Digital Chemical Engineering 7 (2023) 100100

Contents lists available at ScienceDirect

Digital Chemical Engineering

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

Original article

A reachable set-based scheme for the detection of false data injection cyberattacks on dynamic processes

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

Department of Chemical Engineering, University of California, Davis, Davis, CA 95616, USA

ARTICLE INFO

Keywords:
False data injection cyberattacks
Cyberattack detection for dynamic operation
Process control systems

ABSTRACT

Recent cyberattacks targeting process control systems have demonstrated that reliance on information technology-based approaches alone to address cybersecurity needs is insufficient and that operational technology-based solutions are needed. An attack detection scheme that monitors process operation and determines the presence of an attack represents an operational technology-based approach. Attack detection schemes may be designed to monitor a process operated at or near its steady–state to account for the typical operation of chemical processes. However, transient operation may occur; for example, during process start-up and set–point changes. Detection schemes designed or tuned for steady-state operation may raise false alarms during transient process operation. In this work, we present a reachable set-based cyberattack detection scheme for monitoring processes during transient operation. Both additive and multiplicative false data injection attacks (FDIAs) that alter data communicated over the sensor–controller and controller–actuator communication links are considered. For the class of attacks considered, the detection scheme does not raise false alarms during transient operations. Conditions for classifying attacks based on the ability of the detection scheme to detect the attacks are presented. The application of the reachable set-based detection scheme is demonstrated using two illustrative processes under different FDIAs. For the FDIAs considered, their detectability with respect to the reachable set-based detection scheme is analyzed.

1. Introduction

Modern process control systems (PCSs) utilize networked communication between the sensors, the controller, and the control actuators to operate chemical manufacturing processes. In recent years, cyberattacks that maliciously alter data communicated over the PCS communication links have increased in frequency and severity (Duo et al., 2022). Traditionally, information technology-based approaches have been exclusively responsible for maintaining the cybersecurity of PCSs. However, recent attacks have highlighted the need to augment these cybersecurity approaches with operational technology-based solutions. Efforts to enhance cybersecurity through operational technology have focused on attack detection, identification, and mitigation (Duo et al., 2022).

Several detection schemes and methods have been proposed to monitor a process for cyberattacks. A common type of attack detection involves monitoring a residual, defined as the difference between a measured process variable and its estimate or prediction (Mo and Sinopoli, 2009; Hashemi et al., 2019; Trapiello and Puig, 2020; Liu et al., 2021; Oyama et al., 2021; Rangan et al., 2021; Ahmed et al., 2022; Narasimhan et al., 2022a,b; Oyama et al., 2022; Renganathan et al., 2022; Umsonst et al., 2022). Residual-based detection approaches include those using standard anomaly detection schemes (e.g., CUSUM and χ^2 detection schemes) (Mo and Sinopoli, 2009; Hashemi et al., 2019; Ahmed et al., 2022; Renganathan et al., 2022; Umsonst et al., 2022). Such detection schemes are tuned based on attack-free operation, often considering when the process operates at or near its steady state. Neural network-based detection schemes utilize a neural network trained with operational data under attack-free and under various cyberattacks to classify operation as either attack-free or not (Chen et al., 2020, 2021; Wu and Christofides, 2021; Zedan and El-Farra, 2021).

For a process subject to bounded disturbances, the minimum robust positively-invariant set is the neighborhood of the steady-state that the process states asymptotically converge to. In a prior work, a detection scheme was developed to monitor processes when the state evolves within the minimum invariant set (Narasimhan et al., 2022, 2022a). The detection scheme was tuned using the minimum invariant set of the closed-loop process, since chemical processes are typically operated at steady–state for long periods of time. An analysis of the closed-loop process under a cyberattack revealed a relationship between the

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

https://doi.org/10.1016/j.dche.2023.100100
Received 4 March 2023; Received in revised form 19 April 2023; Accepted 20 April 2023
Available online 24 April 2023

choice of PCS parameters and the ability to detect an attack, i.e., attack detectability. Specifically, the analysis revealed that PCS parameters can be selected to enable attack detection. However, operating the process with such parameters may degrade the attack-free closed-loop performance relative to the performance achieved with parameters chosen based on conventional performance-based approaches. An active PCS parameter switching-enabled detection method was proposed to manage the potential tradeoff between attack detectability and closed-loop performance (Narasimhan et al., 2022a). Switching the controller parameters may induce transients during which the process states evolve outside the minimum invariant set of the process operated under the new parameters, potentially triggering false alarms in the detection scheme. A state-dependent switching condition was proposed to minimize false alarms resulting from parameter switching (Narasimhan et al., 2022b). A smooth transition between controller parameters occurs when the switching condition is satisfied, meaning false alarms are not raised. Switching to probe for the presence of cyberattacks may be desirable to enable the detection of an attack, regardless of whether the condition is satisfied. Therefore, developing an attack detection scheme that can effectively monitor the process during transient operation is important.

Some approaches for the detection of cyberattacks on chemical processes during transient operation have been proposed (Chen et al., 2020, 2021; Oyama et al., 2021; Rangan et al., 2021; Wu and Christofides, 2021; Oyama et al., 2022). When extensive closed-loop data for the attack-free and the attacked process during transient operation is available, neural network-based detection schemes may be utilized to detect and identify the attack during transient operation (Chen et al., 2020, 2021; Wu and Christofides, 2021). However, extensive operational or simulation data for the closed-loop possible attacks may not be available. For processes operated using Lyapunov-based economic model predictive control (LEMPC), which may result in dynamic process operation, several integrated cyberattack detection and handling strategies have been proposed (Oyama et al., 2021; Rangan et al., 2021; Oyama et al., 2022). The detection strategies utilize a threshold approach for monitoring a residual in addition to monitoring that the process state evolves within its expected region of operation. However, selecting the detection threshold may be difficult, potentially requiring extensive data.

Reachability analysis has been used for analyzing processes under cyberattacks, and designing systems and methods that improve PCS cyberattack resilience (Mo and Sinopoli, 2012; Murguia et al., 2017; Kwon and Hwang, 2018; Trapiello and Puig, 2020). Specifically, the state and estimation error reachable sets under a stealthy attack were used as a measure of the system resilience (Mo and Sinopoli, 2012). A linear matrix inequality-based approach for performance-based controller parameter selection that minimizes the size of the reachable set under attack was developed (Murguia et al., 2017). Reachable sets have been used to compute all the possible states reached under stealthy cyberattacks (Kwon and Hwang, 2018)). An approach to design an input signal using reachable sets was proposed to ensure an attack is detected (Trapiello and Puig, 2020). While reachability analysis may be used to characterize the behavior of a process during transient operation, the use of reachable sets for detecting and classifying attacks targeting dynamic processes has not been previously explored.

In the present work, a reachable set-based detection scheme is proposed to monitor transient process operations for false data injection attacks (FDIAs) that alter the variable value communicated over the PCS communication links. Both sensor–controller and controller–actuator link FDIAs are considered. The proposed detection scheme verifies whether the value of a generalized monitoring variable at a given time step is contained within its reachable set for the attack-free process. The proposed detection scheme monitors the process without requiring extensive closed-loop data. It also does not raise false alarms during transient operation. Conditions that lead to an attack being detectable or undetectable with respect to the proposed detection scheme are characterized. The proposed detection scheme and the classification approach are applied to two illustrative examples. The detectability of different FDIAs is analyzed, and the applicability of the reachable set-based detection scheme and attack classification to a nonlinear chemical process is demonstrated.

2. Preliminaries

2.1. Notation

The set of non-negative integers is denoted by $\mathbb{Z}^+$. Given $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^n$, the Minkowski sum of $\mathcal{X}$ and $\mathcal{Y}$ is given by $\mathcal{X} \oplus \mathcal{Y} := \{x + y \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$. For the set $\mathcal{X} \subseteq \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{m\times n}$, $A\mathcal{X} := \{Ax \mid x \in \mathcal{X}\}$. Given a matrix $A \in \mathbb{R}^{m\times n}$ and a set $\mathcal{X} \subseteq \mathbb{R}^n$, $\bigoplus_{i=0}^{n-1} A^i\mathcal{X}$ represents the Minkowski sum given by $\mathcal{X}\oplus A\mathcal{X}\oplus\cdots\oplus A^{n-1}\mathcal{X}$. For a square matrix A, $\lambda_i(A)$ represents the ith eigenvalue of A. The identity matrix is denoted by I.

2.2. Class of attack-free processes

We consider in this work discrete-time linear processes with the following state–space dynamics:

$$x_{k+1} = Ax_k + B^u u_k + B^w w_k \tag{1}$$

where $A \in \mathbb{R}^{n_x\times n_x}$, $B^u \in \mathbb{R}^{n_x\times n_u}$, $B^w \in \mathbb{R}^{n_x\times n_w}$, $k \in \mathbb{Z}^+$ is the time step, $x_k \in \mathbb{R}^{n_x}$ is the state vector, $u_k \in \mathbb{R}^{n_u}$ is the manipulated input vector, and $w_k \in \mathcal{W} \subset \mathbb{R}^{n_w}$ is the process disturbance vector. Without loss of generality, the initial time step is assumed to be $k = 0$. Measurements from the process are available and are given by:

$$y_k = Cx_k + v_k \tag{2}$$

where $y_k \in \mathbb{R}^{n_y}$ is the measured output vector and $v_k \in \mathcal{V} \subset \mathbb{R}^{n_y}$ is the measurement noise vector. The sets $\mathcal{W}$ and $\mathcal{V}$ are the sets of admissible process disturbances and measurement noise, respectively, and are assumed to be convex polytopes. A Luenberger observer is synthesized to compute state estimates as follows:

$$\begin{aligned} \hat{x}_{k+1} &= A\hat{x}_k + B^u u_k + L(y_k - \hat{y}_k) \\ \hat{y}_k &= C\hat{x}_k \end{aligned} \tag{3}$$

where $L \in \mathbb{R}^{n_x\times n_y}$ is the observer gain, $\hat{x}_k \in \mathbb{R}^{n_x}$ is the estimated state, and $\hat{y}_k \in \mathbb{R}^{n_y}$ is the estimated output. The estimation error, defined as the difference between the process state and the estimate ($e_k := x_k - \hat{x}_k$), has the following dynamics:

$$e_{k+1} = (A - LC)e_k + B^w w_k - Lv_k \tag{4}$$

The observer gain L is selected such that all eigenvalues of the matrix $A-LC$ lie within the unit circle. The control objective is to stabilize the closed-loop process around its steady-state, assumed to be the origin of the unperturbed system. To achieve the control objective, a linear control law of the following form is synthesized:

$$u_k = -K\hat{x}_k \tag{5}$$

where $K \in \mathbb{R}^{n_u\times n_x}$ is the controller gain. The controller gain K is selected to ensure that all eigenvalues of the matrix $A - BK$ lie within the unit circle.

The dynamics of the process state and the estimation error collectively capture the attack-free closed-loop process dynamics. An augmented state vector is defined and denoted by $\xi_k := [x_k^T \; e_k^T]^T$, with dynamics:

$$\xi_{k+1} = \underbrace{\begin{bmatrix} A - B^uK & B^uK \\ 0 & A - LC \end{bmatrix}}_{=: A^\xi} \xi_k + \underbrace{\begin{bmatrix} B^w & 0 \\ B^w & -L \end{bmatrix}}_{=: B^d} d_k \tag{6}$$

where $d_k \in D := \mathcal{W} \times \mathcal{V}$ is a concatenated vector that includes the process disturbance and measurement noise vectors, i.e., $d_k :=$

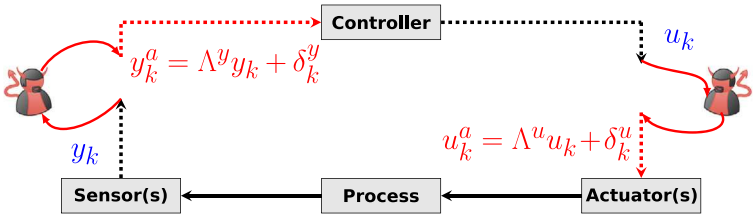

Fig. 1. A block diagram illustrating a process control system under a false data injection attack that simultaneously alters the data over the sensor–controller and controller–actuator communication links.

$[w_k^T \ v_k^T]^T$. The input d_k is called the disturbance for simplicity. The augmented system described in Eq. (6) is referred to as the attack-free closed-loop process. For the closed-loop process, its initial set $\mathcal{R}_0^\xi \subset \mathbb{R}^{2n_x}$ is defined as the region in state–space that contains the value of the augmented state at time step $k = 0$, i.e., $\xi_0 \in \mathcal{R}_0^\xi \subset \mathbb{R}^{2n_x}$. The initial set is assumed to be a polytope. Provided a set of initial states $\mathcal{R}_0^\xi \subset \mathbb{R}^{2n_x}$, the k-step forward reachable set, denoted by $\mathcal{R}_k^\xi(\mathcal{R}_0^\xi)$, for the closed-loop process is the set consisting of all states that can be reached in k time steps under any admissible disturbance, and is given by (e.g., the unlabeled equation preceding Eq. (2) in Girard et al. (2006)):

$$\mathcal{R}_k^\xi(\mathcal{R}_0^\xi) = A^{\xi k}\mathcal{R}_0^\xi \bigoplus_{i=0}^{k-1} A^{\xi i} B^d \mathcal{D} \tag{7}$$

The k-step reachable set for the attack-free closed-loop process depends on the controller and observer gains (K, L), the disturbance set $\mathcal{D}$, and the initial set $\mathcal{R}_0^\xi$. As $k \to \infty$, the k-step forward reachable sets converge to the minimum invariant set ($\mathcal{R}_\infty^\xi := \bigoplus_{i=0}^{\infty} A^{\xi i} B^d \mathcal{D}$), which is the limit set for all trajectories of the process (Raković et al., 2005).

Remark 1. The initial set $\mathcal{R}_0^\xi$ and the disturbance set $\mathcal{D}$ are assumed to be polytopes. With this assumption, Eq. (7) can be computed by recursively applying the following two properties: (1) for two polytopes $\mathcal{D}_1$ and $\mathcal{D}_2$, $\mathcal{D}_1 \oplus \mathcal{D}_2$ can be computed by adding the vertices of $\mathcal{D}_1$ to the vertices of $\mathcal{D}_2$ where the resulting vectors form the vertices of $\mathcal{D}_1 \oplus \mathcal{D}_2$, and (2) for a polytope $\mathcal{D}$, $A^{\xi i} B^d \mathcal{D}$ is a polytope that can be computed by pre-multiplying all vertices of $\mathcal{D}$ by $A^{\xi i} B^d$ and taking the convex hull of the resultant vectors. The assumption that $\mathcal{R}_0^\xi$ and $\mathcal{D}$ are polytopes enables the calculation of Eq. (7) with a finite number of computations.

2.3. False data injection attacks

False data injection attacks (FDIAs) refer to cyberattacks that alter the output or input values communicated over a communication link so that the receiver, i.e., the controller or the actuators, receives the altered value. In the present work, both additive and multiplicative FDIAs that alter data communicated over the sensor–controller and controller–actuator communication links are considered. In the presence of an attack, the value of the variable altered by the attack is given by:

$$\phi_k^a = \Lambda^\phi \phi_k + \delta_k^\phi \tag{8}$$

where $\phi_k \in \mathbb{R}^{n_\phi}$ is the unaltered value of the variable, ϕ_k^a is the altered value of the variable ϕ_k, $\Lambda^\phi \in \mathbb{R}^{n_\phi \times n_\phi}$ is a multiplicative factor to represent multiplicative FDIAs, and $\delta_k^\phi \in \mathbb{R}_k^{n_\phi}$ is the additive bias to represent additive FDIAs. For sensor–controller link FDIAs, ϕ_k represents the sensor measurements (y_k); for controller–actuator link FDIAs, ϕ_k represents the controller output (u_k). Fig. 1 illustrates the block diagram of a process control system under a false data injection attack that simultaneously alters the data over the sensor–controller and controller–actuator links. In the presence of an attack, the values of the measured output and the control input (y_k and u_k shown in blue text) are altered by the attacker and reported over the compromised communication links as $y_k^a = \Lambda^y y_k + \delta_k^y$ and $u_k^a = \Lambda^u u_k + \delta_k^u$, respectively (shown in red text).

FDIAs alter the closed-loop behavior of the process. The augmented state dynamics of the closed-loop process subject to an additive and multiplicative FDIA are given by:

$$\xi_{k+1} = \underbrace{\begin{bmatrix} A - B^u \Lambda^u K & B^u \Lambda^u K \\ L(I - \Lambda^y)C & A - LC \end{bmatrix}}_{=: A^{\xi a}} \xi_k + \underbrace{\begin{bmatrix} B^w & 0 \\ B^w & -L\Lambda^y \end{bmatrix}}_{=: B^{d a}} d_k + \underbrace{\begin{bmatrix} 0 & B^u \\ -L & 0 \end{bmatrix}}_{=: B^{\delta a}} \delta_k \tag{9}$$

where $\delta_k = \left[(\delta_k^y)^T \quad (\delta_k^u)^T\right]^T$. The closed-loop process described by Eq. (9) is referred to as the attacked closed-loop process. Similar to the attack-free process, the k-step reachable set under an FDIA is given by:

$$\mathcal{R}_k^{\xi_a}(\mathcal{R}_0^\xi) = A^{\xi_a k}\mathcal{R}_0^\xi \bigoplus_{i=0}^{k-1} A^{\xi_a i} \mathcal{D}_k^a \tag{10}$$

where $\mathcal{D}_k^a = B^{d_a}\mathcal{D} \oplus B^{\delta_a}\{\delta_k\}$. The attack is generally unknown, so the k-step reachable sets of the attacked process may not be computable for purposes of online attack detection. However, the k-step reachable sets of the attacked process can be used for (offline) classification of specific attacks as detectable or not (this point is discussed further in Section 3.2).

3. Attack detection for processes during transient operation

In this section, a class of reachable set-based attack detection schemes utilizing a generalized monitoring variable are presented to monitor the closed-loop process during transient operation. A method for classifying attacks as detectable, potentially detectable, or undetectable under the proposed detection scheme is also presented.

3.1. Reachable set-based detection scheme

Cyberattack detection schemes often use the measured output, estimated output, or the residual vector ($r_k := y_k - \hat{y}_k$) as the monitoring variable(s) to detect an attack (e.g., Na and Eun, 2018; Hashemi et al., 2019; Cómbita et al., 2022; Narasimhan et al., 2022a; Renganathan et al., 2022). Some attacks may evade detection by a scheme that uses only one of the three variables, but may be detected using a detection scheme based on another variable (Narasimhan et al., 2022b). In this work, a generalized monitoring variable that may be expressed as a linear combination of the measured output and its estimate generated by the observer is considered:

$$\eta_k := H^y y_k + H^{\hat{y}} \hat{y}_k \tag{11}$$

where $\eta_k \in \mathbb{R}^{n_\eta}$ is the generalized monitoring variable and the matrices H^y and $H^{\hat{y}}$ are design parameters of the detection scheme. When H^y and $H^{\hat{y}}$ are chosen such that $H^y = I$ and $H^{\hat{y}} = -I$, the monitoring variable becomes the residual vector ($\eta_k = r_k$). A choice of $H^y = I$ and $H^{\hat{y}} = 0$, on the other hand, results in the monitoring variable being the measured output. Expressing the monitoring variable in terms of the augmented state and the disturbance vector gives:

$$\eta_k = \underbrace{\left[(H^y - H^{\hat{y}})C \quad H^{\hat{y}}C\right]}_{=: C^\xi} \xi_k + \underbrace{\left[0 \quad H^y\right]}_{=: D^d} d_k \tag{12}$$

To address the problem of attack detection during transient operation, we consider in this work the reachable sets of the monitoring variable for the attack-free closed-loop process. For the attack-free closed-loop process and initial set $\mathcal{R}_0^\xi$, the augmented state is contained within the k-step reachable set for all $k \in \mathbb{Z}^+$. From Eqs. (7) and (12), the generalized monitoring variable of the attack-free process is contained in the set:

$$\mathcal{R}_k^\eta(\mathcal{R}_k^\xi) := C^\xi \mathcal{R}_k^\xi(\mathcal{R}_0^\xi) \oplus D^d D \tag{13}$$

The containment of the monitoring variable within the k-step reachable sets of the attack-free process may be verified to monitor the process for attacks as follows:

$$h(\eta_k, \mathcal{R}_k^\xi) = \begin{cases} 1, & \eta_k \notin \mathcal{R}_k^\eta(\mathcal{R}_k^\xi) \\ 0, & \eta_k \in \mathcal{R}_k^\eta(\mathcal{R}_k^\xi) \end{cases} \tag{14}$$

where the mapping h returns the output of the detection scheme. An output of 1 indicates that an attack is detected, and the detection scheme is said to raise an alarm. An output of 0 indicates that no attack is detected. To implement the reachable set-based detection scheme, knowledge of the initial set is required. For a process transitioning from one steady-state to another, the minimum invariant set of the process at the initial steady-state may be used as the initial set. The initial set for process start-up may also be known.

The reachable set-based detection scheme in Eq. (14) is designed to detect an attack if there is a discrepancy between the observed value of the monitoring variable and its expected attack-free value, i.e., the reachable sets are computed for the attack-free process. In the absence of an attack, the values of the generalized monitoring variable are contained within k-step reachable sets for the attack-free process, and the detection scheme generates an output of 0 for all $k \in \mathbb{Z}^+$. Therefore, a necessary condition for attack-free operation is that the monitoring variable must be contained within its k-step reachable set, implying that no attacks are detected. This is formalized in the following proposition.

Proposition 1. *Consider the closed-loop process in Eq. (9) monitored by the reachable set-based detection scheme in Eq. (14), with an initial set $\mathcal{R}_0^\xi$. The closed-loop process is attack-free only if the output of the detection scheme in Eq. (14) is $h(\eta_k, \mathcal{R}_k^\xi) = 0$ for all $k \in \mathbb{Z}^+$.*

Proof. For the attack-free closed-loop process with $\xi_0 \in \mathcal{R}_0^\xi$, the augmented state is contained within the k-step reachable set, i.e., $\xi_k \in \mathcal{R}_k^\xi(\mathcal{R}_0^\xi)$ for all $k \in \mathbb{Z}^+$. From Eq. (13), the generalized monitoring variable of the attack-free process is contained within its k-step reachable set, i.e., $\eta_k \in \mathcal{R}_k^\eta(\mathcal{R}_k^\xi)$ for all $k \in \mathbb{Z}^+$. From Eq. (14), the output of the detection scheme is $h(\eta_k, \mathcal{R}_k^\xi) = 0$ for all $k \in \mathbb{Z}^+$. □

A direct implication of Proposition 1 is that the detection scheme does not raise false alarms during transient operation. While the reachable set-based detection scheme is designed on the basis of attack-free process behavior, an attack may be detected if the detection scheme returns a value of 1 at some $k \in \mathbb{Z}^+$.

Corollary 1. *Consider the closed-loop process in Eq. (9) monitored by the reachable set-based detection scheme in Eq. (14), with an initial set $\mathcal{R}_0^\xi$. If the output of the detection scheme is $h(\eta_{k_d}, \mathcal{R}_{k_d}^\xi) = 1$ for some $k_d \in \mathbb{Z}^+$, then the process cannot be attack-free.*

3.2. Classification of attack detectability

Attacks can be classified based on the ability or inability of the reachable set-based detection scheme to detect an attack. Defining attack detectability requires certain considerations, including the dependence of reachable sets on the initial set $\mathcal{R}_0^\xi$. An attack is detected at time k_d if $h(\eta_{k_d}, \mathcal{R}_{k_d}^\xi) = 1$. An attack is detectable with respect to the reachable set-based detection scheme and the initial set $\mathcal{R}_0^\xi$ if the attack is detected in finite time for all $\xi_0 \in \mathcal{R}_0^\xi$ (and $d_k \in D$). An attack is undetectable with respect to the reachable set-based detection scheme and the initial set $\mathcal{R}_0^\xi$ if the attack is not detected in finite time for all $\xi_0 \in \mathcal{R}_0^\xi$ (and $d_k \in D$). For simplicity of presentation, detectable and undetectable attacks with respect to the detection scheme and initial set $\mathcal{R}_0^\xi$ are called detectable and undetectable attacks, respectively. An attack is potentially detectable if it is neither detectable nor undetectable.

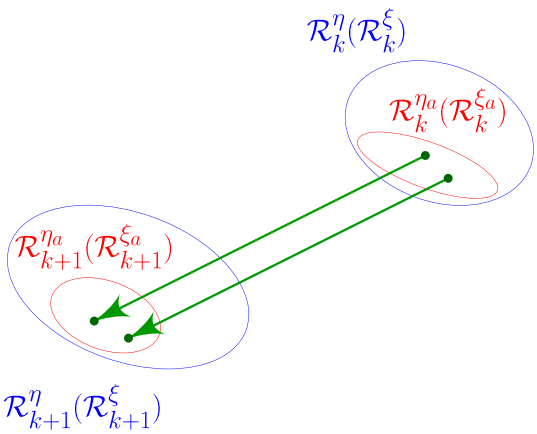

Fig. 2. Illustrative example showing the reachable sets of the monitoring variable for the attack-free (blue sets) and the attacked (red sets) process in the presence of an undetectable attack, with two example trajectories (green lines) for the attacked process.

With the definitions above, conditions based on the relationship between the reachable sets of the attacked and the attack-free process can be established and used for classifying attacks. In the propositions below, an FDIA that begins at $k = 0$ is considered. The results may be extended to an attack occurring at any time. The first proposition establishes that if all possible values of the monitoring variable of the attacked process are contained within the reachable sets for the attack-free process, then the attack is undetectable.

Proposition 2. *Consider the closed-loop process in Eq. (9), with an initial set $\mathcal{R}_0^\xi$, under an FDIA beginning at $k = 0$. The attack is undetectable with respect to the detection scheme in Eq. (14) and the initial set $\mathcal{R}_0^\xi$ if and only if the reachable sets of the monitoring variable for the attacked and the attack-free process satisfy $\mathcal{R}_k^{\eta_a}(\mathcal{R}_k^{\xi_a}) \subseteq \mathcal{R}_k^\eta(\mathcal{R}_k^\xi)$ for all $k \in \mathbb{Z}^+$.*

Proof (*Sufficiency*)**.** Consider the attacked closed-loop process and the initial set $\mathcal{R}_0^\xi$. Let the reachable set of the monitoring variable for the attacked process be a subset of, or equal to, the reachable set for the attack-free process; i.e., $\mathcal{R}_k^{\eta_a}(\mathcal{R}_k^{\xi_a}) \subseteq \mathcal{R}_k^\eta(\mathcal{R}_k^\xi)$ for all $k \in \mathbb{Z}^+$. This implies that the monitoring variable values are contained within the reachable sets for the attack-free process ($\eta_k \in \mathcal{R}_k^{\eta_a}(\mathcal{R}_k^{\xi_a}) \subseteq \mathcal{R}_k^\eta(\mathcal{R}_k^\xi)$), and the detection scheme generates an output of $h(\eta_k, \mathcal{R}_k^\xi) = 0$ for all $k \in \mathbb{Z}^+$. Therefore, the attack is undetectable.

Necessity: Consider the attacked closed-loop process with the initial set $\mathcal{R}_0^\xi$. Let the FDIA begin at $k = 0$ and be undetectable with respect to the detection scheme in Eq. (14) and the initial set $\mathcal{R}_0^\xi$. By definition of an undetectable attack, $h(\eta_k, \mathcal{R}_k^\xi) = 0$ for all $k \in \mathbb{Z}^+$, $\xi_0 \in \mathcal{R}_0^\xi$, and $d_k \in D$. From Eq. (14), this implies that $\eta_k \in \mathcal{R}_k^\eta(\mathcal{R}_k^\xi)$ for all $k \in \mathbb{Z}^+$. However, the process is subjected to the FDIA, so $\eta_k \in \mathcal{R}_k^{\eta_a}(\mathcal{R}_k^{\xi_a})$ for all $k \in \mathbb{Z}^+$, implying that $\mathcal{R}_k^{\eta_a}(\mathcal{R}_k^{\xi_a}) \subseteq \mathcal{R}_k^\eta(\mathcal{R}_k^\xi)$ for all $k \in \mathbb{Z}^+$. □

The condition presented in Proposition 2 is a necessary and sufficient condition for an undetectable attack. Fig. 2 provides a pictorial interpretation of the result of Proposition 2. It illustrates the reachable sets of the attacked process (sets in red) and the attack-free process (sets in blue) over two time steps for a process under an undetectable attack.

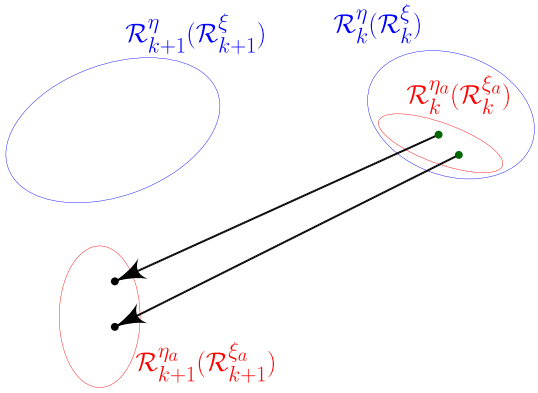

Fig. 3. Illustrative example showing the reachable sets of the monitoring variable for the attack-free (blue sets) and the attacked (red sets) process in the presence of a detectable attack, with two example trajectories for the attacked process.

The figure also illustrates two example trajectories of the monitoring variable for the attacked process (green lines). As illustrated, the values of the monitoring variable at the time steps k and $k+1$ (shown by the green circle markers) are contained within the intersection of the reachable sets for the attack-free and attacked process, leading to an output of 0 from the detection scheme. Therefore, the attack is not detected by the detection scheme. While only two time steps are illustrated in Fig. 2, the reachable sets of the process under an undetectable attack must be contained within the attack-free reachable sets for all time steps $k \in \mathbb{Z}^+$.

If the reachable set of the monitoring variable for the attacked process does not intersect the reachable set of the attack-free process at some time $k \in \mathbb{Z}^+$, the attack will be detected at time k, and is detectable. This is formally stated in the following proposition.

Proposition 3. *Consider the closed-loop process in Eq. (9), with an initial set $\mathcal{R}^{\xi}_0$, under an FDIA beginning at $k = 0$. The attack is detectable if the reachable sets of the monitoring variable for the attacked and the attack-free process satisfy $\mathcal{R}^{\eta_a}_k(\mathcal{R}^{\xi_a}_k) \cap \mathcal{R}^{\eta}_k(\mathcal{R}^{\xi}_k) = \emptyset$ for some $k \in \mathbb{Z}^+$.*

Proof. If the reachable sets of the generalized monitoring variable for the attacked and the attack-free process do not intersect at some $k \in \mathbb{Z}^+$, i.e., $\mathcal{R}^{\eta_a}_k(\mathcal{R}^{\xi_a}_k) \cap \mathcal{R}^{\eta}_k(\mathcal{R}^{\xi}_k) = \emptyset$, no value of the monitoring variable that is contained within the attacked reachable set is contained within the attack-free reachable set, i.e., $\eta_k \notin \mathcal{R}^{\eta}_k(\mathcal{R}^{\xi}_k)$. The output of the detection scheme in this case is $h(\eta_k, \mathcal{R}^{\xi}_k) = 1$, and the attack is detected. Hence, the attack is detectable. □

Fig. 3 provides an illustration of the idea behind Proposition 3. The figure shows the reachable sets of the monitoring variable for the attack-free process (blue sets) and those of the process under a detectable attack (red sets) over two time steps. At time step k, the reachable set for the attacked process is contained entirely within the reachable set for the attack-free process. At time step $k+1$, the reachable set of the attacked process does not intersect the reachable set of the attack-free process. For all initial values, no value of the monitoring variable of the attacked process is contained in the reachable set of the attack-free process (illustrated by the black circle markers in Fig. 3). As a result, the attack is detected at time step $k+1$ with the detection scheme generating an output of 1, i.e., $h(\eta_{k+1}, \mathcal{R}^{\xi}_{k+1}) = 1$ for all $\xi_0 \in \mathcal{R}^{\xi}_0$.

Attacks that do not satisfy the conditions in Proposition 2 or Proposition 3 are also possible. For such attacks, the reachable sets of the attacked process intersect with the reachable sets of the attack-free process for all time steps, and the reachable sets of the attacked process are not contained in the corresponding reachable sets of the attack-free process for at least one time step, i.e., $\mathcal{R}^{\eta}_k(\mathcal{R}^{\xi}_k) \cap \mathcal{R}^{\eta_a}_k(\mathcal{R}^{\xi_a}_k) \neq \emptyset$ for all $k \in \mathbb{Z}^+$ and $\mathcal{R}^{\eta_a}_k(\mathcal{R}^{\xi_a}_k) \not\subseteq \mathcal{R}^{\eta}_k(\mathcal{R}^{\xi}_k)$ for some $k \in \mathbb{Z}^+$. While the attack cannot be undetectable by Proposition 2, the attack may be detectable or potentially detectable. For example, consider an attack on the process such that the monitoring variable of all possible trajectories leaves its attack-free reachable set. In this case, the attack is detectable. This is illustrated by the following example:

$$\begin{aligned} x_{k+1} &= 0.9x_k + \delta^u_k \\ y_k &= x_k \end{aligned} \tag{15}$$

where $x_k \in \mathbb{R}$ is the state, $y_k \in \mathbb{R}$ is the measurement, and $\delta^u_k \in \mathbb{R}$ is an additive controller–actuator link FDIA. Consider the initial set of $\{0\}$ and let the measured output be the monitoring variable, i.e., $\eta_k = y_k$. For the attack-free process, the monitoring variable takes a value of 0 for all k, and the reachable sets of the monitoring variable are $\{0\}$ for all $k \in \mathbb{Z}^+$. Let the attack δ^u_k be a bounded random variable such that $|\delta^u_k| \leq \bar{\delta}$ for all $k \in \mathbb{Z}^+$, where $\bar{\delta} > 0$. Moreover, let δ^u_k take a non-zero value for at least one time step. The reachable sets of the attacked process contain the origin, so they intersect with the attack-free reachable sets for all $k \in \mathbb{Z}^+$. When δ^u_k takes a non-zero value, the state and monitoring variable will move away from 0, so the attack is detected. Thus, the attack is detectable.

An attack that does not satisfy the conditions in Proposition 2 or Proposition 3 may be potentially detectable if there are some trajectories where the attack is detected and others where the attack is not detected. For example, consider the following process:

$$\begin{aligned} x_{k+1} &= 0.9x_k + d_k \\ y_k &= \Lambda^y x_k \end{aligned} \tag{16}$$

where d_k is the process disturbance taking values in the set $[-1, 1]$ and $\Lambda^y = 1.1$ is a multiplicative FDIA altering the data over the sensor–controller link. For the attack-free process with $A^{\xi} = 0.9$, $B^d = 1$, and disturbances bounded as $\mathcal{D} = [-1, 1]$, the minimum invariant set (computed based on the method presented in Raković et al. (2005)) is $[-10, 10]$, meaning that for any $d_k \in \mathcal{D}$, $x_{k+1} \in [-10, 10]$ if $x_k \in [-10, 10]$. Let the initial set be equal to the minimum invariant set of the attack-free process, i.e., $[-10, 10]$, and let the monitoring variable be the measured output. For a process evolving from an initial set $\mathcal{R}^{\xi}_0 \in [-10, 10]$, the k-step forward reachable set of the process is the minimum invariant set itself. Therefore, the reachable sets of the monitoring variable are $[-10, 10]$ for all $k \in \mathbb{Z}^+$. If the initial state of the attacked process is $x_0 = 0$ and the disturbance takes a value of zero for all time, i.e., $d_k = 0$ for all $k \in \mathbb{Z}^+$, the monitoring variable of the attacked process takes a value of 0 for all time, and the attack will not be detected. For some other initial states and disturbances, the attack will be detected. If $x_0 = 10$ and $d_0 = 1$, for example, the value of the monitoring variable is not contained within the reachable set at $k = 1$, since $\eta_1 = 11 \notin [-10, 10]$. In this case, the attack is detected. The attack is potentially detectable because there are some trajectories for which the attack will be detected and other trajectories for which the attack is not detected.

Remark 2. From Eq. (10), the reachable sets of the attacked process depend on the initial state and the matrices A^{ξ_a}, B^{d_a}, and B^{δ_a}, which depend on the controller and observer gains. Therefore, the detectability of an attack is influenced by the controller and observer gains and the initial set. This dependence may be exploited to design methods that help attack detection.

Remark 3. With respect to the reachable set-based detection scheme, the detectability of an attack depends on how the reachable sets of the monitoring variable for the attacked process evolve with respect to the evolution of the reachable sets of the monitoring variable for the attack-free process. The detectability of an attack may vary with the monitoring variable (i.e., the choice of H^y and $H^{\hat{y}}$). From Eq. (14), the parameters H^y and $H^{\hat{y}}$ influence the reachable sets of the monitoring variable for the attack-free process. For the attacked process evolving

from an initial set $\mathcal{R}_0^{\xi} \subset \mathbb{R}^{2n_x}$, the reachable sets of the monitoring variable are also influenced by the parameters H^y and $H^{\hat{y}}$:

$$\mathcal{R}_k^{\eta_a}(\mathcal{R}_k^{\xi_a}) = C^{\xi_a}\mathcal{R}_k^{\xi} \oplus D^{\eta_a} \tag{17}$$

where $C^{\xi_a} = \begin{bmatrix}(H^y\Lambda^y + H^{\hat{y}})C & -H^{\hat{y}}C\end{bmatrix}$, $D^{\eta_a} = \begin{bmatrix}0 & H^y\Lambda^y\end{bmatrix} d_k \oplus \begin{bmatrix}H^y & 0\end{bmatrix}\{\delta_k\}$. Based on Eqs. (14) and (17), the parameters H^y and $H^{\hat{y}}$ influence the evolution of the reachable sets of the monitoring variable for the attacked and the attack-free process. Therefore, H^y and $H^{\hat{y}}$ influence attack detectability.

Remark 4. An additional factor that may influence attack detectability is the closed-loop stability of the attacked process. Specifically, when the magnitudes of the multiplicative components of an attack (Λ^y and Λ^u) are such that $\max_i|\lambda_i(A^{\xi_a})| \geq 1$, the attack destabilizes the process and may cause an unbounded growth in the norm of the augmented state. If an additional observability condition is satisfied (Narasimhan et al., 2022a), the attack may be detected because the generalized monitoring variable may not be contained within its k-step reachable set for the attack-free process at some time step $k \in \mathbb{Z}^+$.

Remark 5. For the attacked closed-loop process, the computation of the k-step reachable sets requires knowledge of the attack, which is unknown in general. Therefore, the detectability-based classification of attacks may be performed (offline) for various attacks.

4. Numerical results: Scalar process example

In this section, the proposed reachable set-based detection scheme, as well as the detectability-based classification of attacks, are applied to a scalar process during transient operation. All polytope computations are performed using the MPT 3.0 toolbox (Herceg et al., 2013). A scalar process with the following process dynamics, measurement output, and control action is considered:

$$\begin{aligned} x_{k+1} &= x_k + u_k + w_k \\ u_k &= -\Lambda^u K\hat{x}_k + \delta_k^u \\ y_k &= \Lambda^y(x_k + v_k) + \delta_k^y \end{aligned}$$

where $x_k \in \mathbb{R}$ is the state, $u_k \in \mathbb{R}$ is the control action received by the actuator, $w_k \in \mathcal{W} := \{w' \mid |w'| \leq 1\}$ is the process disturbance, $y_k \in \mathbb{R}$ is the measurement output received by the controller, and $v_k \in \mathcal{V} := \{v' \mid |v'| \leq 1\}$ is the measurement noise. The process disturbance and measurement noise are modeled as random variables following a uniform distribution bounded between -1 and 1. The process may be subject to an FDIA that simultaneously alters the data communicated over the controller–actuator and the sensor–controller links. To monitor the process for attacks, a monitoring variable that is a concatenation of the measured output and the residual vector is chosen, i.e., $\eta_k = [y_k \; r_k]^T$. The monitoring variable fits the model for the generalized monitoring variable in Eq. (11) with $H^y = [1\;1]^T$ and $H^{\hat{y}} = [0\;-1]^T$.

The process evolving from an initial set to the minimum invariant set is considered. A detection scheme tuned for steady-state operation (e.g., the detection scheme presented in Narasimhan et al. (2022a)) is not applicable to monitor the process because it may raise alarms as the process evolves from its initial condition to the minimum invariant set during attack-free operation. Instead, the reachable set-based detection scheme in Eq. (14) is utilized. Three case studies are presented in this section. Each case study considers the process under a different attack. In the first case study, the application of the reachable set-based detection scheme is demonstrated. Additionally, the detectability-based classification of a simultaneous additive and multiplicative FDIA, which alters the data over the sensor–controller and controller–actuator links, is presented. In the second and third case studies, an additive FDIA and a multiplicative FDIA are considered, respectively. In all cases below, the polytope representing the initial set considered is the attack-free minimum invariant set by shifted a vector (i.e., $\mathcal{R}_0^{\xi} = \mathcal{R}_{\infty}^{\xi} \oplus \{\xi'\}$ where ξ' is the shifting vector).

4.1. Application of the reachable set-based detection scheme and detectability-based classification of attacks

The process evolving from an initial set that is the attack-free minimum invariant set shifted by $\xi' = [100\;\;-50]^T$ ($\mathcal{R}_0^{\xi}$ in Fig. 4(a)) is considered. The attack-free process (first simulation set) and attacked process (second simulation set) are considered to demonstrate the reachable set-based detection scheme. Each simulation set consists of 1000 simulations of the process evolving from $\xi_0 = [103\;\;-48]^T \in \mathcal{R}_0^{\xi}$ to the minimum invariant set. The total length of each simulation is 5000 time steps. For the attacked process, the cyberattack begins at $k = 0$, and is an FDIA with multiplicative factors $\Lambda^y = 0.9$ and $\Lambda^u = 1.05$ and additive biases, which are random variables drawn from a uniform distribution, where $\delta_k^y \in [-0.1, 0.1]$ and $\delta_k^u \in [-0.1, 0.1]$ for all $k \in \mathbb{Z}^+$. For demonstration purposes, the controller and observer gains are chosen as $K = 0.5$ and $L = 1.5$ because the attack on the process operated with $K = 0.5$ and $L = 1.5$ is found to be detectable, as described below.

In the first simulation set, the attack-free process is considered. In every simulation, the values of the state and the estimation error are contained within the reachable sets of the attack-free process. Similarly, the values of the monitoring variable are always contained within their reachable sets. Therefore, the output of the reachable set-based detection scheme is equal to 0 in all simulations, indicating a lack of attack detection. Fig. 4(a) illustrates the values of the state and estimation error over one simulation of the attack-free process, and Fig. 4(b) illustrates the values of the output and estimation error over the same simulation. The values of all variables are contained within their corresponding reachable sets over the simulation, and no alarms are raised. The result demonstrates that the reachable set-based detection scheme does not raise false alarms during dynamic operation.

The second simulation set considers the attacked process. The attack is classified based on its detectability. Applying Proposition 3, the attack is detectable because $\mathcal{R}_k^{\eta_a}(\mathcal{R}_k^{\xi_a})$ and $\mathcal{R}_k^{\eta}(\mathcal{R}_k^{\xi})$ do not intersect at $k = 0$ and $k = 1$, as depicted in Fig. 5(a). Several closed-loop simulations are performed to verify that the attack is detected in all simulations. The detection scheme raises an alarm in all simulations at $k = 0$ and $k = 1$. For some simulations, an alarm is raised over subsequent time steps, but the attack is no longer detected over time once the augmented state converges to the minimum invariant set, i.e., the alarm goes away over time. This behavior occurs because $\mathcal{R}_{\infty}^{\eta} \subset \mathcal{R}_{\infty}^{\eta_a}$ (albeit this is difficult to see from Fig. 5(a)), but the non-intersecting area between the two sets is small. Fig. 5(b) illustrates the values of the monitoring variable over one simulation of the attacked process. Over this simulation, the attack is detected at all $k \in [0, 4]$. For $k \in [5, 5000]$, the monitoring variable evolves within the reachable sets of the attack-free process, and no alarms are raised.

4.2. Factors influencing the detectability of a multiplicative FDIA

In this case study, a multiplicative attack that alters the data communicated over the sensor–controller and controller–actuator links with pre-multiplication factors $\Lambda^y = 1.1$ and $\Lambda^u = 0.9$ and biases $\delta_k^y = \delta_k^u = 0$ for all $k \in \mathbb{Z}^+$ is considered. The impact of the initial set and controller and observer gains on the detectability of this attack is explored. The process evolving from two different initial sets is considered to explore the impact of the initial set on the attack detectability. The first initial set is the attack-free minimum invariant set shifted by $\xi' = [10\;\;-5]^T$, and the second is the attack-free minimum invariant set shifted by $\xi' = [50\;\;-20]^T$. The process is operated with controller and observer gains of $K = 1$ and $L = 0.9$. Fig. 6(a) and Fig. 6(b) illustrate the reachable sets of the monitoring variable for the attack-free and the attacked processes for a few time steps starting from the first and second initial set, respectively. For the given controller and observer gains, the attack is potentially detectable or detectable with respect to the first initial set because the two sets intersect for all time, and the

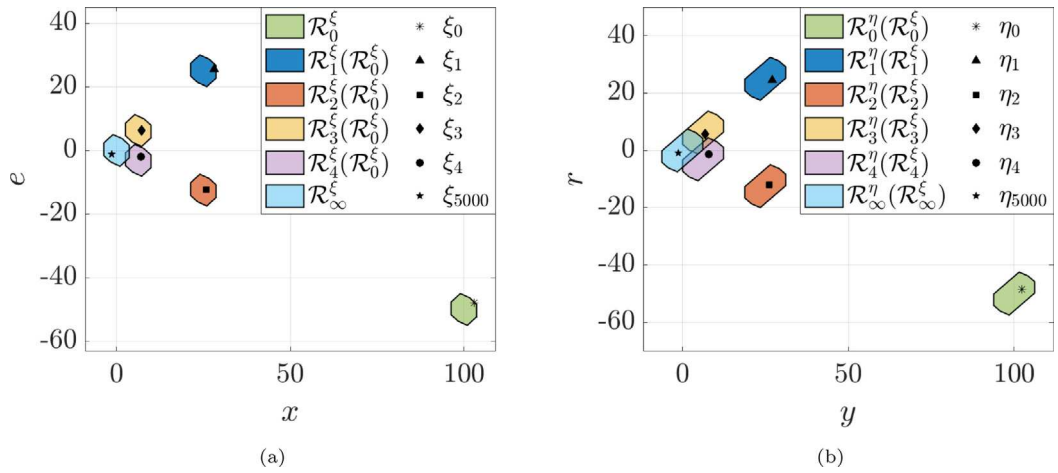

Fig. 4. (a) The state and estimation error values, (b) the monitoring variable values, used in the reachable set-based scheme, and (a)–(b) their corresponding reachable sets for the attack-free process over five time steps.

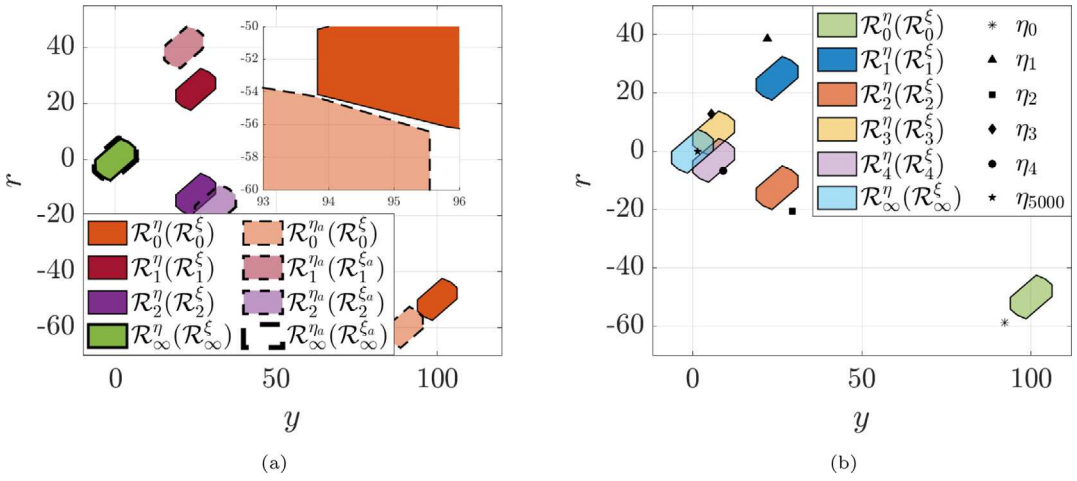

Fig. 5. (a) Evolution of the reachable sets of the monitoring variable for the attack-free and the attacked process over a few time steps. At $k = 0$ and $k = 1$, $\mathcal{R}_k^\eta \cap \mathcal{R}_k^{\eta_a} = \emptyset$, indicating the attack is detectable. The localized zoom in the figure illustrates that the reachable sets at $k = 0$ do not intersect. (b) The values of the monitoring variable of the attacked process and the reachable sets used in the detection scheme over a few time steps.

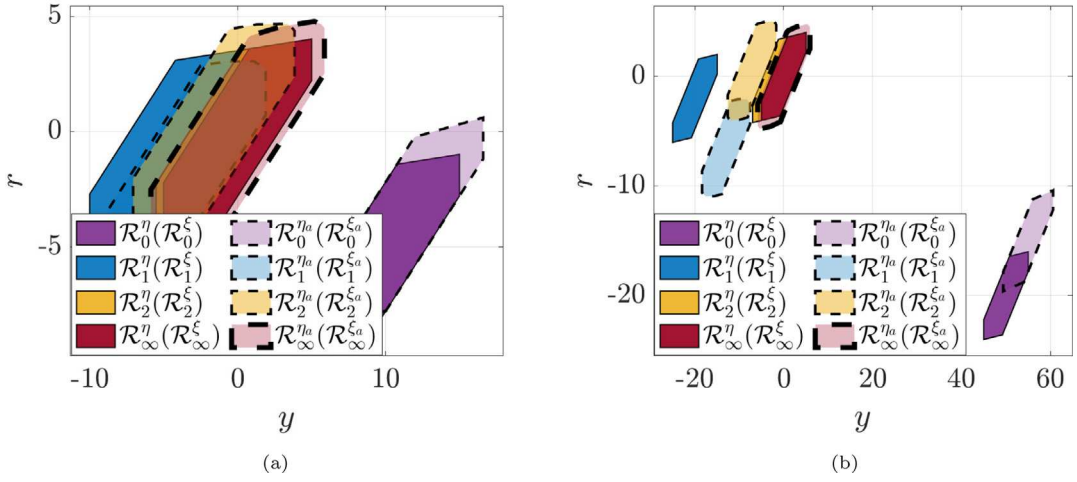

Fig. 6. Evolution of the reachable sets over a few time steps of the attack-free process and process under a multiplicative FDIA. Two initial sets are considered: (a) an initial set that is the attack-free minimum invariant set shifted by $\xi' = [10 \ \ -5]^T$ and (b) an initial set that is the attack-free minimum invariant set shifted by $\xi' = [50 \ \ -20]^T$.

attacked reachable set is not a subset of the attack-free reachable set (Fig. 6(a)). The attack, however, is detectable with respect to the second initial set because the reachable sets do not intersect at $k = 1$ and $k = 2$ (Fig. 6(b)).

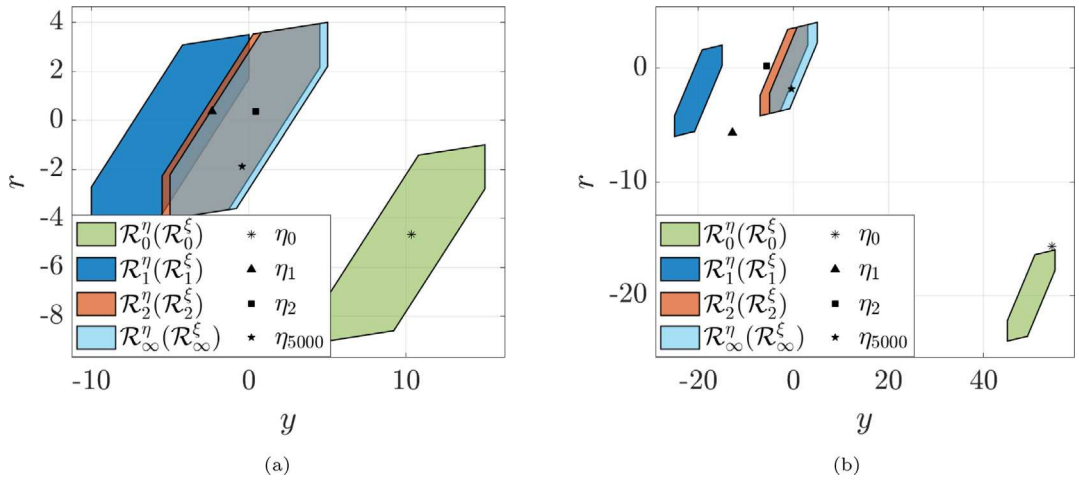

Fig. 7. The monitoring variable values and reachable sets used in the detection scheme over a few time steps for the process under a multiplicative FDIA. Two initial sets are considered: (a) an initial set that is the attack-free minimum invariant set shifted by $\xi' = [10 \ \ -5]^T$ and (b) an initial set that is the attack-free minimum invariant set shifted by $\xi' = [50 \ \ -20]^T$.

To further confirm these findings, two sets of simulations of the attacked process are performed. In the first simulation set, the process evolving from the initial condition $\xi_0 = [10 \ \ -5]^T$, contained in the first initial set, is considered. In the second simulation set, the process evolving from the initial condition $\xi_0 = [50 \ \ -20]^T$, contained in the second initial set, is considered. Each simulation set consists of 1000 simulations of the attacked process. In the first simulation set, the attack is detected over 474 of the 1000 simulations. For the simulations where the attack is detected, the first detection time ranged from $k = 1$ to $k = 4970$, indicating a range of detection times. The monitoring variable values and reachable sets used in the detection scheme are shown in Fig. 7(a) for one simulation. Over this simulation, the monitoring variable values over the time steps shown are contained within the reachable sets used in the detection scheme, and the attack is not detected. In the second simulation set, the attack is detected in all simulations at $k = 1$ and $k = 2$. The monitoring variable values and reachable sets used in the detection scheme over a few time steps are shown in Fig. 7(b) for one simulation where the attack is detected at $k = 0$, $k = 1$, and $k = 2$. These results demonstrate the dependence of the attack detectability on the initial set.

The impact of the controller and observer gains on attack detectability is also analyzed by considering process operation for two choices of the controller and observer gains: $(K, L) = (1, 0.9)$ and $(K, L) = (0.2, 1.5)$. The process evolving from an initial set that is the attack-free minimum invariant set shifted by $\xi' = [50 \ \ -20]^T$ is considered. As described above, the attack is detectable when the process is operated with $(K, L) = (1, 0.9)$. Applying the attack classification scheme, the attack is detectable or potentially detectable when the process is operated with $(K, L) = (0.2, 1.5)$. An additional 1000 simulations of the attacked process under the second choice of gains are performed. The attack is detected in 638 of the 1000 simulations. However, the attack is detected in all 1000 simulations when the process is operated with the first choice of gains. These results indicate that the choice of controller and observer gains can also influence the ability to detect attacks.

4.3. Factors influencing the detectability of an additive FDIA

In this case study, additive FDIAs (i.e., attacks with $\Lambda^y = 1$ and $\Lambda^u = 1$) are considered. First, the detectability of two additive attacks with respect to the reachable set-based detection scheme is analyzed. Next, the impact of the initial set on attack detectability is analyzed. Finally, the influence of the controller and observer gains on the detectability of an additive attack is analyzed.

The detectability of two additive FDIAs that alter the variable values over the sensor–controller and controller–actuator links is analyzed. Both attacks involve randomly varying δ_k^u and δ_k^y where both values are drawn from a uniform distribution at every time step. For the first attack, both numbers are drawn from the interval $[0, 1]$, and for the second attack, both numbers are drawn from the interval $[5, 7]$. The process is operated with $K = 1$ and $L = 0.9$ and an initial set that is the attack-free minimum invariant set shifted by $\xi' = [50 \ \ -20]^T$. The reachable sets of the attack-free process converge to its minimum invariant set at the time step $k = 7$. Fig. 8(a) and Fig. 8(b) illustrate the reachable sets of the monitoring variable for the attack-free process with respect to the reachable sets of the process under the first attack and the second attack, respectively. As illustrated in Fig. 8(a), the first attack is either detectable or potentially detectable because the attacked reachable sets intersect, but are not contained within the attack-free reachable sets at all time steps. However, the second attack is detectable because the attacked and the attack-free reachable sets do not intersect at $k = 1$ (Fig. 8(b)).

To investigate attack detectability further, two sets of closed-loop simulations of the process under an attack are performed. In the first simulation set, the process under the first attack is considered. In the second simulation set, the process under the second attack is considered. Each simulation set consists of 1000 simulations of the process. All simulations are initialized at $\xi_0 = [50 \ \ -20]^T$, which is within the initial set. The first attack is detected in all simulations, with the detection time ranging from $k = 2$ to $k = 1402$. Fig. 9(a) illustrates the attack-free reachable sets for a few time steps, and the monitoring variable values over one simulation. Over this simulation, the monitoring variable values are contained within the reachable sets from $k = 0$ to $k = 7$. The attack is detected at time step $k = 8$. On the other hand, the second attack is detected at time step $k = 1$ in all simulations. Fig. 9(b) illustrates the attack-free reachable sets and the monitoring variable values over one simulation. Over this simulation, the attack is detected at all time steps shown (i.e, from $k = 0$ to $k = 7$). The results demonstrate that an additive attack of this nature, where the attack bias is treated as a random number within a compact interval, may be detectable or potentially detectable.

Next, the impact of the initial set on the detectability of an additive attack is analyzed by considering the process operated with $K = 1$ and $L = 0.9$. The process evolving from three different initial sets is considered: first from an initial set that is the attack-free minimum invariant set shifted by $\xi_1' = [10 \ \ -10]^T$, second from an initial set that is the attack-free minimum invariant set shifted by $\xi_2' = [100 \ \ -50]^T$,

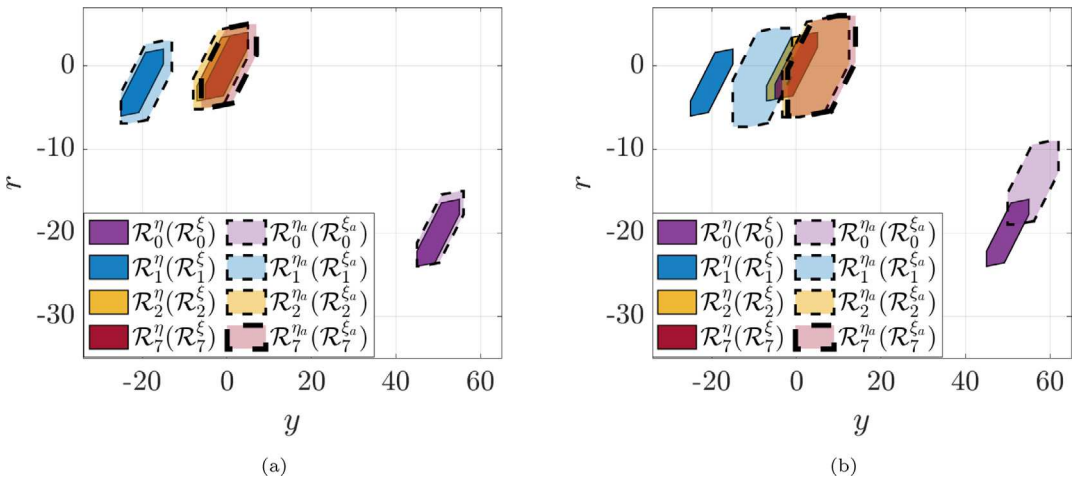

Fig. 8. Evolution of the reachable sets over a few time steps for the attack-free process with respect to the reachable sets for the process under (a) the first additive attack and (b) the second additive attack.

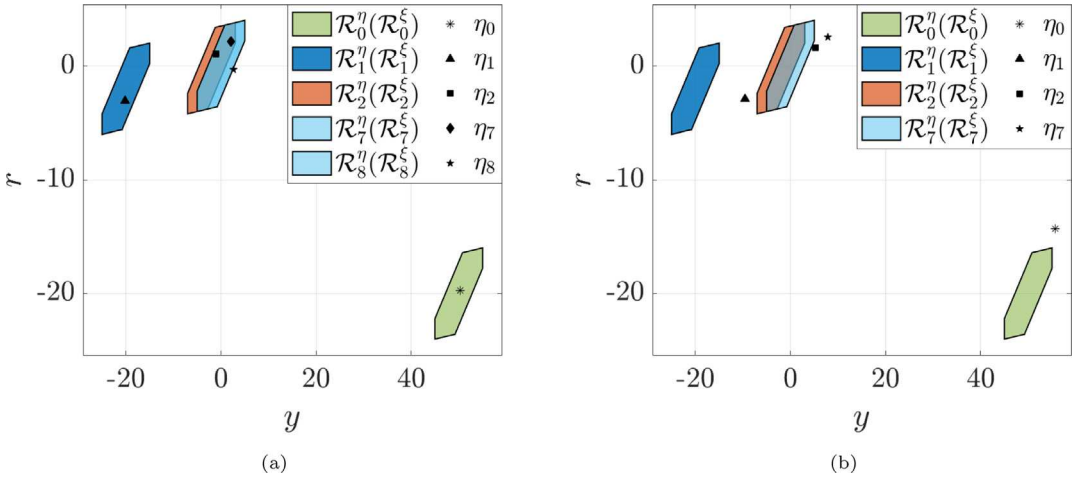

Fig. 9. The monitoring variable values and reachable sets used in the detection scheme over a few time steps for the attacked process. The monitoring variable values shown are observed over one simulation of the process under: (a) the first additive attack and (b) the second additive attack.

and third from an initial set that is the attack-free minimum invariant set shifted by $\xi_3' = [50 \ \ -20]^T$. For each initial set, the detectability of the two additive attacks considered previously is analyzed. For all three initial sets considered, the first attack where the random attack biases are bounded in $[0, 1]$ is found to be either potentially detectable or detectable. For all three initial sets considered, the second attack where the random attack biases are bounded in $[5, 7]$ is detectable, because the reachable sets of the attacked and the attack-free process do not intersect at time step $k = 1$. The results demonstrate that for the process evolving from any of the initial sets considered, the detectability of the two additive attacks is consistent.

Finally, the impact of the controller and observer gains on the detectability of an attack is explored by considering the process evolving from the initial set that is the attack-free minimum invariant set shifted by $\xi' = [50 \ \ -20]^T$ and operated with two controller and observer gains: first with $K = 0.2$ and $L = 1.5$ and second with $K = 1$ and $L = 0.9$. For the process operated with each choice of controller and observer gains, the detectability of the additive attack where the attack biases are bounded in $[5, 7]$ is analyzed. For the process operated with $K = 0.2$ and $L = 1.5$, the attack may either be detectable or potentially detectable. However, the attack on the process operated with $K = 1$ and $L = 0.9$ is detectable with the reachable sets of the attacked and the attack-free process having zero intersection at time step $k = 1$. The results demonstrate that the controller and observer gains influence the detectability of an additive FDIA.

5. Numerical results: Chemical process example

In this section, the proposed reachable set-based detection scheme, as well as the detectability-based classification of attacks, are applied to a chemical process example during transient operation. All polytope computations are performed using the MPT 3.0 toolbox (Herceg et al., 2013). We consider a chemical process example consisting of a well-mixed continuously stirred-tank reactor (CSTR) where a second-order exothermic reaction $A \rightarrow B$ occurs. Under standard modeling assumptions, the process dynamics are described by its mass and energy balances:

$$\begin{aligned} \frac{dC_A}{dt} &= \frac{F}{V}(C_{A0} + \Delta C_{A0} - C_A) - k_0 e^{\frac{-E}{RT}} C_A^2 \\ \frac{dT}{dt} &= \frac{F}{V}(T_0 + \Delta T_0 - T) - \frac{\Delta H k_0}{\rho C_p} e^{\frac{-E}{RT}} C_A^2 + \frac{Q}{\rho C_p V} \end{aligned} \tag{18}$$

where C_{A0} and T_0 are the reactant feed concentration and feed temperature, respectively; and C_A and T are the concentration of the reactant in the reactor and the temperature of the reactor, respectively. The manipulated input is the heat supplied to or removed from the reactor Q. The process is subject to bounded disturbances modeled as the variation in the concentration of the reactant A in the feed ΔC_{A0}, and the variation of the temperature of the feed to the reactor ΔT_0. The measured variables available to the controller are the concentration of the reactant C_A and the temperature of the reactant T. The process

Table 1
Process parameters of the CSTR (Alanqar et al., 2015).

Volumetric flow rate (F)	$5.0\,\mathrm{m^3\,h^{-1}}$
Reactor volume (V)	$1.0\,\mathrm{m^3}$
Feed concentration of A (C_{A0})	$4.0\,\mathrm{kmol\,m^{-3}}$
Activation energy (E)	$5.0\times10^4\,\mathrm{kJ\,kmol^{-1}}$
Pre-exponential factor (k_0)	$8.46\times10^6\,\mathrm{m^3\,h^{-1}\,kmol^{-1}}$
Gas constant (R)	$8.314\,\mathrm{kJ\,kmol^{-1}\,K}$
Feed temperature (T_0)	300 K
Density of reactor liquid hold-up (ρ)	$1000\,\mathrm{kg\,m^{-3}}$
Heat of reaction (ΔH)	$-1.15\times10^4\,\mathrm{kJ\,kmol^{-1}}$
Heat capacity (C_p)	$0.231\,\mathrm{kJ\,kg\,K^{-1}}$
Steady-state heat rate added/removed from the reactor (Q_s)	$0\,\mathrm{kJ\,h^{-1}}$
Steady-state reactant concentration (C_{As})	$1.22\,\mathrm{kmol\,m^3}$
Steady-state temperature (T_s)	438.2 K

is subject to bounded measurement noise acting on all sensors. The process disturbances are bounded such that $|\Delta C_{A0}| \leq 0.01\,\mathrm{kmol\,m^{-3}}$ and $|\Delta T_0| \leq 0.2\,\mathrm{K}$. Similarly, the measurement noise acting on the concentration sensor (v_1) is bounded as $|v_1| \leq 0.01\,\mathrm{kmol\,m^{-3}}$, and the measurement noise on the temperature sensor (v_2) is bounded as $|v_2| \leq 0.2\,\mathrm{K}$. The definitions and values of the other process parameters are listed in Table 1.

The control objective is to stabilize the closed-loop process at its open-loop stable steady–state given by $C_{As} = 1.22\,\mathrm{kmol\,m^{-3}}$, $T_s = 438\,\mathrm{K}$, and $Q_s = 0\,\mathrm{kW}$. A continuous-time linear time-invariant state–space model is obtained via linearization around the desired operating steady-state of the CSTR, and defining the deviation variables $x_1 = C_A - C_{As}$, $x_2 = T - T_s$, and $u = Q - Q_s$. Using a sampling interval of $\Delta = 1\times10^{-2}$ h, a discrete-time state–space model of the form in Eq. (1) is obtained. A monitoring variable that is the concatenation of the measured output and the residual vectors is considered, i.e., $\eta_k := [y_k^T \; r_k^T]^T$. In the case studies that follow, the process is simulated using its continuous-time nonlinear model in Eq. (18) with the control input applied in a sample-and-hold fashion. Euler's method with an integration step size of 1×10^{-4} h is used to integrate the ordinary differential equations. Two case studies are performed. In the first case study, the reachable set-based detection scheme is applied to monitor the CSTR during a transient phase induced by switching the controller and observer gains during operation. In the second case study, the detectability of a simultaneous additive and multiplicative FDIA is analyzed using the reachable set-based attack detectability classification scheme. For both case studies, the linearized process model is used to design the control law and compute the reachable sets. However, the CSTR is simulated using its nonlinear model. Therefore, the case studies presented in this section consider the application of the attack classification and the detection scheme to a nonlinear process.

5.1. Application of the reachable set-based detection scheme

In a prior work (Narasimhan et al., 2022b), controller and observer gain switching between (K_i, L_i), with controller poles at $[-0.2 \;\; -0.3]$ and observer poles at $[-0.2 \;\; -0.3]$, to (K_f, L_f), with controller poles at $[0.2 \;\; -0.1]$ and observer poles at $[0.2 \;\; 0.3]$, was considered as a way to enhance attack detection capabilities of a detection scheme monitoring the process. In this case study, gain switching occurs on the process operating initially with its states bounded in the minimum invariant set of the attack-free process under (K_i, L_i). The controller switch may induce a transient operation, so the reachable set-based detection scheme is applied to monitor the process. The forward reachable sets of the attack-free process from the minimum invariant set of the attack-free process under (K_i, L_i), which is taken to be the initial set $\mathcal{R}_0^\xi$, are computed, and the reachable sets converge to the minimum invariant set of the attack-free process under (K_f, L_f), which is denoted by $\mathcal{R}_\infty^\xi$. To design the reachable set-based detection scheme, the reachable sets of the monitoring variable for the attack-free process are computed from the initial set $\mathcal{R}_0^\eta(\mathcal{R}_0^\xi)$ to its terminal set $\mathcal{R}_\infty^\eta(\mathcal{R}_\infty^\xi)$. Two sets of simulations are considered. The first set considers the attack-free process, and the second set considers the process under a multiplicative sensor–controller link attack of magnitude $\Lambda^y = \mathrm{diag}(1, 0.85)$. Each simulation set consists of 1000 simulations of the process, and each simulation has a total length of 5 h, spanning 500 time steps. All simulations are initialized with $\xi_0 = [0.005\;5 \;\; -0.01\;\; 0.2]^T \in \mathcal{R}_0^\xi \setminus \mathcal{R}_\infty^\xi$.

No attacks are detected using the detection scheme when monitoring the attack-free process. Fig. 10(a) and Fig. 10(b) illustrate the output and the residual values over a few time steps for one of the simulations of the attack-free process. The monitoring variable values are contained within their corresponding reachable sets for all time. At 0.02 h ($k = 2$), the monitoring variable values converge to the terminal set for the attack-free process, where they remain. As a result, the reachable set-based detection scheme generates an output of 0 for all time steps in the simulation.

Considering the process under a multiplicative attack, the attack is detected in 854 out of the 1000 simulations. The detection times ranged from 0.01 h ($k = 1$) to 4.59 h ($k = 459$) for the simulations where the attack is detected. Fig. 11(a) and Fig. 11(b) illustrate the output and residual values over a few time steps over a simulation of the attacked process. From Fig. 11(b), the attack is detected at 0.01 h ($k = 1$). These simulations demonstrate that the reachable set-based detection scheme can monitor the nonlinear process during transient operation without raising false alarms for the attack-free process, and can successfully detect attacks on a nonlinear process.

Remark 6. In a prior work (Narasimhan et al., 2022b), a controller and observer gain switching was utilized to enable attack detection on the nonlinear CSTR process monitored by a terminal set-based detection scheme. However, the attack detection method presented previously has a non-zero false alarm rate because the terminal set-based detection scheme is not designed to account for transient operation. Based on the results in this section, the reachable set-based detection scheme proposed in this work may be used to eliminate false alarms in the controller and observer gain switching-based attack detection method.

5.2. Application of detectability-based classification of an attack

In this case study, the ability to classify attacks using the reachability analysis is demonstrated for the nonlinear CSTR. Specifically, the detectability of a simultaneous multiplicative and additive FDIA that alters the data over both the sensor–controller and controller–actuator links is analyzed. The attack parameters are $\Lambda^y = \mathrm{diag}(1, 0.85)$, $\Lambda^u = 0.9$, $\delta_k^{y_{C_A}} \in [0.1, 0.2]\;\mathrm{kmol\,m^{-3}}$, and $\delta_k^{y_T} \in [0.1, 0.2]\;\mathrm{K}$. The parameters $\delta_k^{y_{C_A}}$ and $\delta_k^{y_T}$ are the attack biases added to the concentration and temperature measurements, respectively, and are modeled as random variables drawn from a uniform distribution. The process is operated with controller and observer gains selected via pole placement using the linearized process model and by placing the poles at $[-0.2 \;\; -0.3]$ to determine K and $[-0.2 \;\; -0.3]$ to determine L. For the attack-free process, the minimum invariant set of the closed-loop system is the initial set, so the terminal set of the monitoring variable is the k-step forward reachable set for all time steps $k \in \mathbb{Z}^+$.

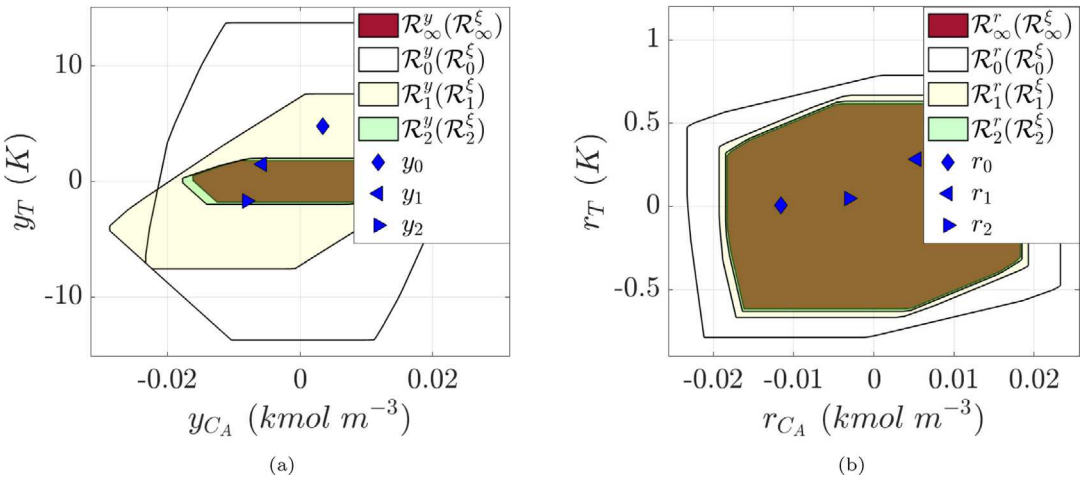

Fig. 10. The monitoring variable values, including (a) the output values and (b) the residual values, and reachable sets used in the detection scheme over a few time steps for the attack-free process. In this case, there are no false alarms. The brown central region represents the intersection of all reachable sets shown.

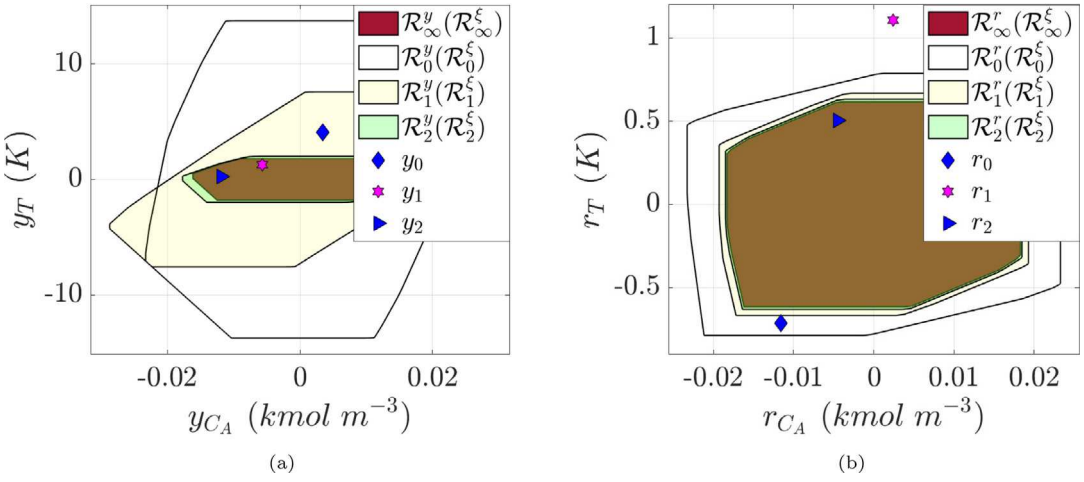

Fig. 11. The monitoring variable values, including (a) the output values and (b) the residual values, and reachable sets used in the detection scheme over a few time steps for the attacked process. In this case, the attack is detected at $k = 1$. The brown central region represents the intersection of all reachable sets shown.

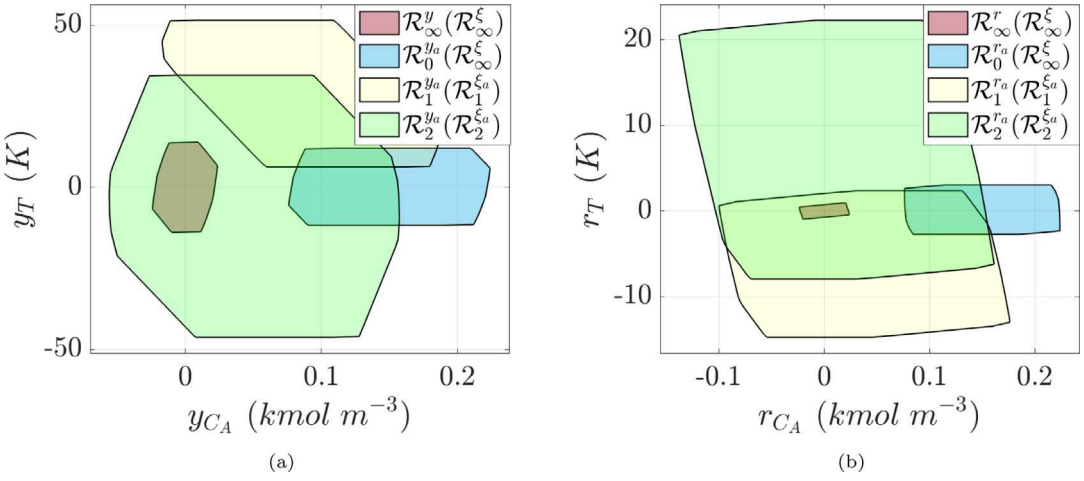

Fig. 12. The reachable sets for (a) the measured output and (b) the residual for the CSTR under an attack.

The closed-loop process under the FDIA is unstable ($\max_i|\lambda_i(A^{\xi_a})| = 1.1371 > 1$). The reachable sets of the attacked process are compared to the terminal set of the attack-free process to classify the attack. Fig. 12(a) illustrates the reachable sets of the measured output for the attacked process for a few time steps and the measured output terminal set for the attack-free process. Fig. 12(b) illustrates the reachable sets of the residual for the attacked process and the residual terminal set for the attack-free process. As illustrated, the attack is detectable with

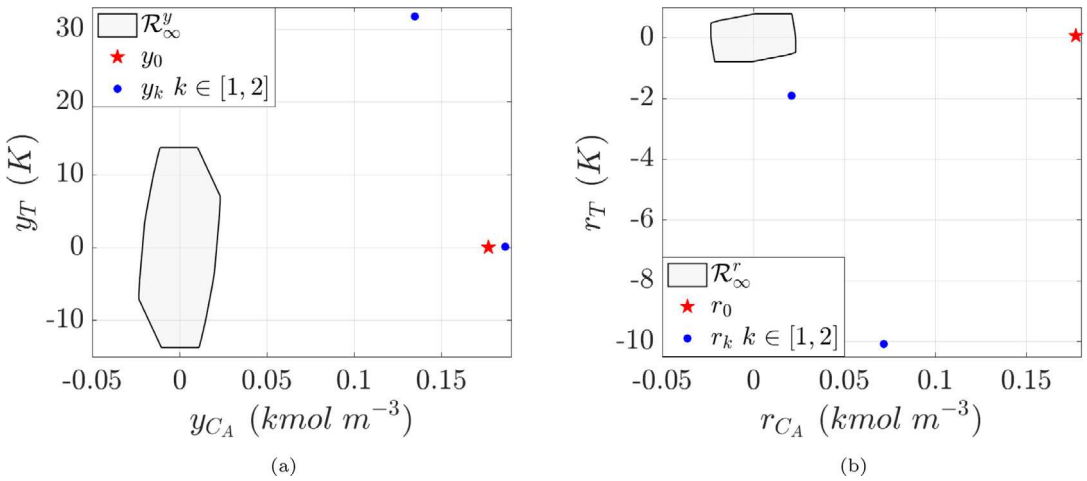

Fig. 13. The values of (a) the measured output and (b) the residual and their corresponding terminal sets over one simulation of the CSTR process under an attack.

respect to the initial set because the reachable set of the attacked process and the terminal set of the attack-free process do not intersect at time step $k = 0$.

Two simulation sets are performed to confirm that the reachability analysis correctly classified the attack. The attack-free process is considered in the first set, and the attacked process is considered in the second set. Each set consists of 1000 simulations of the process, and each simulation simulates the CSTR over a 5 h period (total of 500 time steps). All simulations are initialized with the augmented state at the origin. The detection scheme in Eq. (14) is designed to monitor the process with respect to the reachable sets, which are the terminal set of the attack-free process ($\mathcal{R}^\eta_k(\mathcal{R}^\xi_k) = \mathcal{R}^\eta_\infty(\mathcal{R}^\xi_\infty)$ for all time steps $k \in \mathbb{Z}^+$).

For the attack-free simulations, the detection scheme does not raise any alarms. For the simulations of the attacked process, the attack is detected at $k = 0$ in all simulations, as expected from the reachability analysis. The measured output and the residual (monitoring variable) values for the attacked process over one simulation are shown in Figs. 13(a) and 13(b), respectively. Over this simulation, the monitoring variable values evolve outside the terminal set of the attack-free process but stay within the attacked process reachable sets. The attack is detected at the first three time steps. The results demonstrate that the detectability classification based on reachable sets can be applied to classify attacks for the nonlinear CSTR.

Remark 7. For the attack-free process, the terminal set of the monitoring variable is the forward reachable set for all time steps $k \in \mathbb{Z}^+$. Therefore, the terminal set-based detection scheme is a special case of the reachable set-based detection scheme with $\mathcal{R}^\eta_k(\mathcal{R}^\xi_k) = \mathcal{R}^\eta_\infty(\mathcal{R}^\xi_\infty)$ (for all $k \in \mathbb{Z}^+$) in Eq. (14), and the reachable set-based classification of attacks presented in Section 3 can be used to analyze attack detectability for a process monitored by the terminal set-based detection scheme.

6. Conclusions

A reachable set-based detection scheme was proposed to monitor dynamic processes under false data injection attacks targeting the sensor–controller and controller–actuator communication links. A rigorous characterization of the conditions that render an attack to be undetectable or detectable with respect to the proposed detection scheme was presented. An approach for classifying attacks based on their detectability with respect to the reachable set-based detection scheme was presented. The proposed detection scheme was applied to two illustrative examples. The detectability of various attacks was analyzed, and the applicability of the detection scheme and classification method to monitor and classify attacks on a nonlinear chemical process was demonstrated.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Acknowledgements

Financial support from the National Science Foundation CBET-2137281 is gratefully acknowledged.

References

Ahmed, C.M., Palleti, V.R., Mishra, V.K., 2022. A practical physical watermarking approach to detect replay attacks in a CPS. J. Process Control 116, 136–146. http://dx.doi.org/10.1016/j.jprocont.2022.06.002.

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

Chen, S., Wu, Z., Christofides, P.D., 2020. Cyber-attack detection and resilient operation of nonlinear processes under economic model predictive control. Comput. Chem. Eng. 136, 106806. http://dx.doi.org/10.1016/j.compchemeng.2020.106806.

Chen, S., Wu, Z., Christofides, P.D., 2021. Cyber-security of centralized, decentralized, and distributed control-detector architectures for nonlinear processes. Chem. Eng. Res. Des. 165, 25–39. http://dx.doi.org/10.1016/j.cherd.2020.10.014.

Cómbita, L.F., Quijano, N., Cárdenas, Á.A., 2022. On the stability of cyber-physical control systems with sensor multiplicative attacks. IEEE Access 10, 39716–39728. http://dx.doi.org/10.1109/ACCESS.2022.3164424.

Duo, W., Zhou, M., Abusorrah, A., 2022. A survey of cyber attacks on cyber physical systems: Recent advances and challenges. IEEE/CAA J. Autom. Sin. 9, 784–800. http://dx.doi.org/10.1109/JAS.2022.105548.

Girard, A., Guernic, C.L., Maler, O., 2006. Efficient computation of reachable sets of linear time-invariant systems with inputs. In: Proceedings of the International Workshop on Hybrid Systems: Computation and Control. Santa Barbara, CA, pp. 257–271. http://dx.doi.org/10.1007/11730637_21.

Hashemi, N., German, E.V., Pena Ramirez, J., Ruths, J., 2019. Filtering approaches for dealing with noise in anomaly detection. In: Proc. IEEE Conf. Decis. Control. Nice, France, pp. 5356–5361. http://dx.doi.org/10.1109/CDC40024.2019.9029258.

Herceg, M., Kvasnica, M., Jones, C.N., Morari, M., 2013. Multi-Parametric Toolbox 3.0. In: Proceedings of the European Control Conference. Zürich, Switzerland, pp. 502–510. http://dx.doi.org/10.23919/ECC.2013.6669862.

Kwon, C., Hwang, I., 2018. Reachability analysis for safety assurance of cyber-physical systems against cyber attacks. IEEE Trans. Automat. Control 63, 2272–2279. http://dx.doi.org/10.1109/tac.2017.2761762.

Liu, H., Mo, Y., Johansson, K.H., 2021. Active detection against replay attack: A survey on watermark design for cyber-physical systems. In: Lect. Notes Control. Inf. Sci, Springer, pp. 145–171. http://dx.doi.org/10.1007/978-3-030-65048-3_8.

Mo, Y., Sinopoli, B., 2009. Secure control against replay attacks. In: Proc. Annu. Allerton Conf. Commun. Control Comput. Monticello, Illinois, USA, pp. 911–918. http://dx.doi.org/10.1109/ALLERTON.2009.5394956.

Mo, Y., Sinopoli, B., 2012. Integrity attacks on cyber-physical systems. In: Proc. IEEE Int. Conf. Intell. Comput. Beijing, China, pp. 47–54. http://dx.doi.org/10.1145/2185505.2185514.

Murguia, C., Van de Wouw, N., Ruths, J., 2017. Reachable sets of hidden CPS sensor attacks: Analysis and synthesis tools. In: Proc. of the IFAC World Congress. Toulouse, France, pp. 2088–2094. http://dx.doi.org/10.1016/j.ifacol.2017.08.528.

Na, G., Eun, Y., 2018. A multiplicative coordinated stealthy attack and its detection for cyber physical systems. In: Proc. IEEE Conf. Cont. Techn. Appl. Copenhagen, Denmark, pp. 1698–1703. http://dx.doi.org/10.1109/ccta.2018.8511631.

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

Narasimhan, S., El-Farra, N.H., Ellis, M.J., 2022a. Active multiplicative cyberattack detection utilizing controller switching for process systems. J. Process Control 116, 64–79. http://dx.doi.org/10.1016/j.jprocont.2022.05.014.

Narasimhan, S., El-Farra, N.H., Ellis, M.J., 2022b. A control-switching approach for cyberattack detection in process systems with minimal false alarms. AIChE J. 68, e17875. http://dx.doi.org/10.1002/aic.17875.

Oyama, H., Messina, D., Rangan, K.K., Durand, H., 2022. Lyapunov-based economic model predictive control for detecting and handling actuator and simultaneous sensor/actuator cyberattacks on process control systems. Front. Chem. Eng. 4, 810129. http://dx.doi.org/10.3389/fceng.2022.810129.

Oyama, H., Rangan, K.K., Durand, H., 2021. Handling of stealthy sensor and actuator cyberattacks on evolving nonlinear process systems. J. Adv. Manuf. Process. 3, e10099. http://dx.doi.org/10.1002/amp2.10099.

Raković, S.V., Kerrigan, E.C., Kouramas, K.I., Mayne, D.Q., 2005. Invariant approximations of the minimal robust positively invariant set. IEEE Trans. Automat. Control 50, 406–410. http://dx.doi.org/10.1109/tac.2005.843854.

Rangan, K.K., Oyama, H., Durand, H., 2021. Integrated cyberattack detection and handling for nonlinear systems with evolving process dynamics under Lyapunov-based economic model predictive control. Chem. Eng. Res. Des. 170, 147–179. http://dx.doi.org/10.1016/j.cherd.2021.03.024.

Renganathan, V., Gravell, B.J., Ruths, J., Summers, T.H., 2022. Anomaly detection under multiplicative noise model uncertainty. IEEE Control Syst. Lett. 6, 1873–1878. http://dx.doi.org/10.1109/LCSYS.2021.3134944.

Trapiello, C., Puig, V., 2020. Input design for active detection of integrity attacks using set-based approach. In: Proc. of the IFAC World Congress. Berlin, Germany, pp. 11094–11099. http://dx.doi.org/10.1016/j.ifacol.2020.12.254.

Umsonst, D., Ruths, J., Sandberg, H., 2022. Finite sample guarantees for quantile estimation: An application to detector threshold tuning. IEEE Trans. Control Syst. Technol. 1–8. http://dx.doi.org/10.1109/TCST.2022.3199668.

Wu, Z., Christofides, P.D., 2021. Process Operational Safety and Cybersecurity: A Feedback Control Approach. Springer, http://dx.doi.org/10.1007/978-3-030-71183-2.

Zedan, A., El-Farra, N.H., 2021. A machine-learning approach for identification and mitigation of cyberattacks in networked process control systems. Chem. Eng. Res. Des. 176, 102–115. http://dx.doi.org/10.1016/j.cherd.2021.09.016.