2022 American Control Conference
Atlanta, USA, June 8-10, 2022

Controller Switching-Enabled Active Detection of Multiplicative Cyberattacks on Process Control Systems

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

Abstract— **This work focuses on the problem of enhancing cyberattack detection capabilities in process control systems subject to multiplicative cyberattacks. First, the relationship between closed-loop stability and attack detectability with respect to a class of residual-based detection schemes is rigorously analyzed. The results are used to identify a set of controller parameters (called "attack-sensitive" controller parameters) under which an attack can destabilize the closed-loop system. The selection of attack-sensitive controller parameters can enhance the ability to detect attacks, but can also degrade the performance of the attack-free closed-loop system. To balance this trade-off, a novel active attack detection methodology employing controller parameter switching between the nominal controller parameters (chosen on the basis of standard control design criteria) and the attack-sensitive controller parameters, is developed. The proposed methodology is applied to a chemical process example to demonstrate its ability to detect multiplicative sensor-controller communication link attacks.**

I. Introduction

Traditionally, information technology (IT) infrastructure is responsible for the cybersecurity of industrial control systems (ICSs). However, the increased reliance on networked components in recent years has made ICSs targets for frequent and increasingly complex cyberattacks [1], [2], [3], demonstrating that IT-based approaches alone are insufficient. Consequently, as an additional layer of protection against cyberattacks, the integration of cybersecurity considerations into traditional ICS design methods has received increased attention (e.g., [4], [5] and the references therein).

Process control systems (PCSs) are ICSs designed to control chemical manufacturing processes and may be particularly attractive to attackers due to their inherent risks. An attack on a PCS may maliciously modify or falsify either the control law or the data transmitted over the PCS communication links. The design of a successful communication link attack resulting in loss of closed-loop stability while remaining undetected may require some process knowledge on the part of the attacker [6]. Irrespective of the attacker's process knowledge, stealthy attacks may be designed to be detrimental to the process while evading detection by the anomaly detection schemes monitoring the process. These attacks are challenging to detect because the behavior of the attacked process is difficult to differentiate from the behavior of the attack-free process [7]. Consequently, the design of detection schemes aimed at monitoring a process for several types of stealthy attacks on PCSs has received significant attention [8], [9], [10], [11], [12], [13], [14].

Attack detection schemes may be grouped into two categories: passive attack detection schemes and active attack detection schemes. Passive attack detection schemes monitor a process for anomalies without employing external intervention or inducing a perturbation in the process (e.g., [8], [9], [14]). Depending on the cyberattack design, passive approaches for attack detection may not always be successful in detecting an attack (Sec. III, [10]). As an alternative, active detection schemes that involve external intervention to actively probe for cyberattacks may be used. Examples of such approaches include watermarking and moving target approaches [10], [11], [12], [13].

This work focuses on the detection of multiplicative sensor-controller link cyberattacks. These cyberattacks may be designed to be stealthy without requiring intimate knowledge of process dynamics or without simultaneously targeting the actuator-controller communication link. In a prior work [15], the connection between the PCS design and the detectability of multiplicative sensor-controller link attacks was analyzed. The analysis revealed that some controller parameter choices could prevent attack detection. Consequently, a controller screening methodology to identify and discard controller parameters that mask the impact of an attack from the detection scheme was proposed. However, an attack may still go undetected even when the controllers are chosen so that they do not mask the impact of an attack.

In the present work, the connection between attack detectability with respect to a residual-based detection scheme and closed-loop stability is analyzed. The analysis reveals that it is advantageous to select controller parameters that render the attacked closed-loop process unstable. However, operating with these "attack-sensitive" controller parameters may not be desirable from the standpoint of the performance of the attack-free closed-loop system. To balance the trade-off between the enhancement in attack detection capability and the possible degradation in closed-loop performance, an active attack detection methodology that employs occasional controller parameter switching between the nominal and attack-sensitive parameters is developed and demonstrated using a chemical process example.

II. Preliminaries

A. Notation and Definitions

The closed Euclidean ball and infinity ball centered at the origin with radius $R > 0$ are denoted by $B^n(R) := \{x \in$

Shilpa Narasimhan, Nael H. El-Farra, and Matthew J. Ellis are with the Department of Chemical Engineering, University of California, Davis, CA 95616, USA. Emails: shnarasimhan@ucdavis.edu, nhelfarra@ucdavis.edu, and mjellis@ucdavis.edu. Corresponding author: M. J. Ellis. Financial support from the College of Engineering at UC Davis is gratefully acknowledged.

$\mathbb{R}^n \mid \|x\| \leq R\}$ and $B_\infty^n := \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq R\}$. 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 $X \oplus Y = \{x + y \mid x \in X, y \in Y\}$. 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 $t \geq 0$.

B. Process Model and Control System

We consider processes modeled by discrete-time linear time-invariant systems subject to bounded disturbances:

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

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. In this work, multiplicative sensor-controller communication link attacks are considered. The value of the measured output received by the controller 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 representing the magnitude of the multiplicative sensor-controller link attack on the process, where $\Lambda = I$ indicates the absence of an attack and $\Lambda = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_{n_y}) \neq I$ indicates the presence of an attack. Here, $\alpha_i \neq 1$ is the attack magnitude on the i^{th} sensor-controller link. The matrix Λ is referred to as the magnitude of the attack. The matrix C is of appropriate dimensions.

The pair (A, B) is assumed to be controllable, and the 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. To steer the state to the origin, which is assumed to be the desired operating steady-state, 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. The gain is 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:

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

Defining the augmented state vector as $\xi(t) = [x^T(t)\ e^T(t)]^T$, the overall closed-loop system consisting of the process in (1) with the feedback control law in (4) using the estimated state from the observer in (3) can be written as:

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

where $d(t) := \begin{bmatrix} w^T(t) & v^T(t) \end{bmatrix}^T \in F$ is the augmented vector of disturbances and measurement noise, and $F = \left\{ \begin{bmatrix} w \\ v \end{bmatrix} \mid w \in W, v \in V \right\}$ is the set of disturbances. Here, $\xi(t)$ denotes the augmented state and $A_\xi(\Lambda, K, L)$ and $B_\xi(\Lambda, L)$ are the closed-loop system matrices for the augmented state dynamics. In what follows, 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 of time, 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 system has converged to its terminal set, which is the minimum invariant set for the closed-loop system. The minimum invariant set for the augmented closed-loop system in (6) such that $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| < 1$ is the solution to the infinite Minkowski sum [16]:

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

Based on (7), the minimum invariant sets are 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 process in (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 the remainder, the term closed-loop process refers to the process in (1)-(2) under the feedback law given by (4) using the estimates of states generated by the observer in (3).

III. Active Attack Detection via Controller Parameter Switching

In this section, a residual-based detection scheme is introduced, followed by detectability-based classifications of attacks. Several results relating the detectability-based classifications to closed-loop stability are then discussed, and the active detection strategy, using controller parameter switching, is presented. Due to space constraints, proofs of the theoretical results are omitted.

A. 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 and its estimate generated by the observer, i.e., $r(t) := y(t) - \hat{y}(t)$. Writing the residual in terms of the augmented state vector, $\xi(t)$, and the disturbance vector, $d(t)$, yields:

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

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 [16] and compact (Sec. 4 in [17]), and the residual is ultimately bounded within a terminal set. From (8), the residual terminal set is equal to the Minkowski sum of the projections of the minimum invariant set, $D_\xi(\Lambda, K, L)$, and the disturbance set, F, onto the residual space, and 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 $d(t) \in F$, all possible realizations of the residual will be contained within its terminal residual set, i.e., $r(t) \in D_r(\Lambda, K, L)$.

Residual-based anomaly detection schemes are commonly used in model-based detection schemes for process monitoring [18], [19], [20], [21], [22]. A set membership-based detection scheme of the form:

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

is considered in the present work. Here $z(t)$ 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. With the detector, attacks can be classified into three categories based on the ability to detect them. These definitions are specific to the detection scheme considered. While one could consider more general systems-theoretic definitions of attack detectability, the following definitions enable the classification of attacks as detectable and undetectable for the detection scheme considered.

For the closed-loop process monitored by the detection scheme in (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 $z(t_d) = 1$. With respect to the detection scheme in (10), an attack is defined as a detectable attack, if it is detected in finite time for all $\xi(0) \in D_\xi(\Lambda, K, L)$ and $\mathbf{d} \in \mathcal{F}$. An attack is defined as an undetectable attack, 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 $\xi(0) \in D_\xi(\Lambda, K, L)$. Finally, an attack is defined as potentially detectable, if the attack is neither detectable nor undetectable. 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 $\mathbb{R}^{2n_x}$. The set of initial conditions considered is $D_\xi(\Lambda, K, L)$ because steady-state operation is considered. It is possible that, for some initial conditions in $D_\xi(\Lambda, K, L)$ and realizations of $d(0) \in F$, the attack is detected immediately by the detection scheme in (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 (10) are valid for any attack, multiplicative sensor-controller link attacks are considered in the present work.

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

Proposition 1. *Consider the closed-loop process operated at steady-state with the controller-observer parameter pair (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 (10), if and only if $D_r(\Lambda, K, L) \subseteq D_r(I, K, L)$.*

Proposition 2 establishes the relationship between potential detectability and both stability and instability of the closed-loop process with controller-observer parameter pair (K, L), and the possible realizations of the disturbances. This proposition implies that even when the closed-loop process with the pair (K, L) under a multiplicative sensor controller link attack of magnitude Λ is unstable with $\max_i |\lambda_i(A_\xi(\Lambda, K, L))| > 1$, attack detection is not guaranteed for all $\mathbf{d} \in \mathcal{F}$.

Proposition 2. *Consider the closed-loop process operated at steady-state with the controller-observer parameter pair (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 $D_r(\Lambda, K, L) \not\subseteq D_r(I, K, L)$, or (2) the attacked closed-loop process is unstable with $\max_i |\lambda_i(A_\xi(\Lambda, K, L)| > 1$, then the attack is potentially detectable with respect to the detection scheme in (10).*

Proposition 2, indicates that a multiplicative attack may be only potentially detectable even when the closed-loop process is unstable in the sense that some eigenvalues of $A_\xi(\Lambda, K, L)$ lie outside the unit circle (i.e., instability does not imply detectability). The attack may not be detected for some trivial solutions generated from the attacked closed-loop process, such as the one generated when $\xi(0) = 0$ and $\mathbf{d} \equiv 0$. Additionally, pathological cases are possible where the disturbance d effectively acts as stabilizing input, resulting in the closed-loop augmented state remaining bounded.

Under an observability condition, if the augmented state under an attack is such that $\|\xi(t)\| \to \infty$ as $t \to \infty$ the attack will be detected in finite-time. This result is formally stated in Proposition 3.

Proposition 3. *Consider the closed-loop process with the controller-observer parameter pair (K, L) under a multiplicative attack of magnitude $\Lambda \neq I$. Let the pair (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, then the attack is detected in finite time with respect to the detection scheme in (10).*

B. Active Attack Detection Methodology

Proposition 3 suggests that choosing controller parameters that render the closed-loop system unstable in the presence of an attack is potentially advantageous from the standpoint of enhancing the monitoring scheme's ability to detect the attack. However, sustained operation with these controller parameters may not be desirable because the resulting closed-loop performance could be worse than that achieved under the nominal controller parameters.

To strike a balance between the two conflicting objectives, the proposed active detection methodology involves occasional switching between the nominal controller parameters, determined by traditional design approaches (e.g., [23], [24]), and the parameters that are sensitive to cyberattacks. Attack-sensitive controller parameters with respect to a multiplicative attack of magnitude Λ are defined as parameters that maintain the closed-loop stability of the attack-free process, but in the presence of the attack, the closed-loop process is rendered unstable. Additionally, they are such that the pair $(A_\xi(\Lambda, K^*, L^*), A_r(\Lambda)))$ is observable. The attack-sensitive controller parameters exploit the dependence of the terminal residual set on the controller parameters. The presented approach has two design considerations to address the potential trade-off between attack detection and closed-loop performance, including (1) how often to switch to the attack-sensitive controller parameters, and (2) the duration of the window of operation under the attack-sensitive controller parameters. The nominal controller parameters are denoted by (K^*, L^*), while the attack-sensitive controller parameters are denoted by (K_Λ, L_Λ).

From the definition of attack-sensitive controller parameters, some eigenvalues of the augmented system matrix lie outside the unit circle, i.e., $\max_i |\lambda_i(A_\xi(\Lambda, K_\Lambda, L_\Lambda))| > 1$. Even though the attacked closed-loop process with attack-sensitive controller parameters is unstable, it is possible that some realizations of the disturbances could stabilize the closed-loop process under an attack. For this reason, multiplicative sensor-controller link attacks are potentially detectable under attack-sensitive controller parameters (Proposition 2). From a practical point of view, disturbances do not normally act in a way that stabilizes the closed-loop system because they are exogenous perturbations.

Under the proposed 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. The detection scheme is therefore 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 anomaly-free operation and $z(t) = 1$ indicates anomalous process operation. For the closed-loop process with the nominal parameter pair $(K(t), L(t)) = (K^*, L^*)$, while for the closed-loop process with the attack-sensitive parameter pair, $(K(t), L(t)) = (K_\Lambda, L_\Lambda)$.

The attack-sensitive controller parameters are chosen to be sensitive to a (preferably large) range of attack magnitudes. Under the proposed active detection scheme, the controller parameter switching occurs occasionally. Considering that a switch from the nominal pair, (K^*, L^*), to the attack-sensitive pair, (K_Λ, L_Λ), occurs at t_s, the process is operated in the "attack-sensitive mode" for a certain period of time, $T_c > 0$. In the absence of an attack, the residual trajectory after the switch is expected to evolve in $D_r(I, K_\Lambda, L_\Lambda)$. At the end of the chosen time window, $[t_s, t_s + T_c]$, the controller parameters are switched back to the nominal controller parameters. In the presence of a multiplicative attack, the residual trajectory may evolve outside $D_r(I, K_\Lambda, L_\Lambda)$ for some time resulting in a detection alarm by the detection scheme in (10). If there is a detection of an attack at some time t_d, attack identification and mitigation strategies could be employed to cope with the attack, but these strategies are beyond the scope of the present work. Since attack detection cannot be guaranteed in general over the period T_c, an attack may go undetected over the period that the process operates in the attack-sensitive mode. While a more rigorous assessment is needed, it is expected that the probability of detection would increase with increasing T_c.

Based on these considerations, the implementation of the active detection methodology is initiated with a controller parameter switch from (K^*, L^*) to (K_Λ, L_Λ) implemented on the closed-loop process at some time $t_s \geq 0$. If $r(t_d) \notin D_r(I, K_\Lambda, L_\Lambda)$, an attack is detected at $t = t_d$. Following this, attack identification and mitigation strategies may be implemented. If no attack is detected for the duration of operation of the closed-loop process with (K_Λ, L_Λ), then at time $t = t_s + T_c$, a controller parameter switch from (K_Λ, L_Λ) to (K^*, L^*) is implemented on the closed-loop process.

IV. Application to a Chemical Process Example

A chemical process consisting of a continuously stirred tank reactor (CSTR) is considered, where a second order, exothermic reaction of the form $A \to B$ 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 a submerged heat exchanger coil. A dynamic process model is obtained from the appropriate mass and energy balances under standard assumptions, and is given by the following

system of ordinary differential equations:

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

The parameter definitions and the simulation data are as described in Section 4.2 in [15]. For the simulation study, a discrete LTI model (of the form of Eq 1) is dervied for the process in Eq. 12, with $A = \begin{bmatrix} 0.7364 & -0.0041 \\ 10.6953 & 1.1560 \end{bmatrix}$, $B = \begin{bmatrix} -9.0708 \times 10^{-8} \\ 4.6741 \times 10^{-5} \end{bmatrix}$, and $G = \begin{bmatrix} 0.0433 & -0.0001 \\ 0.2724 & 0.0540 \end{bmatrix}$. For the calculation of the minimum invariant and residual terminal sets, the Multi-Parametric Toolbox (MPT) 3.0 [25] is used. Numerical approximations of the minimum invariant sets are computed based on the algorithm in [26] with an error bound of 5×10^{-5}. In comparing the numerical estimates of the terminal residual sets $D_r(I, K^*, L^*)$ and $D_r(\Lambda, K^*, L^*)$, the technique presented in [15] is used.

The nominal controller parameters (K^*, L^*) are chosen to ensure closed-loop stability using pole placement with the controller poles placed at $[0.5\ -0.1]$, and the observer poles placed at $[0.48\ 0.48]$. It was verified, based on extensive closed-loop simulations, that the resulting control system is stabilizing when applied to the nonlinear process. Furthermore, the nominal controller parameters are not attack-sensitive to multiplicative sensor-controller link attacks of magnitudes in the set $\{\text{diag}(1, \alpha) \mid \alpha \in [0.1, 0.8]\}$.

In the simulations below, an attack of magnitude 0.8 on the temperature sensor (i.e., $\Lambda = \text{diag}(1, 0.8)$) is considered. In the presence of such an attack, the maximum absolute eigenvalue of the closed-loop process with (K^*, L^*) is $\max_i |\lambda_i(A_\xi(\Lambda, K^*, L^*))| = 0.6843 < 1$, indicating that closed-loop stability is maintained under the attack. The estimates of the terminal set of the residuals of the attack-free closed-loop process with (K^*, L^*) $(D_r(I, K^*, L^*))$, and for the attacked closed-loop process with (K^*, L^*) $(D_r(\Lambda, K^*, L^*))$, are computed and it is observed that $D_r(\Lambda, K^*, L^*) \not\subseteq D_r(I, K^*, L^*)$. Thus, the attack is potentially detectable with this choice of parameters (Proposition 2). However, the two residual sets are found to be almost equal, and the attack may be difficult to detect. One hundred closed-loop simulations of the attacked process operated with (K^*, L^*) are performed by varying the seed of the random number generator used to generate the process disturbance and measurement noise for each simulation. The attack is not detected over all these simulations.

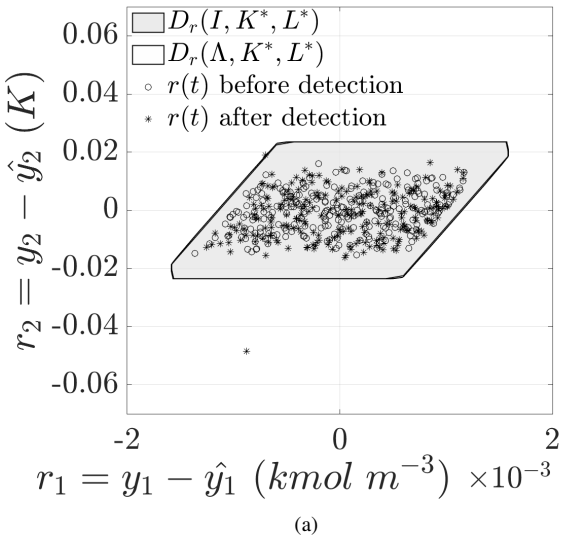
The attack-sensitive controller parameters (K_Λ, L_Λ) are selected such that the controller poles are placed at $[-0.4\ 0.1]$, and the observer poles are placed at $[0.3\ -0.1]$. In the presence of an attack of magnitude $\Lambda = \text{diag}(1, 0.8)$, the maximum absolute eigenvalue for the closed-loop process is $\max_i |\lambda_i(A_\xi(\Lambda, K_\Lambda, L_\Lambda))| = 1.1937 > 1$, indicating that the parameters are sensitive with respect to this attack. Furthermore, the matrix pair $(A_\xi(\Lambda, K_\Lambda, L_\Lambda), A_r(\Lambda))$ is observable. In addition to an attack of magnitude $\Lambda = \text{diag}(1, 0.8)$, the attack-sensitive controller parameters are found to be sensitive to attacks of magnitudes contained in the set $\{\text{diag}(1, \alpha) \mid \alpha \in [0.1, 0.8]\}$. The controller parameter switch for active detection methodology implementation occurs at $t_s = 2.5\,\text{h}$ and the detection cycle is chosen to be $T_c = 1\,\text{h}$. The proposed active attack detection methodology is applied to the CSTR for one hundred closed-loop simulations with the same realizations of the random variables as the previous set of simulations. Over all these simulations, implementing the active detection methodology resulted in the detection of the attack within a maximum of 10 time steps from the controller parameter switch.

In what follows, results from two simulation case studies with the proposed active detection strategy are discussed. The first case considers an attack-free closed-loop simulation. With the nominal controller parameters (K^*, L^*), $r(t)$ is maintained in $D_r(I, K^*, L^*)$ for $t \in [0, 2.5)$ h. After the controller parameters are switched, $r(t)$ is maintained in $D_r(\Lambda, K^*, L^*)$ for $t \in [2.5, 3.5)$ h, resulting in zero alarms generated by the detection scheme. At $t = t_s + T_c = 3.5\,\text{h}$, due to absence of an alarm in the detection scheme, the controller parameters are switched from (K_Λ, L_Λ) back to (K^*, L^*). Thereafter, $r(t) \in D_r(I, K^*, L^*)$ for $t \in (3.5, 5]$ h.

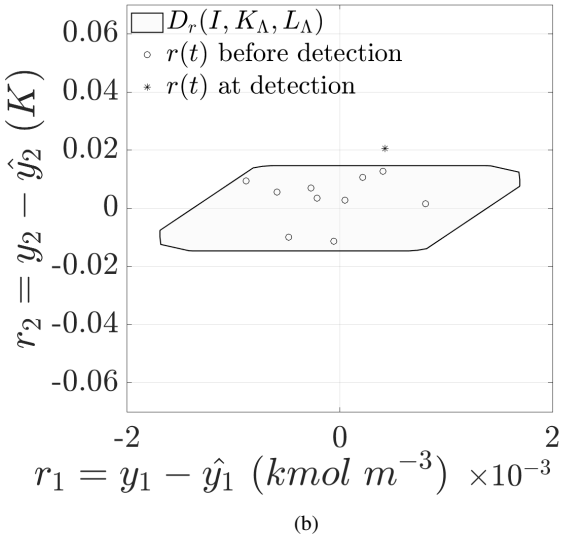
In the second case study, a simulation of the attacked closed-loop process is considered using the same realization of the random variables, where the attack begins at time $t = 0\ h$. Because the residuals satisfy $r(t) \in D_r(I, K^*, L^*)$ for all time $t \in [0, 2.5)$ h (Fig. 1a), the output of the detection scheme is $z(t) = 0$, and the attack is not detected. At time $t_s = 2.5\ h$ when the controller parameters are switched from (K^*, L^*) to (K_Λ, L_Λ), the attack is detected in 10 time steps, i.e., at time $t_d = 2.6h$ with the residual of the process $r(t_d) = [0.0004\ \ 0.0205]^T \notin D_r(I, K_\Lambda, L_\Lambda)$. Since the nominal controller parameters are stabilizing for the attack magnitudes that (K_Λ, L_Λ) are sensitive to, the controller parameters are switched back to the nominal parameters immediately after the detection of the attack. This allows for process stabilization while attack identification and mitigation schemes may be implemented. Fig. 1a illustrates that after the switch from (K_Λ, L_Λ) back to (K^*, L^*), there are alarms generated for 2 more time steps by the detection scheme due to residuals satisfying $r(t) \notin D_r(I, K^*, L^*)$ for $t \in (2.6, 2.63)\ h$.

V. Conclusions

This work presented an active attack detection methodology utilizing controller parameter switching to proactively probe a closed-loop process for sensor-controller communication link attacks. The methodology features switching between the nominal controller parameters, selected on the basis of standard control design criteria, and attack-sensitive controller parameters, selected by exploiting the interdependence of the attack detectability and closed-loop stability. The application of the proposed methodology to enhance attack detection capabilities was demonstrated using a chemical process example.



(a)



(b)

Fig. 1: (a) The residual values for the attacked closed-loop process with (K^*, L^*) showing $r(t) \in D_r(I, K^*, L^*)$ for all $t \in [0, t_s)$. (b) The residual values for the attacked closed-loop process showing $r(t_d) \notin D_r(I, K_\Lambda, L_\Lambda)$ at the detection time $t_d = 2.6\ h$.

REFERENCES

[1] K. E. Hemsley and R. E. Fisher, "History of industrial control system cyber incidents," Idaho National Lab., Idaho Falls, ID (United States), Tech. Rep. INL/CON-18-44111, 2018.

[2] T. Miller, A. Staves, S. Maesschalck, M. Sturdee, and B. Green, "Looking back to look forward: Lessons learnt from cyber-attacks on industrial control systems," *International Journal of Critical Infrastructure Protection*, vol. 35, p. 100464, 2021.

[3] A. Di Pinto, Y. Dragoni, and A. Carcano, "TRITON: The first ICS cyber attack on safety instrument systems," pp. 1–26, 2018. [Online]. Available: https://i.blackhat.com/us-18/Wed-August-8/us-18-Carcano-TRITON-How-It-Disrupted-Safety-Systems-And-Changed-The-Threat-Landscape-Of-Industrial-Control-Systems-Forever-wp.pdf

[4] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams, and A. Hahn, "Guide to industrial control systems (ICS) security," U.S. Department of Commerce, Tech. Rep. NIST Special Publication 800-82 Revision 2, 2015.

[5] M. S. Chong, H. Sandberg, and A. M. Teixeira, "A tutorial introduction to security and privacy for cyber-physical systems," in *Proceedings of the 18th European Control Conference*, Naples, Italy, 25-28 June 2019, pp. 968–978.

[6] M. Krotofil and 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, 18-21 October 2013, pp. 166–182.

[7] K. Nieman, H. C. Oyama, M. Wegener, and H. Durand, "Predict the impact of cyberattacks on control systems," *Chemical Engineering Progress*, vol. 116, no. 9, pp. 52–57, 2020.

[8] Y. Hu, H. Li, H. Yang, Y. Sun, L. Sun, and Z. Wang, "Detecting stealthy attacks against industrial control systems based on residual skewness analysis," *European Association for Signal Processing (EURASIP) Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–14, 2019.

[9] C. Murguia and J. Ruths, "CUSUM and chi-squared attack detection of compromised sensors," in *Proceedings of the IEEE Conference on Control Applications*, Buenos Aires, Argentina, 19-22 September 2016, pp. 474–480.

[10] S. Weerakkody, O. Ozel, P. Griffioen, and 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, 27-30 Aug 2017, pp. 1306–1312.

[11] M. Ghaderi, K. Gheitasi, and W. Lucia, "A blended active detection strategy for false data injection attacks in cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 168–176, 2020.

[12] G. Na and 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, 21-24 August 2018, pp. 1698–1703.

[13] A. Teixeira, I. Shames, H. Sandberg, and 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, 1-5 October 2012, pp. 1806–1813.

[14] S. Chen, Z. Wu, and P. D. Christofides, "A cyber-secure control-detector architecture for nonlinear processes," *AIChE Journal*, vol. 66, no. 5, p. e16907, 2020.

[15] S. Narasimhan, N. H. El-Farra, and M. J. Ellis, "Detectability-based controller design screening for processes under multiplicative cyberattacks," *AIChE Journal*, vol. 68, p. e17430, 2022.

[16] V. M. Kuntsevich and B. N. Pshenichnyi, "Minimal invariant sets of dynamic systems with bounded disturbances," *Cybernetics and Systems Analysis*, vol. 32, no. 1, pp. 58–64, 1996.

[17] I. Kolmanovsky and E. Gilbert, "Theory and computation of disturbance invariant sets for discrete-time linear systems," *Mathematical Problems in Engineering*, vol. 4, pp. 317–367, 1998.

[18] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *Journal of Process Control*, vol. 7, no. 6, pp. 403–424, 1997.

[19] S. Simani, C. Fantuzzi, and R. Patton, *Model-based Fault Diagnosis in Dynamic Systems using Identification Techniques*. Springer: London, 2003.

[20] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. Springer-Verlag Berlin Heidelberg, 2003.

[21] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer-Verlag Berlin Heidelberg, 2006.

[22] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Computers & Chemical Engineering*, vol. 27, pp. 293–311, 2003.

[23] P. D. Christofides and N. H. El-Farra, *Control of Nonlinear and Hybrid Process Systems: Designs for Uncertainty, Constraints and Time-Delays*. Springer Science & Business Media, 2005, vol. 324.

[24] J. Romagnoli and A. Palazoglu, *Introduction to Process Control*. CRC press, 2020.

[25] M. Kvasnica, P. Grieder, and M. Baotić, "Multi-Parametric Toolbox (MPT)," 2004. [Online]. Available: http://control.ee.ethz.ch/ mpt/

[26] S. V. Raković, E. C. Kerrigan, K. I. Kouramas, and D. Q. Mayne, "Invariant approximations of the minimal robust positively invariant set," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 406–410, 2005.