 1$. Furthermore, the attack-sensitive parameters are found to be "sensitive" to multiplicative sensor–controller link attacks with magnitudes in the set $\{\Lambda \mid \text{diag}(1, \alpha) \mid \alpha \in [0.6, 0.95]\}$. This range is numerically verified by parameterizing the attack magnitude with a parameter α where $\Lambda = \text{diag}(1, \alpha)$. The value of α is varied beginning at 0.6 and incremented by 0.01 until a value of $\alpha = 0.95$ is reached. For each Λ, the control system parameters are sensitive to the attack if any of the eigenvalues of $A_\xi(\Lambda, K_\Lambda, L_\Lambda)$ are outside the unit circle and the observability matrix for the pair $(A_\xi(\Lambda, K_\Lambda, L_\Lambda), A_r(\Lambda))$ is full rank. A similar analysis is performed using the nominal parameters. The nominal parameters are not sensitive to any attack in the set $\{\Lambda \mid \text{diag}(1, \alpha) \mid \alpha \in [0.6, 0.95]\}$.

Three sets of simulations are performed, and the results are compared. First, the closed-loop process with nominal parameters and without the active detection methodology (without switching) is considered. Second, the active detection methodology is applied to the attacked closed-loop process. The first and second simulation sets are used to evaluate the enhanced detection capabilities of the proposed active detection methodology. Third, the active detection methodology is applied to the attack-free process to identify if false alarms are raised resulting from control system parameter switching.

Each simulation set consists of 1000 simulations of the closed-loop process. The process disturbances and measurement noise are modeled as random variables drawn from a uniform distribution on the interval defined by the bounds of the appropriate admissible set. The value of the random variables modeling the process disturbances and measurement noise are varied every sample time, and different realizations of the random variables are used in each simulation. The same realizations of random variables are used across simulation sets to compare the results across simulation sets. For each simulation, the process states are initialized at 0, and a period of 5 h is simulated.

For simulating the attacked process, an attack magnitude of $\Lambda = \text{diag}(1, 0.9)$ is considered. Under the nominal parameters, the attack is potentially detectable, which can be observed from Fig. 5(a) since $D_r(\Lambda, K^*, L^*) \not\subseteq D_r(I, K^*, L^*)$. Under the attack-sensitive parameters, the attack is detectable, and the terminal residual set for the attack-free process under the attack-sensitive parameters is shown in Fig. 5(b). To demonstrate the enhancement of attack detection capabilities of the residual-based detection scheme, the proposed active detection methodology is applied and the control system switches to attack-sensitive parameters at time $t_s = 2.5$ h. A cycle time of $T_c = 1$ h is used, i.e., in the absence of attack detection, a second switch from attack-sensitive to nominal parameters is implemented at time $t_s + T_c = 3.5$ h.

The residual values for one of the simulations from the first set (with the active detection methodology) are depicted in Fig. 5. From Fig. 5(a), the residual values of the closed-loop process with nominal parameters are in the attack-free terminal residual set before the switch occurs, i.e., $r(t) \in D_r(I, K^*, L^*)$, $t \in [0, 2.5]$ h. As a result, no alarms are raised by the detection scheme, and the attack is not detected during this period. After the switch to attack-sensitive parameters, the attack is detected at time $t_d = 2.74$ h because the residual value is outside the terminal residual set, i.e., $r(t_d) \notin D_r(I, K_\Lambda, L_\Lambda)$ (Fig. 5(b)). Following the detection of the attack, the control system switches back to nominal parameters to stabilize the closed-loop process. In practice, attack identification and mitigation measures would be activated after detection. After switching back to the nominal parameters, no alarms are raised (Fig. 5(a)).

The results obtained from the first and second simulation sets are compared. For the second simulation set (passive detection), the attack is detected in 43 out of 1000 simulations. In 19 of these 43 simulations, the attack is detected before the switching instance ($t_s = 2.5$ h). For the first and second simulation sets, the process evolves the same during the period $t = 0$ h to $t = 2.5$ h because the same control system, attack, disturbance, measurement noise, and detection scheme are applied to the process during this period. For the 19 simulations, switching to attack-sensitive parameters is not needed as the attack is detected before the switch. For the remaining 981 simulations, the attack is detected after switching to the attack-sensitive parameters within a maximum of 24 sample times in all simulations. Considering the 981 remaining simulations with passive detection (second simulation set), the attack is detected in 33 sample times after $t_s = 2.5$ h in the best case and never detected over the simulated 5 h operating period in the worst case. The results demonstrate an enhancement of detection capabilities of the residual-based detection scheme by applying the active detection methodology.

In the third simulation set, the false alarm rate under the proposed active detection methodology is evaluated. The attack-free process is considered for the analysis. False alarms are not raised after switching into and out of the attack-sensitive parameters in any simulation. Further analysis is performed to address the possibility of false alarms. The containment of the augmented state at the switching instances in the minimum invariant set is verified. When switching to the attack-sensitive parameters, the state is verified to be in the minimum invariant set associated with the attack-sensitive parameters, i.e., $\xi(t_s) \in D_\xi(I, K_\Lambda, L_\Lambda)$. When switching back to the nominal parameters, the state is verified to be in the minimum invariant set associated with the nominal parameters, i.e., $\xi(t_s + T_c) \in D_\xi(I, K^*, L^*)$. When the control system switches from nominal parameters to attack-sensitive parameters, the augmented state is in the minimum invariant set of the attack-free process with attack-sensitive parameters in all simulations. When the control system switches from attack-sensitive parameters to nominal parameters, the augmented state is not contained within the minimum invariant set of the attack-free process with nominal parameters in 958 out of the 1000 simulations. In the 958 simulations, the augmented state evolves briefly outside the minimum invariant set of the attack-free closed-loop process with nominal parameters. The augmented state converges to the minimum invariant set within two sample times. No false alarms are observed in any of these cases. Although this analysis confirms the possibility of a false alarm, false alarms are not raised in these cases.

4.2. Application of the active detection methodology to the nonlinear CSTR

In this section, we apply the active detection methodology to the nonlinear CSTR process model in Eq. (12). To this end, the state is maintained within a region around the origin when the process disturbances and measurement noise are small. In this region, the nonlinear process may be approximated by its linearized model. As the magnitude of the disturbances and measurement noise increases, the impact of the nonlinearities increases. While the theoretical results in this work are developed strictly for linear systems, the objective of this study is to assess the method's applicability to the nonlinear case. The proposed active detection method is therefore applied to the nonlinear process, considering small disturbances. The disturbance set F is described by an admissible process disturbance set (W) given by $\Delta C_{A0} \in [-0.01, 0.01]$ kmol m^{-3} and $\Delta T_0 \in [-0.2, 0.2]$ K, and an admissible measurement noise set (V) described by $[-0.01, 0.01]$ kmol m^{-3} and $[-0.2, 0.2]$ K for the concentration and temperature sensors, respectively. The process disturbances and measurement noise are modeled as random variables drawn from a uniform distribution in the interval specified by the bounds of the admissible set. The same realizations of random variables are used across simulation sets.

The closed-loop simulations of the continuous-time CSTR process use the explicit Euler's method with a step size of 1×10^{-4} h to integrate the ordinary differential equations in Eq. (12). Extensive simulations are employed to verify that further reduction in the integration time step did not lead to substantial changes in the computed solution of the nonlinear ordinary differential equations. To steer the process states to the origin, a linear control law (Eq. (4)) is used with state estimates generated by a Luenberger observer (Eq. (3)) based on the linearized process model using the matrices A, B, and G (Eq. (13)). The sampling period of the control system is 10^{-2} h, which is the same as that used in Section 4.1. In general, the sampling period should be sufficiently small so that the continuous-time process in Eq. (12) may be stabilized with the discrete-time controller in Eq. (4). Two sets of simulations, each consisting of 1000 simulations of the attack-free closed-loop process with nominal parameters and the attack-free closed-loop process with attack-sensitive parameters are performed. In all simulations, the attack-free closed-loop process is found to be stable, verifying that the sampling period is appropriately chosen.

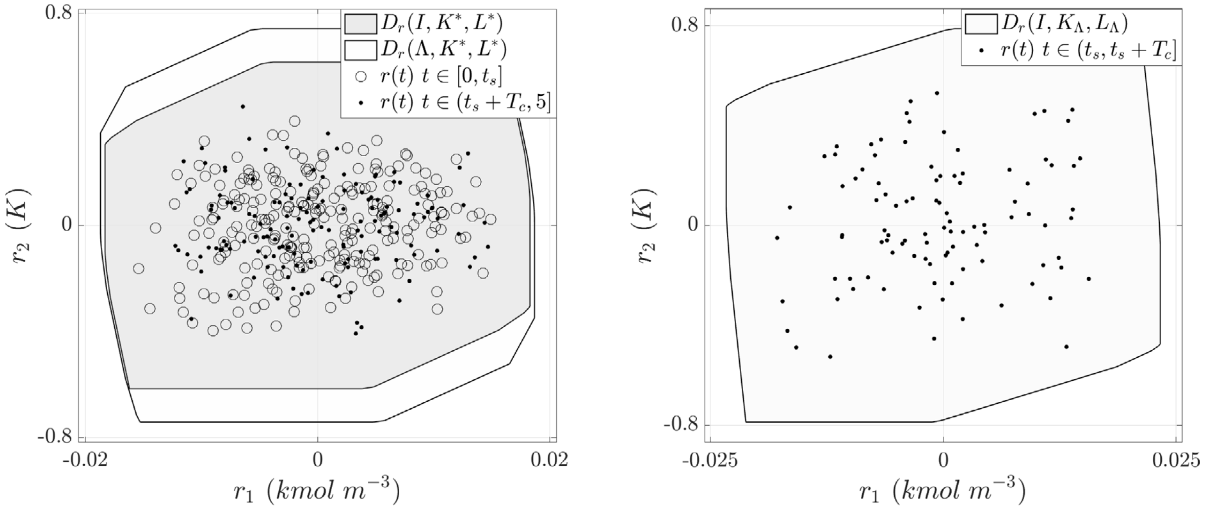
Numerical approximations of the attack-free terminal residual sets for the closed-loop process with nominal parameters and with the attack-sensitive parameters are computed from the linearized CSTR model. To verify that the terminal residual set approximated from the linear model is a suitable approximation for the nonlinear process, the evolution of the residuals are considered under the nominal and attack-sensitive parameters. Considering the same two sets of simulations used for verifying closed-loop stability, the residuals of the attack-free process are bounded within the appropriate terminal residual set in all simulations. Based on this, the computed terminal residual sets are suitable approximations for the nonlinear process.

A similar study as that performed in Section 4.1 consisting of three simulations sets is carried out for the nonlinear process. In each simulation set, 1000 simulations are conducted. These simulations enable the evaluation of the detection capabilities and potential of false alarms under the proposed active detection methodology for the nonlinear process. For the first two simulation sets, an attack of magnitude $\Lambda = \text{diag}(1, 0.9)$ is considered. Based on the linearized model, the attack-sensitive parameters are "sensitive" to this attack, while the nominal parameters are not. For the third simulation set, attack-free operation is considered. In the first simulation set (with active detection methodology), the control system switches from nominal parameters to attack-sensitive parameters at time instance $t_s = 2.5$ h. The cycle time under the attack-sensitive parameters is $T_c = 1$ h.

The residual values from one simulation in the first simulation set (with active detection methodology) are depicted in Fig. 6. From Fig. 6(b), the attack is detected at time $t_d = 2.69$ h after the control system switches from nominal parameters to attack-sensitive. Following the detection of the attack, the control system switches back to nominal parameters to stabilize the closed-loop process. From Fig. 6(a), the residual is outside the attack-free terminal residual set for one sample time after switching back to nominal parameters. As a result, another alarm is raised. Thereafter, no alarms are raised because the residual converges to the terminal residual set.

For the 1000 simulations of the attacked process under nominal parameters monitored by the residual-based detection scheme in Eq. (10) (passive only detection), the attack is detected in only one simulation. In the active detection simulation set, the attack is detected after the control system switches from nominal parameters to attack-sensitive parameters in all cases. After switching to attack-sensitive parameters, the attack is detected within a minimum of 2 sample times and a maximum of 19 sample times after the switching instance. Thus, the active detection methodology also enhances the detection capabilities of the residual-based detection scheme for the nonlinear process.

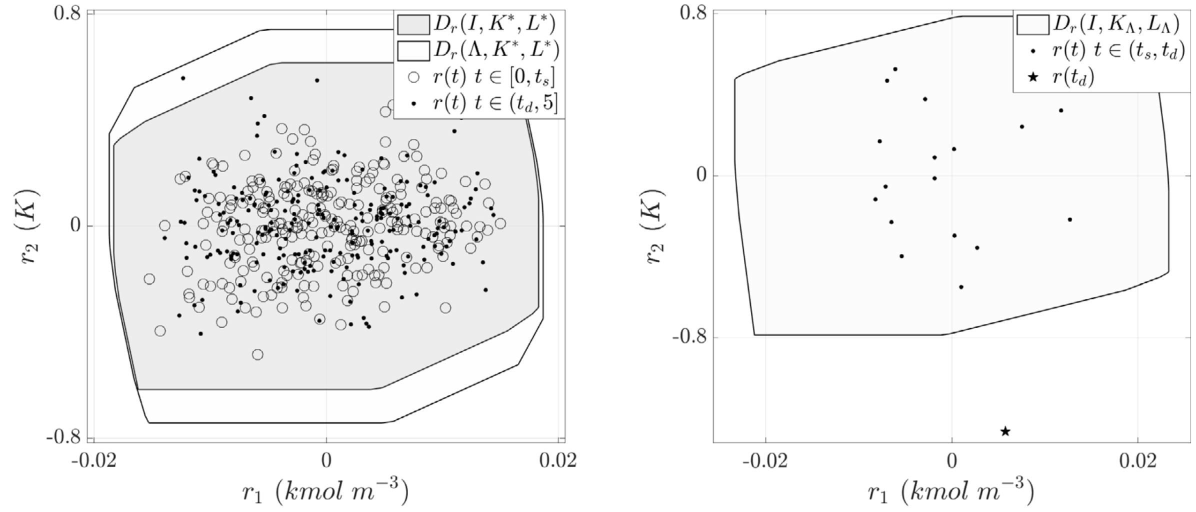
In the third simulation set, 1000 simulations of the attack-free process with the active detection methodology are performed to analyze the false alarm rate. No false alarms are raised in any of the simulations. Similar to the analysis for the third simulation set in Section 4.1, the containment of the augmented state at the switching instance within the minimum invariant set with the updated parameters is verified. At the switching instance t_s, the state is always contained in the minimum invariant set $D_\xi(\Lambda, K^*, L^*)$ in all cases. At the switching instance $t_s + T_c$, the augmented state is not within the minimum invariant set under nominal parameters in 921 out of 1000 simulations. Over these 921 simulations, the augmented state evolves outside the



(a)

(b)

Fig. 6. (a) The residual values of the attacked nonlinear CSTR process with nominal parameters before and after a switch to the attack-sensitive parameters. (b) The detection of the attack at the time $t_d = 2.69$ h after a switch from nominal parameters to attack-sensitive parameters.



(a)

(b)

Fig. 7. (a) The residual values of the attack-free nonlinear CSTR process with nominal parameters before and after a switch to the attack-sensitive parameters. (b) The residual values of the attack-free nonlinear CSTR process with attack-sensitive parameters.

minimum invariant set, but converges to it in 2 sample times. However, no false alarms are observed in any of the 921 simulations. Fig. 7 illustrates the residual values for the attack-free process with the active detection methodology over one simulation. Over this simulation, no false alarms are observed when the control system switches from the nominal mode into the attack-sensitive mode (Fig. 7(a)). Similarly, no false alarms are observed after the control system switches back to the nominal mode (Fig. 7(b)).

4.2.1. Comparison between active and passive detection

The application of active detection methodology to enhance the detection capabilities of the CUSUM detection scheme, another residual-based detection scheme, is demonstrated. The CUSUM detection scheme is a statistical change detection scheme that monitors a process based on the deviation of the detection metric from a predefined baseline value. Application of the CUSUM detection scheme using the residual vector as the detection scheme has been considered as a passive attack detection scheme previously in the literature [22,26,32]. The CUSUM detection scheme monitoring a process based on the 2-norm of the residual may be represented by:

$$S(t) = \max\{S(t-1) + \|r(t)\| - b, 0\};\ \ S(-1) = 0; \tag{14}$$

where $S(t)$ is the CUSUM statistic, which is the detection scheme output, $r(t)$ is the residual of the process at the time $t \geq 0$, and b is the baseline parameter. An attack on the process is detected by the scheme if the CUSUM statistic exceeds the tolerance value, which is the alarm threshold τ, i.e.,

$$S(t) \leq \tau;\ \ \text{No Attack}$$
$$S(t) > \tau;\ \ \text{Attack}$$

The CUSUM detection scheme is chosen as the detection scheme in place of the set membership-based detection scheme considered earlier to monitor the CSTR. An attack is considered in the temperature sensor–controller link and has the same magnitude as that previously considered, i.e., $\Lambda = \text{diag}(1, 0.9)$.

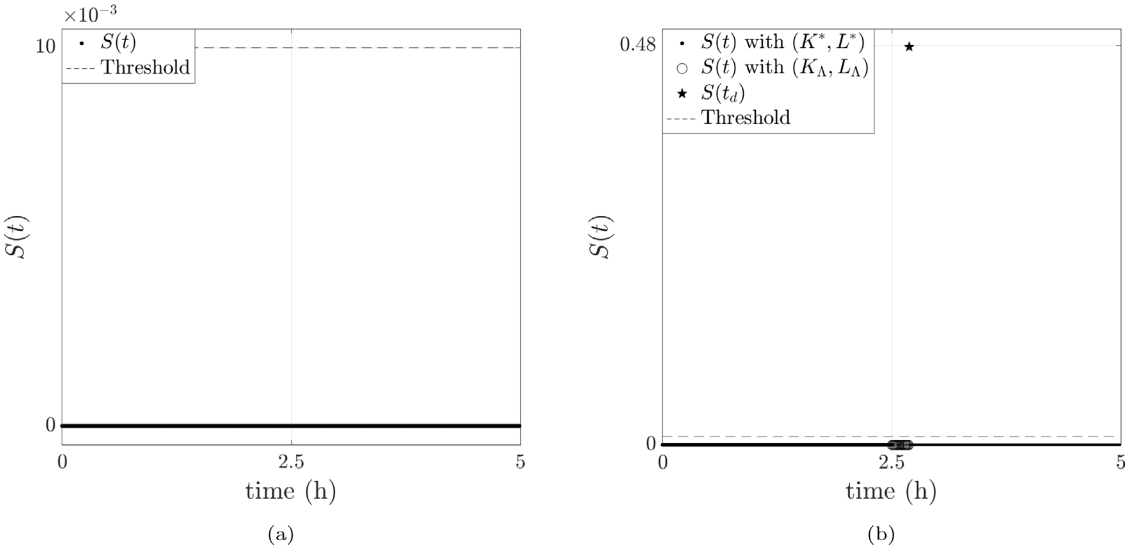


Fig. 8. (a) The CUSUM statistic for the attacked CSTR without the active detection methodology implemented. (b) The CUSUM statistic for the attacked CSTR process with the active detection methodology implemented, showing that the attack is detected at time $t = 2.69$ h.

To tune the CUSUM detection scheme for a zero false alarm rate in the absence of an attack (and without switching) when monitoring the closed-loop process, the approach presented in [32] is adopted. Because the residuals of the attack-free closed-loop process are always contained within the terminal residual set, they are also contained within the 2-norm ball enclosing the terminal residual set. Consequently, the norm of the residual vector of the attack-free process is always less than the radius of the ball. The baseline parameter is selected as the radius of the 2-norm ball enclosing the terminal residual set of the attack-free process. With the nominal parameters, the radius is $R_{D_r(I,K^*,L^*)} = 0.6169$, and with the attack-sensitive parameters, the radius is $R_{D_r(I,K_\Lambda,L_\Lambda)} = 0.7884$. Based on Eq. (14), the CUSUM statistic for the attack-free process always remains at zero with this choice of the baseline parameter, and any non-zero CUSUM statistic value may be considered indicative of an attack. In this case, the CUSUM detection may be tuned with an alarm threshold choice of $\tau = 0$. To maintain a zero false alarm rate when there are small variations in the process that are not necessarily due to an attack, the alarm threshold for the detection scheme is set at $\tau = 0.01$. Furthermore, the CUSUM detection scheme is implemented so that upon detection of an attack, the CUSUM statistic is reset to 0 at the next time step, i.e., $S(t) > \tau$ implies $S(t+1) = 0$.

To enhance the attack detection capability of the CUSUM detection scheme, the active detection methodology is implemented. Because the parameter b is dependent on the terminal residual set, which depends on the control system parameters, the baseline parameter switches from the value with the nominal parameters to the value corresponding to the attack-sensitive parameters when operating with the attack-sensitive parameters:

$$b(t) = \begin{cases} R_{D_r(I,K_\Lambda,L_\Lambda)} = 0.7884; & t \in (t_s, t_s + T_c] \\ R_{D_r(I,K^*,L^*)} = 0.6169; & \text{Otherwise} \end{cases}$$

In this case, the control system switches back to using the nominal parameters if an attack is detected during operation with the attack-sensitive parameters.

Two sets of simulations are performed for the process monitored by the CUSUM detection scheme in Eq. (14). First, the attacked closed-loop process with nominal parameters (no parameter switching) and monitored by the CUSUM detection scheme is considered. Next, the attacked closed-loop process with the active detection methodology and monitored by the CUSUM detection scheme is considered. The attack is not detected in any of the simulations using passive detection (with no switching). Fig. 8(a) illustrates the output of the CUSUM scheme for one simulation with passive detection. With the active detection methodology, however, the attack is detected in all cases after the control system switches from nominal parameters to attack-sensitive parameters. The attack is detected in a minimum of 2 sample times and a maximum of 19 sample times after switching.

The residual values from one simulation with the active detection methodology are shown in Fig. 9. Before the switch to attack-sensitive parameters occurs, the residual values for the attacked closed-loop process with nominal parameters are in the 2-norm ball enclosing the attack-free terminal residual set, i.e., $r(t) \in B^2(R_{D_r(I,K^*,L^*)})$ for all $t \in [0, 2.5]$ h (Fig. 9(a)). As a result, no alarms are raised by the detection scheme. After the switch to attack-sensitive parameters, the attack is detected at time $t_d = 2.69$ h because the residual value is outside the 2-norm ball enclosing the terminal residual set (Fig. 9(b)). From Fig. 8(b), an alarm is raised by the detection scheme at the detection time. Following the detection of the attack, the control system switches back to nominal parameters to stabilize the closed-loop process. From Fig. 9(a), no alarms are raised by the detection scheme thereafter.

Remark 9. The CUSUM detection scheme is a dynamic detection scheme measuring the cumulative deviation of the 2-norm of the residual from the baseline parameter over time. While not observed in the simulations presented in this section, in some cases, the CUSUM detection scheme may not detect an attack immediately after the residual of the closed-loop process leaves the 2-norm ball enclosing its attack-free terminal residual set. However, the set membership-based detection scheme in Eq. (10) detects an attack as soon as the residual leaves the terminal residual set of the attack-free process. Based on this, it may appear that the CUSUM detection scheme is not as sensitive to the drifts in the detection parameter, as the set membership-based detection scheme. However, the sensitivity of the CUSUM detection scheme to the drifts in the detection metric is dependent on its tuning parameters (i.e., on the threshold τ and the parameter b). The tuning approach is fundamentally different from the set membership-based detection scheme. Consequently, the detection performance of the CUSUM detection scheme with a given choice of τ and b may not be directly compared with that of the set membership-based detection scheme.

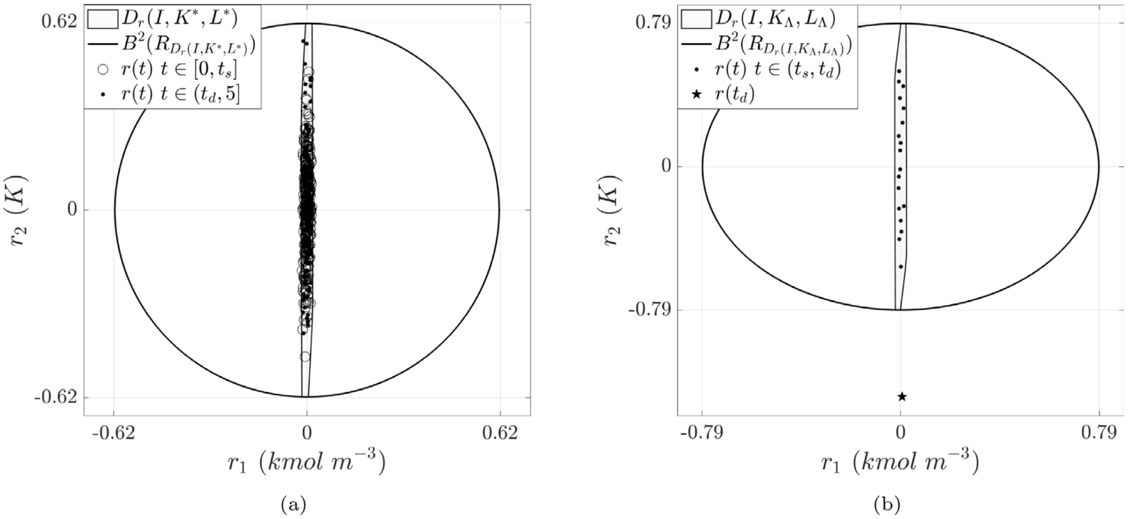


Fig. 9. (a) The residual values of the attacked nonlinear CSTR with nominal parameters before and after a switch to attack-sensitive parameters. (b) The residual values of the attacked nonlinear CSTR with attack-sensitive parameters showing attack detection at time $t_d = 2.69$ h.

5. Conclusions

In this work, an active attack detection methodology that enhances the attack detection capabilities of residual-based detection schemes was developed. The methodology utilizes control system parameter switching to probe for, and elicit detection of, multiplicative sensor–controller attacks. In this approach, the control system switches occasionally between nominal control system parameters, selected on the basis of standard control design criteria, and attack-sensitive control system parameters to manage the potential trade-off between closed-loop performance and attack detectability. The relationship between attack detectability with respect to a residual-based detection scheme, the control system parameters, and closed-loop stability was rigorously analyzed and the selection of the attack-sensitive parameters exploited this relationship. The enhancement of attack detection capabilities of two residual-based detection schemes upon application of the active attack detection methodology was demonstrated using a chemical process example. Future work will investigate extensions of the theoretical analysis to nonlinear process systems, and the development of attack identification and mitigation strategies.

CRediT authorship contribution statement

Shilpa Narasimhan: Conceptualization, Methodology, Software, Formal Analysis, Writing, Visualization. **Nael H. El-Farra:** Conceptualization, Methodology, Formal Analysis, Writing, Supervision, Project administration. **Matthew J. Ellis:** Conceptualization, Methodology, Formal Analysis, Writing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Financial support from the College of Engineering at University of California, Davis is gratefully acknowledged.

Appendix. Unstable systems with bounded inputs

For the detectability result, a result on unstable systems with bounded inputs is needed.

Theorem 1. *Consider the system*

$$\begin{aligned} z(t+1) &= A_z z(t) + B_\nu \nu(t) \\ \eta(t) &= C_z z(t) + D_\nu \nu(t) \end{aligned} \tag{A.1}$$

with $z(t) \in \mathbb{R}^n$, $\eta(t) \in \mathbb{R}^{n_\eta}$, and $\nu(t) \in \Gamma \subset \mathbb{R}^{n_\nu}$ for all time $t \geq 0$, where Γ is a compact set. If the pair (A_z, C_z) is observable, and $\|z(t)\| \to \infty$ as $t \to \infty$, then $\|\eta(t)\| \to \infty$ as $t \to \infty$.

Proof. Defining $\eta_n(t)$ and $\nu_n(t)$ as:

$$\eta_n(t) := \begin{bmatrix} \eta(t) \\ \vdots \\ \eta(t+n-1) \end{bmatrix}, \quad \nu_n(t) := \begin{bmatrix} \nu(t) \\ \vdots \\ \nu(t+n-1) \end{bmatrix} \tag{A.2}$$

If the pair (A_z, C_z) is observable, the observability matrix has rank n_z. Provided $\eta_n(t)$ and $\nu_n(t)$, $z(t)$ is the unique solution to the following system of equations if (A_z, C_z) is observable:

$$\begin{aligned} \eta_n(t) &= \underbrace{\begin{bmatrix} C_z \\ C_z A_z \\ \vdots \\ C_z A_z^{n-1} \end{bmatrix}}_{=:\mathcal{O}_n} z(t) + \underbrace{\begin{bmatrix} D_\nu & & & \\ C_z B_\nu & D_\nu & & \\ \vdots & \ddots & \ddots & \\ C_z A_z^{n-2} B_\nu & \cdots & C_z B_\nu & D_\nu \end{bmatrix}}_{=:\mathcal{B}_n} \\ \nu_n(t) &= \mathcal{O}_n z(t) + \mathcal{B}_n \nu_n(t) \end{aligned} \tag{A.3}$$

where $\mathcal{O}_n$ is the observability matrix.

Since the pair (A_z, C_z) is observable, $\mathcal{O}_n$ has full column rank and $\mathcal{O}_n^T \mathcal{O}_n$ is a positive definite matrix. Thus, $\|z\|_{\mathcal{O}_n^T \mathcal{O}_n} := \sqrt{z^T \mathcal{O}_n^T \mathcal{O}_n z}$ is a weighted Euclidean norm. Owing to the equivalence of norms, there exists $c > 0$ such that $\|z(t)\|_{\mathcal{O}_n^T \mathcal{O}_n} \geq c\|z(t)\|$. From Eq. (A.3), the equivalence of norms, and the triangle inequality,

$$\begin{aligned} c\|z(t)\| &\leq \|\mathcal{O}_n z(t)\| = \|\eta_n(t) - \mathcal{B}_n \nu_n(t)\| \\ &\leq \|\eta(t)\| + \|\eta(t+1)\| + \cdots + \|\eta(t+n-1)\| \\ &\quad + \|\mathcal{B}_n \nu_n(t)\| \end{aligned} \tag{A.4}$$

Since $\nu(t)$ is bounded for all $t \geq 0$, $\|\mathcal{B}_n \nu_n(t)\|$ is bounded. Because the last line of Eq. (A.4) is a sum over a finite number of terms, $\|\eta(t)\| \to \infty$ as $t \to \infty$ if $\|z(t)\| \to \infty$ as $t \to \infty$. □

References

[1] M. Abrams, J. Weiss, Malicious Control System Cyber Security Attack Case Study-MAroochy Water Services, Australia, Tech. Rep., 2008.

[2] K.E. Hemsley, R.E. Fisher, History of Industrial Control System Cyber Incidents, Tech. Rep. INL/CON-18-44111, Idaho National Lab. Idaho Falls, ID (United States), 2018.

[3] T. Miller, A. Staves, S. Maesschalck, M. Sturdee, B. Green, Looking back to look forward: Lessons learnt from cyber-attacks on industrial control systems, Int. J. Crit. Infrastruct. Prot. 35 (2021) 100464, http://dx.doi.org/10.1016/j.ijcip.2021.100464.

[4] A. Di Pinto, Y. Dragoni, A. Carcano, TRITON: The First ICS Cyber Attack on Safety Instrument Systems, Tech. Rep., Nozomi Networks, 2018.

[5] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams, A. Hahn, Guide to Industrial Control Systems (ICS) Security, Tech. Rep., NIST Special Publication 800-82 Revision 2, U.S. Department of Commerce, 2015, http://dx.doi.org/10.6028/NIST.SP.800-82r2.

[6] M.S. Chong, H. Sandberg, A.M. Teixeira, A tutorial introduction to security and privacy for cyber–physical systems, in: Proceedings of the 18th European Control Conference, Naples, Italy, 2019, pp. 968–978, http://dx.doi.org/10.23919/ECC.2019.8795652.

[7] US Department of Energy, Office of Energy Efficiency and Renewable Energy, Department of Energy announces $1.5 million to mitigate cybersecurity risks in manufacturing, 2021, https://www.energy.gov/eere/amo/articles/department-energy-announces-15-million-mitigate-cybersecurity-risks-manufacturing.

[8] C.G. Rieger, D.I. Gertman, M.A. Mcqueen, Resilient control systems: Next generation design research, in: Proceedings of the 2nd Conference on Human System Interactions, Catania, Italy, 2009, pp. 632–636, http://dx.doi.org/10.1109/HSI.2009.5091051.

[9] D. Wei, K. Ji, Resilient industrial control system (RICS): Concepts, formulation, metrics, and insights, in: Proceedings of the 3rd International Symposium on Resilient Control Systems, Idaho Falls, ID, 2010, pp. 15–22.

[10] L. Mili, Taxonomy of the characteristics of power system operating states, in: Proceedings of the 2nd NSF-RESIN Workshop, Tucson, AZ, 2011, pp. 1–13.

[11] J. Giraldo, D. Urbina, A. Cárdenas, J. Valente, M. Faisal, J. Ruths, N.O. Tippenhauer, H. Sandberg, R. Candell, A survey of physics-based attack detection in cyber–physical systems, ACM Comput. Surv. 51 (4) (2018) 1–36, http://dx.doi.org/10.1145/3203245.

[12] Y. Hu, A. Yang, H. Li, Y. Sun, L. Sun, A survey of intrusion detection on industrial control systems, Int. J. Distrib. Sens. Netw. 14 (8) (2018) 1550147718794615, http://dx.doi.org/10.1177/1550147718794615.

[13] D. Zhang, Q.G. Wang, G. Feng, Y. Shi, A.V. Vasilakos, A survey on attack detection, estimation and control of industrial cyber–physical systems, ISA Trans. (2021) 1–16, http://dx.doi.org/10.1016/j.isatra.2021.01.036.

[14] S. Hu, D. Yue, Z. Cheng, E. Tian, X. Xie, X. Chen, Co-design of dynamic event-triggered communication scheme and resilient observer-based control under aperiodic DoS attacks, IEEE Trans. Cybern. 51 (9) (2020) 4591–4601, http://dx.doi.org/10.1109/TCYB.2020.3001187.

[15] S. Schlor, M. Hertneck, S. Wildhagen, F. Allgöwer, Multi-party computation enables secure polynomial control based solely on secret-sharing, in: Proceedings of the 60th IEEE Conference on Decision and Control, Austin, Texas, USA, 2021, pp. 4882–4887, http://dx.doi.org/10.1109/CDC45484.2021.9683026.

[16] M. Krotofil, A.A. Cárdenas, Resilience of process control systems to cyber–physical attacks, in: Proceedings of the 18th Nordic Conference on Secure IT Systems, Ilulissat, Greenland, 2013, pp. 166–182, http://dx.doi.org/10.1007/978-3-642-41488-6_12.

[17] I.M. Chapman, S.P. Leblanc, A. Partington, Taxonomy of cyber attacks and simulation of their effects, in: Proceedings of the 2011 Military Modeling & Simulation Symposium, Society for Computer Simulation International, Boston, Massachusetts, 2011, pp. 73–80.

[18] B. Zhu, A. Joseph, S. Sastry, A taxonomy of cyber attacks on SCADA systems, in: Proceedings of the 2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing, Dalian, China, 2011, pp. 380–388, http://dx.doi.org/10.1109/iThings/CPSCom.2011.34.

[19] S. Kim, G. Heo, E. Zio, J. Shin, J. Song, Cyber attack taxonomy for digital environment in nuclear power plants, Nucl. Eng. Technol. 52 (5) (2020) 995–1001, http://dx.doi.org/10.1016/j.net.2019.11.001.

[20] Z. Drias, A. Serhrouchni, O. Vogel, Taxonomy of attacks on industrial control protocols, in: Proceedings of the International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS), Paris, France, 2015, pp. 1–6, http://dx.doi.org/10.1109/NOTERE.2015.7293513.

[21] H.T. Reda, A. Anwar, A. Mahmood, Comprehensive survey and taxonomies of false data injection attacks in smart grids: attack models, targets, and impacts, Renew. Sustain. Energy Rev. 163 (2022) 112423, http://dx.doi.org/10.1016/j.rser.2022.112423.

[22] C. Murguia, J. Ruths, CUSUM and chi-squared attack detection of compromised sensors, in: Proceedings of the IEEE Conference on Control Applications, Buenos Aires, Argentina, 2016, pp. 474–480, http://dx.doi.org/10.1109/CCA.2016.7587875.

[23] G. Na, Y. Eun, A multiplicative coordinated stealthy attack and its detection for cyber physical systems, in: Proceedings of the IEEE Conference on Control Technology and Applications, Copenhagen, Denmark, 2018, pp. 1698–1703, http://dx.doi.org/10.1109/CCTA.2018.8511631.

[24] K. Nieman, H.C. Oyama, M. Wegener, H. Durand, Predict the impact of cyberattacks on control systems, Chem. Eng. Progr. 116 (9) (2020) 52–57.

[25] Y. Hu, H. Li, H. Yang, Y. Sun, L. Sun, Z. Wang, Detecting stealthy attacks against industrial control systems based on residual skewness analysis, Eur. Assoc. Signal Process. J. Wirel. Commun. Netw. 2019 (1) (2019) 1–14, http://dx.doi.org/10.1186/s13638-019-1389-1.

[26] C. Murguia, J. Ruths, Characterization of a CUSUM model-based sensor attack detector, in: Proceedings of the IEEE 55th Conference on Decision and Control, Las Vegas, NV, USA, 2016, pp. 1303–1309, http://dx.doi.org/10.1109/CDC.2016.7798446.

[27] Z. Chu, J. Zhang, O. Kosut, L. Sankar, Unobservable false data injection attacks against PMUs: Feasible conditions and multiplicative attacks, in: Proceedings of the IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, Aalborg, Denmark, 2018, pp. 1–6, http://dx.doi.org/10.1109/SmartGridComm.2018.8587555.

[28] S. Weerakkody, O. Ozel, P. Griffioen, B. Sinopoli, Active detection for exposing intelligent attacks in control systems, in: Proceedings of the IEEE Conference on Control Technology and Applications, Hawai'i, USA, 2017, pp. 1306–1312, http://dx.doi.org/10.1109/CCTA.2017.8062639.

[29] M. Ghaderi, K. Gheitasi, W. Lucia, A blended active detection strategy for false data injection attacks in cyber–physical systems, IEEE Trans. Control Netw. Syst. 8 (1) (2020) 168–176, http://dx.doi.org/10.1109/TCNS.2020.3024315.

[30] A. Teixeira, I. Shames, H. Sandberg, K.H. Johansson, Revealing stealthy attacks in control systems, in: Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, USA, 2012, pp. 1806–1813, http://dx.doi.org/10.1109/Allerton.2012.6483441.

[31] T. Huang, B. Satchidanandan, P.R. Kumar, L. Xie, An online detection framework for cyber attacks on automatic generation control, IEEE Trans. Power Syst. 33 (2018) 6816–6827, http://dx.doi.org/10.1109/TPWRS.2018.2829743.

[32] S. Narasimhan, N.H. El-Farra, M.J. Ellis, Detectability-based controller design screening for processes under multiplicative cyberattacks, AIChE J. 68 (2022) e17430, http://dx.doi.org/10.1002/aic.17430.

[33] S. Chen, Z. Wu, P.D. Christofides, A cyber-secure control-detector architecture for nonlinear processes, AIChE J. 66 (5) (2020) e16907, http://dx.doi.org/10.1002/aic.16907.

[34] P.D. Christofides, N.H. El-Farra, Control of Nonlinear and Hybrid Process Systems: Designs for Uncertainty, Constraints and Time-Delays, Vol. 324, Springer Science & Business Media, 2005, http://dx.doi.org/10.1007/b105110.

[35] D. Popescu, A. Gharbi, D. Stefanoiu, P. Borne, Process Control Design for Industrial Applications, Wiley Online Library, 2017, http://dx.doi.org/10.1002/9781119407461.

[36] J. Romagnoli, A. Palazoglu, Introduction to Process Control, CRC Press, 2020, http://dx.doi.org/10.1201/9780429351396.

[37] V.M. Kuntsevich, B.N. Pshenichnyi, Minimal invariant sets of dynamic systems with bounded disturbances, Cybern. Syst. Anal. 32 (1) (1996) 58–64, http://dx.doi.org/10.1007/BF02366582.

[38] I. Kolmanovsky, E. Gilbert, Theory and computation of disturbance invariant sets for discrete-time linear systems, Math. Probl. Eng. 4 (1998) 317–367, http://dx.doi.org/10.1155/S1024123X98000866.

[39] P.M. Frank, X. Ding, Survey of robust residual generation and evaluation methods in observer-based fault detection systems, J. Process Control 7 (6) (1997) 403–424, http://dx.doi.org/10.1016/S0959-1524(97)00016-4.

[40] S. Simani, C. Fantuzzi, R. Patton, Model-Based Fault Diagnosis in Dynamic Systems using Identification Techniques, Springer, London, 2003, http://dx.doi.org/10.1007/978-1-4471-3829-7.

[41] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, Diagnosis and Fault-TOlerant Control, Springer-Verlag Berlin Heidelberg, 2003, http://dx.doi.org/10.1007/978-3-540-35653-0.

[42] R. Isermann, Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance, Springer-Verlag Berlin Heidelberg, 2006, http://dx.doi.org/10.1007/3-540-30368-5.

[43] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, A review of process fault detection and diagnosis: Part I: Quantitative model-based methods, Comput. Chem. Eng. 27 (2003) 293–311, http://dx.doi.org/10.1016/S0098-1354(02)00160-6.

[44] International Society of Automation, ANSI/ISA-18.2-2016: Management of Alarm Systems for the Process Industries, Standard, International Society of Automation, 2009.

[45] A. Alanqar, M. Ellis, P. Christofides, Economic model predictive control of nonlinear process systems using empirical models, AIChE J. 61 (3) (2015) 816–830, http://dx.doi.org/10.1002/aic.14683.

[46] M. Kvasnica, P. Grieder, M. Baotić, [Multi-parametric toolbox (MPT)], 2004, http://control.ee.ethz.ch/mpt/.

[47] S.V. Raković, E.C. Kerrigan, K.I. Kouramas, D.Q. Mayne, Invariant approximations of the minimal robust positively invariant set, IEEE Trans. Autom. Control 50 (3) (2005) 406–410, http://dx.doi.org/10.1109/TAC.2005.843854.