Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50242-6

Cyberattack Detectability-Based Controller Screening: Application to a Nonlinear Process

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

[a]*Department of Chemical Engineering, University of California Davis, Davis, CA-95616, USA*
mjellis@ucdavis.edu

Abstract

In this work, multiplicative cyberattacks targeting the sensor-controller communication link of a process control system are considered. The interdependence of detectability of an attack with respect to a general class of residual-based detection schemes and the control parameters is characterized. Exploiting this dependence, a controller screening methodology that may be used to incorporate cyberattack detectability into the standard controller design criteria is presented. Using a chemical process example, the application of the controller design screening to a nonlinear process is demonstrated.

Keywords: Multiplicative cyberattack, zero-alarm attack, controller design, cyberattack detectability

1. Introduction

Increased reliance on networked communication has rendered process control systems (PCSs) vulnerable to increasingly complex and frequent cyberattacks (Miller et al., 2018) in the past couple of decades. This has motivated an increasing body of research dedicated to the development of controller-based approaches to attack resilient controller design, cyberattack detection, identification, and mitigation schemes (Giraldo et al., 2018, and Tan et al., 2020 and references therein, Oyama et al., 2020, and Chen et al., 2021). Broadly, cyberattacks may target the PCS by compromising the data integrity of process data in the communication links or by altering the PCS logic. In this work, multiplicative zero-alarm cyberattacks compromising the data integrity of the sensor-controller communication link are considered. These attacks are modeled by a factor multiplied to the measured variable and are particularly threatening as they may be designed to evade detection by keeping the alarms in the detection scheme monitoring the process at zero, with minimal process knowledge.

The detectability of the measured states of a process may be viewed as a systems-theoretic property. In practice, the detectability of a cyberattack on the measured states of a process is dependent on the control parameters (e.g., controller gain and observer gain), and the detection scheme monitoring the process. This interdependence of the detectability of an attack and the controller design has not received much attention in the literature. In a previous work (Narasimhan et al., 2021), an approach to characterizing the interdependence of the attack detectability in terms of terminal set of residuals (a small set containing the origin within which the residual is ultimately bounded) for the attacked and the attack-free process was presented. Based on the characterization, a controller screening methodology that may be used identify and discard control parameters that mask an attack was presented for processes modeled by discrete-time linear time invariant

(LTI) dynamics. This screening may be used to incorporate cyberattack detectability into existing controller design criteria (e.g., closed loop-stability and robustness to uncertainty). The present work explores the application of the controller screening methodology to a nonlinear chemical process. To make the paper self-contained, an abridged version of the screening methodology is presented herein. Interested readers are directed to Narasimhan et al. (2021) for more information.

2. Controller Screening Methodology

Processes modeled by discrete-time linear time invariant (LTI) dynamics, and subject to bounded measurement noise and process disturbances are considered:

$$\begin{aligned} x(t+1) &= Ax(t) + B\,u(t) + Gw(t) \\ y(t) &= \Lambda(Cx(t) + v(t)) \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state of the process, $u(t) \in \mathbb{R}^{n_u}$ is the control input, $y(t) \in \mathbb{R}^m$ is the output from the process, $w(t) \in W \subset \mathbb{R}^{n_w}$ and $v(t) \in V \subset \mathbb{R}^m$ are the bounded process disturbances and measurement noise. The sets W and V are compact and contain the origin in their interior. Multiplicative sensor-controller link attacks are modeled by the diagonal matrix $\Lambda = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_m)$ where $\alpha_i \neq 1$ represents attack on the i^{th} sensor-controller communication link. To generate an estimate of the state (denoted by $\hat{x}(t) \in \mathbb{R}^n$), a Luenberger observer with gain L is used. A linear controller with gain K is used to steer the state to the origin. The eigenvalues of $A - LC$ and $A - BK$ are assumed to lie within the unit circle. For analysis, the augmented state vector is defined as $\xi(t) = [x^T(t) \quad e^T(t)]^T$, where $e(t) = x(t) - \hat{x}(t)$ is the estimation error (the dynamic model is defined in Eq. (5) of Narasimhan et al., 2021). The dynamics of the augmented state is $\xi(t+1) = A_\xi(\Lambda, K, L)\xi(t) + B_\xi(\Lambda, L)d(t)$, where $d(t) = [w^T(t) \quad v^T(t)]^T \in F := \{\binom{w}{v} \mid w \in W, v \in V\}$ ($A_\xi(\Lambda, K, L)$ and $B_\xi(\Lambda, K, L)$ are given in Eq. (6) of Narasimhan et al., 2021).

Due to the presence of bounded noise in the measurement, and process disturbances, when the process is at steady state, its augmented state converges to a small set containing the origin, which is the minimum invariant set. When the closed-loop process is stable with $\max_i |\lambda_i(A_\xi(\Lambda, \mathrm{K}, \mathrm{L})| < 1$ ($\lambda_i\left(A_\xi(\Lambda, K, L)\right)$ is the i^{th} eigenvalue of $A_\xi(\Lambda, K, L)$), the minimum invariant set of the process depends upon the attack magnitude and the control parameters. It is given by the infinite Minkowski sum $D_\xi(\Lambda, K, L) = B_\xi F \oplus A_\xi B_\xi F \oplus A_\xi^2 B_\xi F \oplus \ldots$ (Kuntsevich et al., 1996). The residual vector is defined as $r(t) = y(t) - \hat{y}(t)$. Writing the residual in terms of the augmented state and the disturbance gives $r(t) = [(\Lambda - I)C \quad C]\xi(t) + [0 \quad \Lambda]d(t) = A_r(\Lambda)\xi(t) + B_r(\Lambda)d(t)$. Thus, the terminal set of residuals for the process may be computed as $D_r(\Lambda, K, L) = A_r(\Lambda)D_\xi(\Lambda, K, L) \oplus B_r(\Lambda)F$.

A general class of residual-based detection schemes using the 2-norm of the residual vector as the detection metric is considered. Elaborate discussion on these detection schemes may be found in Section 2.3-2.4 of Narasimhan et al., 2021. With respect to the class of detection schemes considered, a closed-loop stability preserving attack of magnitude $\Lambda \neq I$ is defined as undetectable if the radius of the 2-norm ball enclosing the terminal residual set of the process under an attack is less than or equal to the radius of the 2-norm ball enclosing the terminal residual set of the attack-free process, i.e, $R(\Lambda, K, L) \leq R(I, K, L)$, where $R(\Lambda, K, L) \coloneqq \max_{r' \in D_r(\Lambda, K, L)} ||r'||$. Similarly, any closed-loop

stability preserving attack of magnitude $\Lambda \neq I$ on the process is defined as a potentially detectable attack if the radius of the 2-norm ball enclosing the terminal set of the attack-free process is less than the radius of the 2-norm ball enclosing the terminal set of the process under an attack, i.e., $R(I,K,L) < R(\Lambda,K,L)$. Finally, an attack of magnitude $\Lambda \neq I$ is said to be detectable if it renders the closed-loop system in Eq.(1) unstable. Interested readers are directed to Section 3.1 of Narasimhan et al., 2021, for more discussions.

To derive numerically verifiable conditions for characterizing attack detectability, invariant outer polytopic approximation of the minimum invariant set of the augmented states of the stable process is computed as $D_{\xi}^{est}(\Lambda,K,L)$ (see Raković et al., 2005 for method), such that $D_{\xi}(\Lambda,K,L) \subseteq D_{\xi}^{est}(\Lambda,K,L) \subseteq D_{\xi}(\Lambda,K,L) \oplus B_{\infty}^{2n}(\epsilon)$ $(B_{\infty}^{2n}(\epsilon) \coloneqq \{\xi' \in \mathbb{R}^{2n} \mid \| \xi' \|_{\infty} \leq \epsilon\})$, where ϵ is the error bound. Then, the inner polytopic approximations of the terminal residual sets for the attack-free and the attacked process are computed as $D_{r_e}^{est}(I,K,L)$ and $D_{r_{a,e}}^{est}(\Lambda,K,L)$, respectively. These inner approximations satisfy $D_{r_j}^{est}(\Lambda,K,L) = D_{r}^{est}(\Lambda,K,L) \ominus A_r(\Lambda)B_{\infty}^{2n}(\epsilon) \subseteq D_r(\Lambda,K,L)$ ($j = e$ or $j = a,e$), where, $D_{r}^{est}(\Lambda,K,L)$ is the outer polytopic approximation of $D_r(\Lambda,K,L)$.

With the disturbance set F, the attack magnitude of interest Λ, the controller gain K, and the observer gain L as the input, the controller screening algorithm is as follows. First, it is checked if $\max_i |\lambda_i(A_{\xi}(\Lambda,K,L)| \geq 1$. If this is true, the choice of K and L under a multiplicative attack with attack matrix Λ will render the closed-loop process unstable and the attack is detectable. The screening algorithm is terminated in this case. However, if $\max_i |\lambda_i(A_{\xi}(\Lambda,K,L)| < 1$, then the outer polytopic approximations of the residual sets are computed using the method described in Section 3.2 of Narasimhan et al., 2021. From the outer approximations, the inner approximations of the residual sets are computed. Then, the radii of the 2-norm balls enclosing the sets ($R^{est}(I,K,L)$, $R^{est}(\Lambda,K,L)$, $R_e^{est}(\Lambda,K,L)$, and $R_{a,e}^{est}(\Lambda,K,L)$) are computed. If $R^{est}(\Lambda,K,L) < R_e^{est}(I,K,L)$, the attack is classified as an undetectable attack, else, if $R_{a,e}^{est}(\Lambda,K,L) > R^{est}(I,K,L)$, then the attack is classified as a potentially detectable. However, if it is found that $R^{est}(\Lambda,K,L) > R_e^{est}(I,K,L)$ or $R_{a,e}^{est}(\Lambda,K,L) > R^{est}(I,K,L)$, the test is inconclusive. This may occur when the numerical approximations of the terminal sets satisfy $|R_{a,e}^{est}(\Lambda,K,L) - R^{est}(I,K,L)| \leq \beta$ or $|R_e^{est}(\Lambda,K,L) - R^{est}(I,K,L)| \leq \beta$ where $\beta > 0$ is a small number, implying that the radius estimates are close to each other.

3. Application to a Nonlinear Chemical Process

A chemical process example consisting of a continuously stirred tank reactor (CSTR) with a second-order reaction occurring is considered. The process is subject to bounded process disturbances and measurement noise. In the real-time control of chemical processes, the control actions are usually applied to the nonlinear continuous-time process at discrete time instances. To simulate this, the nonlinear ordinary differential equation model describing the CSTR process is integrated with the linear controller generating a new control action at every 10^{-2} h. To apply the screening methodology, a discrete-time LTI process model is needed. To generate the model, a continuous-time LTI model is first obtained by linearizing the nonlinear model about its open-loop stable steady state. The discrete-time LTI process model is then obtained by applying a zero-order hold discretization. The process models, and process parameter values are given in Narasimhan et al., 2021.

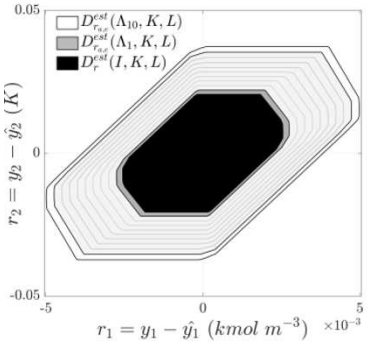

Figure 1. Outer approximation of the terminal residual set for the attack-free process ($D_r^{est}(I,K,L)$), and the inner approximations of the terminal sets for the attacked process ($D_{r_{a,e}}^{est}(\Lambda_i,K,L)$).

In this study, the control parameters are to be chosen such that they do not mask multiplicative attacks of magnitude in this range $[1.1, 2]$ on the concentration sensor-controller link. Since there are infinitely many attack magnitudes in the range, the screening algorithm is carried out with ten values from the range, i.e., with values $\Lambda_i = \text{diag}(\alpha_i, 1)$ where $\alpha_i = 1.1 + 0.1(i-1)$, $i = 1,2,\dots,10$. Using pole placement, the controller and observer gains are selected with the controller poles at $[0.5 - 0.1]$ and observer poles at $[-0.3\ 0.4]$. The screening algorithm is applied to the chosen controller design over all Λ_i. The attacked closed-loop process is stable with $\max_j \left|\lambda_j\left(A_{\xi_a}(\Lambda_i,K,L)\right)\right| < 1$ for all $i = 1,2,\dots 10$. The polytopic approximations of the minimum invariant sets of the attacked process over all attack magnitudes, and for the attack-free process are computed with an error bound of $\epsilon = 5 \times 10^{-5}$. Then, the polytopic approximations of terminal residual sets are computed (Figure 1). The size of the terminal residual set for the attacked process increases with α_i. For all cases, $R^{est}(I,K,L) < R^{est}_{a,e}(\Lambda_i,K,L)$, $i = 1,2,\dots 10$, with $R^{est}(I,K,L) = 0.0207$, $\min_i R^{est}_{a,e}(\Lambda_i,K,L) = 0.0222$ for $\Lambda_1 = \text{diag}(\alpha_1, 1) = \text{diag}(1.1,1)$ and $\max_i R^{est}_{a,e}(\Lambda_i,K,L) = 0.0375$ for $\Lambda_{10} = \text{diag}(2,1)$. Thus, all the attacks in the range of interest are potentially detectable. To verify the potential detectability of each attack magnitude, two sets of simulations of the process are performed, using the (1) discrete-time LTI process model and (2) nonlinear continuous-time process model. One thousand closed-loop simulation pairs of the attacked process are performed with various realizations of the measurement noise and the process disturbance. For each simulation pair, the same realization was applied. For both sets of simulations, the state is estimated using a Luenberger observer designed based on the discrete-time LTI process model. To monitor the process, a CUSUM detection scheme described by Eq. (12) in Narasimhan et al., 2021 is used. The detection scheme is tuned with the choice of $b = R^{est}(I,K,L) = 0.0207$ and an alarm threshold of $\tau = 0.01$ to guarantee a zero-false alarm rate in the absence of an attack.

For attacks with $\alpha_i \in [1.6, 2]$, the attack is detected over most simulations with the LTI and the nonlinear models. Results from simulating an attack of magnitude $\Lambda_{10} = \text{diag}(2,1)$ on the nonlinear process model are presented in Figure 2. The attack is detected in 969 out of 1000 simulations with the nonlinear continuous-time process model. Of these 969 simulations, the attack is not detected in the corresponding linear model simulation in 4 simulations. Similarly, during the simulations with the discrete-time LTI process model, the attack is detected in 970 out of 1000 simulations. For 5 of the 970 simulations, the attack is not detected for the

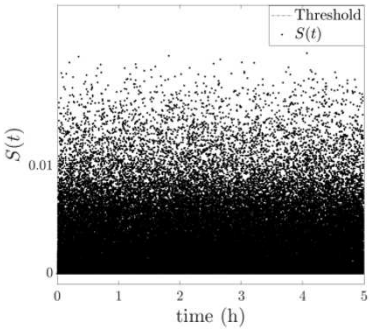

Figure 2. The CUSUM statistic $S(t)$ over 1000 simulations of the closed-loop process under an attack of magnitude $\Lambda_{10} = \text{diag}(2,1)$.

corresponding nonlinear model simulation. In most of the simulation pairs where a discrepancy between attack detection is observed, the difference in the maximum CUSUM statistic value between the two cases was small, meaning that in one case, the value exceeded the threshold by a small amount and in the other, the statistic value was close to the other, but did not exceed the threshold. For example, an attack is detected in the linear case with a CUSUM statistic value of 0.0101, and for the corresponding nonlinear case, the statistic value at the same time is 0.0095, and the attack is not detected. Finally, for the same realization of random variables applied to the nonlinear case and linear case, except for 174 simulations, the time at which the attack is first detected is the same for both process models. Over most of these 174 simulations, the attack detection time for the two process models differed by a single time step.

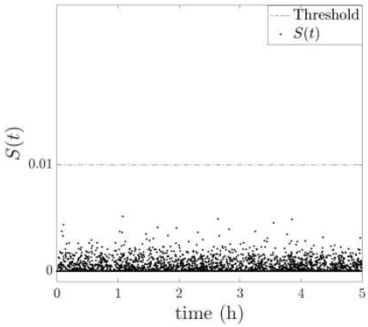

Figure 3. The CUSUM statistic $S(t)$ over 1000 simulations of the closed-loop process under an attack of magnitude $\Lambda_4 = \text{diag}(1.4,1)$.

For attacks with magnitude in range $\alpha_i \in [1.2, 1.5]$, while the CUSUM statistic never breaches the threshold, it is found to be a non-zero value over most simulations for both the nonlinear case and the linear case. The CUSUM statistic resulting from simulating an attack of magnitude $\Lambda_4 = \text{diag}(1.4, 1)$ (with $R_{a,e}^{est}(\Lambda_4, K, L) = 0.027$) on the nonlinear model are given in Figure 3. It is also found that the maximum realization of the CUSUM statistic over each simulation of the nonlinear model is of the same order of magnitude as the maximum realization of the statistic over the corresponding simulation of the linear model. In one case, the maximum CUSUM statistic in simulations with the nonlinear case is 0, however, the maximum statistic for corresponding simulations with the linear case is 3.1×10^{-5}. The mean of the 2-norm of the residual vector over 1000 simulations is 0.0081 with a variance of 2.4563×10^{-5} for the nonlinear case. The mean of the 2-norm of the residual vector over 1000 simulations with the linear case is 0.0081 (same as that for the nonlinear case) and the variance is 2.4553×10^{-5}. While not shown here, statistical distribution of the 2-norm of the residual indicates that tuning the CUSUM detection scheme with a lower alarm threshold may enable the detection of this attack. Note that with the tuning approach adopted in this work, the CUSUM statistic remains at zero for the attack-free process, and any non-zero CUSUM statistic value would indicate an attack. Furthermore, with a decrease in α_i, the total number of simulations with non-zero realizations of the CUSUM statistic decrease, indicating that the attack becomes more difficult to detect.

During closed-loop simulations with an attack of magnitude $\Lambda_1 = \text{diag}(1.1,1)$, the CUSUM statistic remains at zero over all the 1000 simulations for both the linear case and the nonlinear case. Thus, the attack goes undetected, despite the control parameters satisfying $R^{est}(I, K, L) < R_{a,e}^{est}(\Lambda_1, K, L)$. This does not contradict the screening algorithm, as the attack is only potentially detectable. The fact that the attack went undetected may be because $R^{est}(\Lambda_1, K, L) - R^{est}(I, K, L) = 0.0015$ is small. While further analysis is required, for this attack magnitude, a different set of control parameters with a larger separation between the sets $D_r^{est}(I, K, L)$ and $D_r^{est}(\Lambda_1, K, L)$ may enable attack detection. Moreover, for this case, the performance degradation resulting from this attack is small.

The applicability of the controller screening algorithm to the nonlinear case is verified for other control designs and attacks. For example, when the controller screening algorithm predicts that (1) an attack $\Lambda = \text{diag}(0.5,1)$ with control and observer gains selected by placing the poles at $[0.3\ 0.1]$ and $[-0.4\ 0.3]$, respectively, and (2) an attack $\Lambda = \text{diag}(1,0.9)$ for control and observer gains selected by placing poles at $[0.5 - 0.1]$ and $[0.4\ 0.4]$, respectively, are undetectable. During the closed-loop simulations with the nonlinear case and the linear case, the CUSUM statistic remains at zero, and the attacks are not detected. When the controller screening predicts that an attack $\Lambda = \text{diag}(1,1.2)$ is potentially detectable for control and observer gains with poles placed at $[0.5 - 0.1]$ and $[0.2\ 0.2]$, respectively, the CUSUM statistic has non-zero realizations over most simulations. However, it never breaches the threshold (similar to Figure 3). Thus, the controller screening algorithm based on the discrete-time LTI model of the process is applicable to the continuous-time nonlinear model of the process. This may be because the linear process model adequately represents the dynamics of the nonlinear process near the operating steady-state.

4. Conclusions

In this work, the application of a detectability-based controller design screening methodology for a multiplicative sensor-controller link cyberattacks to a nonlinear chemical process example was demonstrated. Future work will focus on controller-based approaches that enable the detection of a multiplicative sensor-controller link attack.

References

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

J. Giraldo, D. Urbina, Á.A. Càrdenas, J. Valente, M. Faisal, J. Ruths, N.O. Tippenhauer, H. Sandberg, R. Candell, 2018, A Survey of Physics-Based Attack Detection in Cyber-Physical Systems, ACM Computing Surveys, vol. 51, issue 4, article 76, pp. 1–36.

S. Tan, J.M. Guerrero, P. Xie, R. Han, J.C. Vasquez, 2020, Brief survey on attack detection methods for cyber-physical systems. IEEE Systems Journal, vol. 14(4), pp. 5329–5239.

H. Oyama, H. Durand, 2020, Integrated cyberattack detection and resilient control strategies using Lyapunov-based economic model predictive control, AIChE Journal, vol. 66, issue 12, pp. e17084.

S. Chen, Z. Wu, P.D. Christofides, Cyber-security of centralized, decentralized, and distributed control-detector architectures for nonlinear processes, 2021, Chemical Engineering Research and Design, vol. 165, pp. 25–39.

V.M. Kuntsevich, B.N. Pshenichnyi, 1996, Minimal invariant sets of dynamic systems with bounded disturbances, Cybernetics and Systems Analysis, vol. 32(1), pp. 58–64.

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

S. Narasimhan, N.H. El-Farra, M.J. Ellis, 2021, Detectability-based controller design screening for processes under multiplicative cyberattacks, AIChE Journal, vol. 68, issue 1, pp. e17430.