Terminal Set-based Cyberattack Detection in Model Predictive Control Systems with Zero False Alarms

By

Rahul Panicker

Thesis

Submitted in partial satisfaction of the requirements for the degree of

Master of Science

in

Chemical Engineering

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Matthew J. Ellis, Chair

Nael H. El-Farra, Co-chair

Ahmet N. Palazoglu

Committee in Charge

2025

Acknowledgements

I would like to begin by expressing my deepest gratitude to the Almighty for His divine presence and the strength that has made this journey possible.

I am grateful to my advisors, Professor Matthew J. Ellis and Professor Nael H. El-Farra, for their guidance and support throughout my graduate studies. I would also like to extend my sincere thanks to Professor Ahmet N. Palazoglu for generously dedicating his time to serve on my thesis committee. Their collective contributions have enriched the quality of my research.

I would also like to thank my lab members—Shilpa, Loren, Aatam, Pranav, Hossein, and Aniket—for their support throughout my time in the lab. Each of you has helped me in different ways, both personally and professionally, and for that, I am truly grateful.

Lastly, I would like to thank my friends and family for their unconditional support and steadfast belief in me, which has been the cornerstone of my academic journey. Words cannot adequately express the sacrifices my parents, Rakesh Panicker and Sunita Panicker, have made in raising me. Their love and dedication have helped me reach where I am today. I am equally grateful to my brother, Suraj, whose unwavering belief in me has given me the strength to push through challenging times. I am forever indebted to all three of you for shaping me into the person I am today.

Terminal Set-based Cyberattack Detection in Model Predictive Control Systems with Zero False Alarms

Abstract

The increased reliance of industrial control systems on networked components has made them more vulnerable to cyberattacks, necessitating cyberattack detection schemes specifically designed for detecting cyberattacks affecting industrial control systems. This thesis presents a set-membership-based detection scheme for systems under model predictive control (MPC). Specifically, we consider steady-state operation because many systems operate over long periods near a desired steady state. Provided the disturbances and measurement noise acting on the system are sufficiently small, we show that the closed-loop system under MPC is equivalent to the closed-loop system under a linear quadratic regulator, formulated with the same stage cost and weighting matrices, in a region containing the desired operating point. This equivalence is leveraged to show that the minimum robust positively invariant (mRPI) sets under both controllers are equivalent, enabling the calculation of the mRPI set for the closed-loop system under MPC. Using the mRPI set of the attack-free system, we present an attack detection scheme for systems under MPC and derive conditions under which the attack detection scheme applied to the attack-free closed-loop system does not raise an alarm. The detection scheme is applied to a simplified (linear) building space-cooling system to demonstrate that it does not raise false alarms during attack-free operation and that it successfully detects attacks when the system is subjected to a multiplicative false-data injection attack altering the data communicated over the sensor-controller link. Furthermore, the detection scheme's applicability to nonlinear systems is assessed. Specifically, the detection scheme is applied to a

nonlinear chemical process to demonstrate that the detection scheme does not raise false alarms during attack-free operation and successfully detects an attack when the process is subjected to a false-data injection attack.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past decade, industrial control systems have increasingly become targets of cyberattacks, compromising the integrity of data being communicated in the closed-loop system and resulting in performance degradation, economic loss, and safety risks [1, 2, 3]. The severity of these attacks became evident with the Stuxnet virus, one of the world's first publicly known malware designed for a cyberattack on industrial control systems [4, 5]. Recent cyberattacks on control systems, including the Ukraine power grid attack, the German steel mill attack, and the Florida water treatment plant attack, have further emphasized the need for resilient cybersecurity measures [6]. Consequently, an increasing body of research has focused on developing resilient cyber-secure control system designs to safeguard against the potentially devastating consequences of cyberattacks on control systems [7]. These attacks demonstrate that information technology (IT)-based approaches to cyberattacks are insufficient and that operation technology (OT)-based approaches are also needed.

Cyberattacks targeting control systems can take various forms, including denial-of-service (DoS) attacks, man-in-the-middle attacks, and false data injection attacks [8]. A

critical approach to mitigating these threats is the integration of cyberattack detection schemes. Numerous detection schemes have been proposed in the literature, including residual-based methods that utilize a residual—a measure of the deviation between two variables, often a measured variable and an estimated or predicted variable (e.g., [9, 10]). For example, a model predictive control (MPC) strategy was introduced with an added constraint to maintain outputs within a time-varying neighborhood of the reference trajectory in [9]. The residual, calculated as the deviation from this trajectory, was analyzed using a cumulative sum (CUSUM) scheme. Several Lyapunov-based MPC formulations and detection schemes that monitor residuals and state values to detect attacks have also been proposed [10]. In residual-based schemes, an attack is detected when the residual's norm exceeds a defined threshold. On the other hand, in state-based schemes, states are compared against their expected operational regions, with attacks identified when states deviate from these regions.

Neural network-based detection schemes represent another class of approaches that have been used for cyberattack detection [11, 12, 13]. In [11], attacks are detected using a neural network-based detector, where the detector is trained with a large data set of states to distinguish between different operational modes representing attack-free operation and attack operations. In [13], to reduce the amount of data required for training neural network-based detectors, residual-based indicator functions are embedded within the loss function of standard neural network-based detectors to distinguish between attack-free operation and attack operations.

An important design consideration for attack detection schemes is minimizing instances of false alarms, i.e., instances when an alarm is raised, but there is no attack [14]. In

our previous work, we designed attack detection schemes to produce zero false alarms for systems controlled by a linear control system using a linear observer and controller without considering state or input constraints [15, 16, 17]. The attack detection scheme considered was a set-membership-based detection scheme, wherein a monitoring variable was defined, and a terminal set of the monitoring variable under attack-free operation was computed. Specifically, in [15], theoretical conditions were established based on the terminal set to classify attacks as detectable, undetectable, or potentially detectable with respect to a general class of set-based detection schemes. This relationship was leveraged in [16] to present an active set-based attack detection scheme utilizing control system parameter switching to facilitate attack detection. In [17], theoretical conditions were presented to guarantee zero false alarms for the active detection scheme proposed in [16]. However, the method proposed for estimating the terminal set used for detection only applies to linear systems under an explicit linear controller (one that is an explicit function of the system's states or outputs) and a linear observer without accounting for state or input constraints.

On the other hand, an implicit controller determines the control action indirectly as an implicit function of the states or outputs. MPC is an implicit optimization-based control strategy because the control action is obtained by solving an optimal control problem to minimize a cost function while accounting for constraints on the state and inputs [18, 19]. One of the main advantages of MPC compared to other control strategies is its ability to handle system constraints explicitly [20]. Several papers have considered various aspects related to cyberattack detection for systems regulated by MPC. In addition to [9, 11, 10, 13], a set-membership-based method using one-step reachable sets has been

presented [21]. However, to the best of the authors' knowledge, no prior work has investigated the integration of terminal sets into a set-membership-based detection scheme for systems regulated by MPC, where the terminal sets can be precisely calculated. Such schemes are capable of achieving zero false alarms during attack-free operation under MPC and guarantee that any alarm raised indicates the presence of an attack.

Motivated by this, we present in this work a terminal set-based cyberattack detection scheme for monitoring systems controlled by MPC. To this end, discrete-time linear time-invariant systems subjected to bounded disturbances and measurement noise and operated near the desired steady-state over long periods are considered. Under attack-free operation, the closed-loop system state asymptotically converges to the minimum robust positively invariant (mRPI) set. We show that provided the disturbances and measurement noise acting on the system are sufficiently small (made precise in this thesis), the closed-loop system under MPC is equivalent to that under a linear quadratic regulator (LQR). Moreover, we show that the mRPI sets under both controllers are equivalent, facilitating the calculation of the mRPI set under MPC. Using the mRPI set of the attack-free system, we present an attack detection scheme for systems under MPC. We derive conditions under which the detection scheme does not raise false alarms when applied to the attack-free closed-loop system and raises alarms only if the closed-loop system is under attack.

The organization of the rest of the thesis is as follows. We provide details on the notations, class of systems, and class of cyberattacks in Chapter 2. Chapter 3 presents a method to compute the mRPI set for closed-loop systems under MPC and the proposed terminal set-based detection scheme. In Chapter 4, we apply the detection scheme to

a building space-cooling system controlled by MPC. We demonstrate that the detection scheme does not raise false alarms during attack-free operations. We consider multiplicative false-data injection attacks altering the data communicated over the sensor-controller link as an illustrative model to demonstrate that the detection scheme successfully detects the attack. Furthermore, we demonstrate the detection scheme's applicability to nonlinear processes using an illustrative nonlinear chemical process. We show that the proposed detection scheme does not trigger any false alarms for the attack-free nonlinear process and that it successfully detects an attack when a multiplicative false-data injection attack is applied. Finally, Chapter 5 provides the conclusions and potential future work.

Chapter 2

Preliminaries

2.1 Notation

The non-negative set of integers is denoted by $\mathbb{I}_+$. The n-dimensional Euclidean space is denoted by $\mathbb{R}^n$. The Minkowski sum of two sets $\mathbb{A} \subset \mathbb{R}^n$ and $\mathbb{B} \subset \mathbb{R}^n$ is denoted by $\mathbb{A} \oplus \mathbb{B} = \{a + b | a \in \mathbb{A}, b \in \mathbb{B}\}$. For a matrix $M \in \mathbb{R}^{m \times n}$ and set $\mathbb{A} \subset \mathbb{R}^n$, $M\mathbb{A}$ denotes the set $\{Ma | a \in \mathbb{A}\}$. For a matrix $Q \in \mathbb{R}^{n \times n}$, $||Q||$ refers to the 2-norm of the matrix Q. The matrix I denotes the identity matrix of appropriate dimensions. For a positive definite function $V : \mathbb{R}^n \to \mathbb{R}_+$, Ω_ρ denotes the sublevel set of the function, i.e., $\{x \in \mathbb{R}^n | V(x) \leq \rho\}$ for $\rho > 0$. A closed norm ball of radius r^* is denoted by $\mathcal{B}_{r^*}$, where $\mathcal{B}_{r^*} := \{\zeta \in \mathbb{R}^n | \|\zeta\| \leq r^*\}$.

2.2 Class of Systems

We consider linear time-invariant systems of the form:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + B_w w(t) \\ y(t) &= x(t) + v(t) \end{aligned} \tag{2.1}$$

where $x(t) \in \mathbb{R}^{n_x}$ is the state, $x(t+1) \in \mathbb{R}^{n_x}$ is the successor state, $u(t) \in \mathbb{R}^{n_u}$ is the manipulated input, $w(t) \in \mathbb{W} \subset \mathbb{R}^{n_w}$ is the disturbance, $y(t) \in \mathbb{R}^{n_x}$ is the measured output, and $v(t) \in \mathbb{V} \subset \mathbb{R}^{n_v}$ is the measurement noise. The initial time is taken to be $t = 0$. The sets $\mathbb{W}$ and $\mathbb{V}$ are compact polytopes containing the origin. The system matrices A, B, and B_w are of appropriate dimensions, and the pair (A, B) is stabilizable. Constraints are imposed on the state and input where the admissible state set, denoted by $\mathbb{X}$, is closed and the admissible input set, denoted by $\mathbb{U}$, is compact. Both sets are assumed to have a non-empty interior. For compactness of notation, we define $\mathbb{Z} := \mathbb{X} \times \mathbb{U}$. We assume that the desired operating point is the origin of the unforced system, i.e., $x_s = 0$ is the desired operating point, which corresponds to the steady-state of the system with manipulated input $u_s = 0$, disturbance $w_s = 0$, and measurement noise $v_s = 0$, and $x_s \in \text{int } \mathbb{X}$ and $u_s \in \text{int } \mathbb{U}$.

2.3 Model Predictive Control

MPC is an implicit control strategy in which the control action is obtained by repeatedly solving an optimal control problem (e.g., [20]). Compared to other control strategies, one advantage of MPC is that it can explicitly account for state and input constraints in its

formulation. The finite-horizon optimal control problem of MPC is given by:

$$\min_{\tilde{\mathbf{u}}} \sum_{k=t}^{t+N-1} \left(\tilde{x}(k)^T Q \tilde{x}(k) + \tilde{u}(k)^T R \tilde{u}(k)\right) + \tilde{x}(t+N)^T P \tilde{x}(t+N) \tag{2.2a}$$

$$\text{s.t. } \tilde{x}(k+1) = A\tilde{x}(k) + B\tilde{u}(k), \tag{2.2b}$$

$$\tilde{x}(t) = y(t) \tag{2.2c}$$

$$\tilde{x}(k) \in \mathbb{X}, \forall\ k \in \{t+1, ..., t+N\} \tag{2.2d}$$

$$\tilde{u}(k) \in \mathbb{U}, \forall\ k \in \{t, ..., t+N-1\} \tag{2.2e}$$

where N is the number of time steps in the prediction horizon, $\tilde{x}(k)$ is the predicted state at time $k \in \{t, \ldots, t+N\}$, $\tilde{u}(k)$ is the predicted input at time $k \in \{t, \ldots, t+N-1\}$, and $\tilde{\mathbf{u}} := \{\tilde{u}(t), \ldots, \tilde{u}(t+N-1)\}$.

In the above formulation, Eq. 2.2a represents a conventional quadratic regulation cost function with a quadratic stage cost and a quadratic terminal cost that penalizes the deviation of the state and input from their corresponding steady-state values. The matrices Q and P are positive semidefinite weighting matrices, and R is a positive definite weighting matrix. The terminal cost can serve two purposes: to summarize the cost-to-go, i.e., the cost incurred beyond the prediction horizon, and for closed-loop stability considerations (see, for example, [20]). The dynamic model is used in Eq. 2.2b to forecast the system behavior over the prediction horizon. The model is initialized with a measurement from the system (Eq. 2.2c). In this work, we consider that the problem is initialized with a state measurement, which could be corrupted by noise. The constraints in Eqs. 2.2d-2.2e are the state and input constraints.

MPC is implemented following the receding horizon principle: at time t, the MPC receives the measured output $y(t)$ from the system sensors, creates and solves an instance

of the problem in Eq. 2.2 to determine the optimal input sequence, denoted by $\tilde{\mathbf{u}}^* := \{\tilde{u}^*(t), \ldots, \tilde{u}^*(t+N-1)\}$, and sends the first element ($\tilde{u}^*(t)$) in the sequence to the control actuators. At time $t+1$, the process is repeated after receiving an updated measured output. This process is repeated indefinitely.

We select the terminal cost weighting matrix P to be the solution to the algebraic Riccati equation (ARE), given by:

$$P = A^T P A - P - (A^T P B)(B^T P B + R)^{-1} B^T P A + Q \tag{2.3}$$

If the pair (A, B) is stabilizable and the pair (A, Q) is detectable, the ARE has a unique positive semidefinite solution (e.g., [22]). This choice of terminal cost is an exact representation of the cost-to-go for the nominal system in the sense that $\tilde{x}^T(t+N)P\tilde{x}(t+N)$ is the optimal value of the infinite horizon cost, i.e.,

$$\min_{\mathbf{u}} \sum_{k=0}^{\infty} x(k)^T Q x(k) + u(k)^T R u(k) \tag{2.4}$$

for the system without perturbations, given by:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= x(k) \end{aligned} \tag{2.5}$$

initialized at $x(0) = \tilde{x}(t+N)$ and with controller:

$$u(k) = -K^* x(k) \tag{2.6}$$

where K^* represents the optimal linear quadratic regulator (LQR) gain, which minimizes

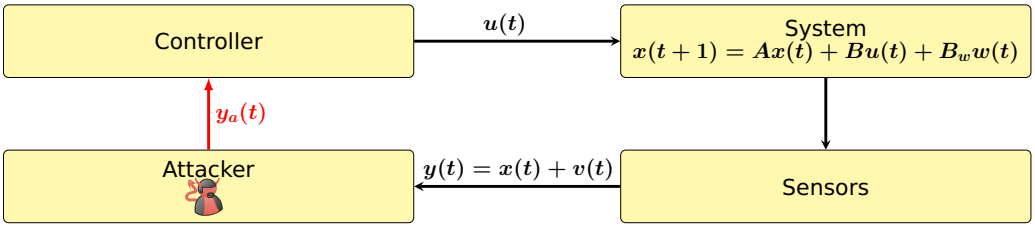


Figure 2.1: Closed-loop system subject to a false-data injection attack altering data communicated on the sensor-controller link.

the infinite horizon quadratic cost. The gain K^* is given by:

$$K^* = (B^T P B + R)^{-1} B^T P A \tag{2.7}$$

and provided that $(x(k), u(k)) \in \mathbb{Z}$ for all k. Under the stated conditions, the eigenvalues of $A - BK^*$ are within the unit circle (e.g., [22]).

2.4 Cyberattacks Targeting Control Systems

Industrial control systems can be targeted by several types of cyberattacks [2, 23, 24]. Specifically, cyberattacks, such as denial-of-service, replay, and false-data injection attacks, can alter or prevent data from being communicated over the control system. False-data injection attacks alter the data communicated over the sensor-controller or controller-actuator link (e.g., [25]). Figure 2.1 illustrates the closed-loop system and how a false-data injection attack alter the data transferred from the sensors to the controller for the case of a sensor-controller link false-data injection attack. In this case, the controller receives the output manipulated by the attacker, $y_a(t)$, instead of the actual outputs, $y(t)$. Two specific models of false-data injection attacks that have been considered in the literature are (1) additive attacks, which alter the data by adding a value to the data being communicated [26, 27], and (2) multiplicative attacks, which alters the data by multiplying the data being communicated over the link by a factor [28, 17, 29].

Chapter 3

Cyberattack Detection Scheme for Systems Under MPC

This chapter presents an attack detection scheme to detect attacks in systems controlled by MPC. The attack detection scheme relies on the computation of the minimum robust positive invariant sets for the closed-loop system under MPC. We begin by outlining a method to compute these sets for systems operating under MPC in the presence of sufficiently small perturbations. To establish this, we first analyze systems under perturbations controlled by an explicit stabilizing controller. We prove that the closed-loop system under the explicit stabilizing controller possesses a robust positive invariant set within which the state and input constraints are satisfied, provided the perturbations are sufficiently small. We then compare the closed-loop systems under LQR and MPC, showing that under sufficiently small perturbations, the minimum robust positive invariant set for the system under MPC is the same as that for LQR. Building on these findings, we present an attack detection scheme for systems regulated by MPC and prove that this scheme operates with zero false alarms.

3.1 Robustness Analysis of Constrained Attack-free Systems using an Explicit Stabilizing Controller

We summarize results on the closed-loop system consisting of the system in Eq. 2.1 under an explicit stabilizing controller. We consider a general stabilizing controller in this section, given that the results presented broadly apply to any controller of the form $u = -Kx$ such that the eigenvalues of $A - BK$ lie within the unit circle. While the results follow from standard stability and robustness analysis, we present them here to make the thesis self-contained. The analysis reveals that under sufficiently small perturbations, characterized below, a robust positive invariant set exists within which state and input constraints are satisfied.

Since we analyze several closed-loop systems under different controllers, we clarify the terminology used: the term nominal system refers to the system in Eq. 2.1 without perturbations ($w(t) = 0$ and $v(t) = 0$ for all $t \geq 0$); the unconstrained system refers to the system in Eq. 2.1 without input and state constraints; the constrained system refers to the system in Eq. 2.1 with input and state constraints; the attack-free system refers to the case when the controller receives the actual output $y(t)$; and the attacked system refers to the case where the controller receives an output value that has been modified by the attack, $y_a(t)$. The results presented in this section apply to the attack-free system. The controller uses feedback measurements of the state for the nominal system because $v(t) = 0$ for all $t \geq 0$. In contrast, the controller uses the measured output to compute the control action for the closed-loop system, consisting of the system in Eq. 2.1 (with perturbations) under any of the controllers considered. Additionally, we present several standard set definitions:

Definition 1. *A set $\mathbb{M}$ is a positive invariant set for a nominal closed-loop system (e.g., the nominal system with controller $u(t) = -Kx(t)$) if*

$$x(t) \in \mathbb{M} \implies x(t+1) \in \mathbb{M} \tag{3.1}$$

for all $x(t) \in \mathbb{M}$.

Definition 2. *A set $\mathbb{M}$ is a robust positive invariant (RPI) set for a closed-loop system (e.g., the system in Eq. 2.1 with controller $u(t) = -Ky(t) = -K(x(t) + v(t))$) if*

$$x(t) \in \mathbb{M} \implies x(t+1) \in \mathbb{M} \tag{3.2}$$

for all $x(t) \in \mathbb{M}$, $w(t) \in \mathbb{W}$, and $v(t) \in \mathbb{V}$.

Definition 3. *The minimum RPI (mRPI) set for a closed-loop system is the RPI set contained in every closed RPI set.*

We consider the nominal system, i.e., the system in Eq. 2.1 without the disturbance and measurement noise. Given that (A, B) is stabilizable, a feedback gain K exists such that eigenvalues of $(A - BK)$ lie within the unit circle. The nominal closed-loop system under this controller (neglecting constraints) is given by:

$$x(t+1) = (A - BK)x(t) \tag{3.3}$$

where $u(t) = -Kx(t)$. While constraints are neglected in writing the closed-loop system in Eq. 3.3, i.e., it is an unconstrained system, the system possesses a positive invariant set where any state within this set satisfies the state constraints. Moreover, the input computed from the controller satisfies the input constraints for any point contained within this positive invariant set. This is precisely stated in the following proposition.

Proposition 1. *For the closed-loop system in Eq. 3.3 where the eigenvalues of $A - BK$ lie within the unit circle, there exists a positive invariant set, denoted by Ω_ρ, such that $(x, -Kx) \in \mathbb{Z}$ for all $x \in \Omega_\rho$.*

The proof of Proposition 1 is provided in Appendix A.

We now consider the perturbed closed-loop system, i.e., the system under the explicit stabilizing controller and subject to bounded process disturbance and measurement noise. In this case, the controller uses the output to compute the control action, i.e., $u(t) = -Ky(t)$. The resulting closed-loop system is given by:

$$x(t+1) = A_{cl}x(t) + B_{cl}f(t) \tag{3.4}$$

where the closed-loop system matrices A_{cl} and B_{cl} are defined as:

$$A_{cl} := A - BK \tag{3.5}$$

$$B_{cl} := \begin{bmatrix} B_w & -BK \end{bmatrix} \tag{3.6}$$

and $f(t) := \begin{bmatrix} w^T(t) & v^T(t) \end{bmatrix}^T \in \mathbb{F} := \mathbb{W} \times \mathbb{V}$.

Using Ω_ρ, we intend to find an RPI set for the closed-loop system in Eq. 3.4 where $(x, -K(x+v)) \in \mathbb{Z}$ for any x in this RPI set and all $v \in \mathbb{V}$. If the state is contained in Ω_ρ, the state constraint is satisfied, which follows from Prop. 1. From Prop. 1, $-Kx \in \mathbb{U}$ for all $x \in \Omega_\rho$. However, the controller uses the measured output to compute its control action, and $y = x + v$ will not be contained within Ω_ρ for all $x \in \Omega_\rho$ and $v \in \mathbb{V}$, so $-Ky$ may not be contained in $\mathbb{U}$ for all $x \in \Omega_\rho$ and $v \in \mathbb{V}$. Therefore, we consider a subset of Ω_ρ such that we can guarantee that $y \in \Omega_\rho$, ensuring that the control action computed by the explicit controller is in $\mathbb{U}$. Moreover, if the perturbations $f(t)$ applied to the system are sufficiently small, this subset of Ω_ρ is an RPI set for the closed-loop system. This is formally stated in the theorem below.

Theorem 1. *Consider the perturbed closed-loop system in Eq. 3.4 where the eigenvalues of A_{cl} lie inside the unit circle. For any $\rho' < \rho$, there exists $\epsilon_f^* > 0$, such that for any $\epsilon_f \in [0, \epsilon_f^*)$, $\Omega_{\rho'}$ is an RPI set if $\mathbb{F} = \mathbb{W} \times \mathbb{V} \subseteq \mathcal{B}_{\epsilon_f}$. Moreover, $(x, -K(x+v)) \in \mathbb{Z}$ for all $x \in \Omega_{\rho'}$ and $v \in \mathbb{V}$.*

The proof of Theorem 1 is provided in Appendix B. For the perturbed, unconstrained system in Eq. 3.4, the state asymptotically converges to the mRPI set, given that the eigenvalues of $A - BK$ lie inside the unit circle [30]. The mRPI set for the closed-loop system in Eq. 3.4, which we denote as $\mathbb{D}_x$, can be expressed as an infinite Minkowski sum [30]:

$$\mathbb{D}_x = B_{cl}\mathbb{F} \oplus A_{cl}B_{cl}\mathbb{F} \oplus A_{cl}^2 B_{cl}\mathbb{F} \oplus A_{cl}^3 B_{cl}\mathbb{F} \oplus \cdots \tag{3.7}$$

By definition, the mRPI set is a subset of, or equal to, $\Omega_{\rho'}$, provided the perturbations are sufficiently small (Theorem 1). Moreover, since $\mathbb{D}_x \subseteq \Omega_{\rho'} \subset \Omega_\rho \subseteq \mathbb{X}$ and $x \in \Omega_{\rho'}$ implies that $y = x + v \in \Omega_\rho$ for all $v \in \mathbb{V}$, the state and input constraints are satisfied for the closed-loop system in Eq. 3.4 for $t \geq 0$ when $x(0) \in \mathbb{D}_x$ and the perturbations are sufficiently small. With the mRPI set, we can establish a set of measured outputs where we expect the measured output to be after long-term operation of the system in Eq. 3.4 (after the state has converged to $\mathbb{D}_x$), which we call the output terminal set, and is given by:

$$\mathbb{D}_y = \mathbb{D}_x \oplus \mathbb{V} \tag{3.8}$$

When the LQR gain is used as the controller gain, we use $\mathbb{D}_x^*$ and $\mathbb{D}_y^*$ to denote the mRPI and output terminal sets.

In our previous work [15], we used a set-membership-based detection scheme to detect attacks in systems operated under a linear control system. Since many systems are

operated at steady-state for long periods, our previous work addressed systems under steady-state operation. When a system is operated for long periods, the state asymptotically converges to $\mathbb{D}_x$. Once the state converges to $\mathbb{D}_x$, it remains bounded within $\mathbb{D}_x$ after that, provided the system is attack-free. Hence, the boundedness of state within $\mathbb{D}_x$ serves to certify attack-free operations. If the state leaves $\mathbb{D}_x$, the system cannot be attack-free [15]. Since the state is not directly measurable, we can instead monitor values of $y(t)$ and their containment within $\mathbb{D}_y$.

3.2 Connections between MPC and LQR

We establish connections between the systems under MPC and LQR and show that locally (in a compact set containing the origin) the same control actions are applied to both systems, allowing us to analyze the closed-loop system under MPC using the closed-loop system under LQR (an explicit controller). We show that under sufficiently small disturbance and measurement noise, the closed-loop system under MPC possesses the same mRPI set as that under LQR. This result provides a method for computing the mRPI set for the system under MPC.

Lemma 1. *Consider the nominal system under the LQR, initialized at $y(t)$, given by:*

$$\begin{aligned} \tilde{x}(k+1) &= A\tilde{x}(k) + B\tilde{u}(k) \\ \tilde{u}(k) &= -K^*\tilde{x}(k),\ k \in \{t, \dots, t+N-1\} \\ \tilde{x}(t) &= y(t) \end{aligned} \tag{3.9}$$

If $\tilde{x}(k) \in \mathbb{X}$ for $k \in \{t+1, \dots, t+N\}$ and $\tilde{u}(k) \in \mathbb{U}$ for $k \in \{t, \dots, t+N-1\}$, then the input sequence $\{\tilde{u}(t), \dots, \tilde{u}(t+N-1)\}$ is the optimal solution of the optimal control problem in Eq. 2.2.

Proof. The state and input sequences resulting from the system in Eq. 3.9 are feasible

with respect to the problem in Eq. 2.2 if $\tilde{x}(k) \in \mathbb{X}$ for all $k \in \{t+1, \ldots, t+N\}$, $\tilde{x}(t) = y(t)$, and $\tilde{u}(k) \in \mathbb{U}$ for all $k \in \{t, \ldots, t+N-1\}$. By construction, the LQR controller is the optimal control policy for the infinite-horizon optimal control problem given by:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \sum_{k=t}^{\infty} x(k)^T Q x(k) + u(k)^T R u(k) \\ \text{s.t.} \quad & x(k+1) = Ax(k) + Bu(k), \\ & x(t) = y(t) \end{aligned} \tag{3.10}$$

where $\mathbf{u} := \{u(t), u(t+1), \ldots\}$ is the decision variable. The problem in Eq. 3.10 is equivalent to the following finite-horizon optimal control problem with a terminal cost (follows from Bellman's principle of optimality), given by:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \sum_{k=t}^{t+N-1} x(k)^T Q x(k) + u(k)^T R u(k) + x(t+N)^T P x(t+N) \\ \text{s.t.} \quad & x(k+1) = Ax(k) + Bu(k), \\ & x(t) = y(t) \end{aligned} \tag{3.11}$$

where P is the solution to the ARE in Eq. 2.3 and with slight abuse of notation, $\mathbf{u} := \{u(t), \ldots, u(t+N-1)\}$.

The finite-horizon optimal control problem in Eq. 2.2 is similar to the finite-horizon optimal control problem in Eq. 3.11, with the problem in Eq. 2.2 having additional constraints on the state and input sequences, i.e., the two problems have the same cost function and differ only in the constraints. Therefore, the input sequence defined in Eq. 3.9 is an optimal solution to Eq. 2.2 if $\tilde{x}(k) \in \mathbb{X}$ for all $k \in \{t+1, \ldots, t+N-1\}$ and $\tilde{u}(k) \in \mathbb{U}$ for all $k \in \{t, \ldots, t+N-1\}$, following from the fact that adding constraints cannot improve the optimal cost function value. Finally, since $Q \geq 0$, $P \geq 0$, and $R > 0$, the cost function of the problems in Eq. 2.2 and Eq. 3.11 are strictly convex. Therefore, the solution is also unique. □

We can build upon Lemma 1 to show that for any positive invariant set for the nominal system under LQR where the state and inputs constraints are satisfied, the optimal input sequence for the problem in Eq. 2.2 is the sequence defined in Eq. 3.9.

Corollary 1. *Let $\mathcal{X}$ be a positive invariant set for the closed-loop system in Eq. 3.9 such that $(x, -K^*x) \in \mathbb{Z}$ for all $x \in \mathcal{X}$. Then, for all $y(t) \in \mathcal{X}$, the input sequence $\{\tilde{u}(t), \ldots, \tilde{u}(t+N-1)\}$ (defined in Eq. 3.9) is the optimal solution to optimal control problem in Eq. 2.2.*

Given that the MPC and the LQR will lead to the same control action in a region of the desired operating steady-state, a reasonable question is whether they possess the same mRPI set, which offers a straightforward way to compute the mRPI set under MPC. We can show that when the disturbance and measurement noise are sufficiently small, $\mathbb{D}_x^*$ is also the mRPI set under MPC, stated formally in the following theorem.

Theorem 2. *Consider the attack-free closed-loop system consisting of the system in Eq. 2.1 under the MPC defined in Eq. 2.2. If $\|f(t)\| < \epsilon_f^*$ for all $t \geq 0$, the mRPI set for the attack-free closed-loop system under LQR, $\mathbb{D}_x^*$, is also the mRPI set for the attack-free closed-loop system under MPC.*

Proof. Since the set $\mathbb{D}_x^*$ is the mRPI set for the closed-loop system consisting of the system in Eq. 2.1 under the LQR, we start by analyzing the closed-loop system under LQR. From Prop. 1, there exists a positive invariant set for the nominal closed-loop system under LQR, which we denote by Ω_ρ, such that $(x, -K^*x) \in \mathbb{Z}$ for all $x \in \Omega_\rho$ (Ω_ρ is a sublevel set of a Lyapunov function for the closed-loop system under the LQR). For any $\rho' < \rho$, there exist $\epsilon_f^* > 0$ such that for any $\epsilon_f \in [0, \epsilon_f^*)$, $\Omega_{\rho'}$ is a RPI set for the closed-loop system under LQR if $\mathbb{F} \subseteq \mathcal{B}_{\epsilon_f}$ (Thm. 1). Since $\mathbb{D}_x^*$ is the mRPI set, it is contained within $\Omega_{\rho'}$ ($\mathbb{D}_x^* \subseteq \Omega_{\rho'}$).

For any $x(t) \in \Omega_{\rho'}$, $y(t) \in \Omega_\rho$ (this follows from criteria on ϵ_f^* requiring $\epsilon_f^* \leq \rho - \rho'$ discussed in the proof of Thm. 1 such that $x(t) \in \Omega_{\rho'}$ implies $y(t) \in \Omega_\rho$). From Corollary 1, the control action computed by the MPC, i.e., the first element in the optimal sequence, is the same control action computed by the LQR, i.e., $u(t) = -K^*y(t)$ for all $y(t) \in \Omega_\rho$, because Ω_ρ is a positive invariant set for the closed-loop system in Eq. 3.9. Therefore, for any $x(t) \in \Omega_{\rho'}$, the control actions under MPC and LQR are identical. Consequently, with the same initial state, identical control actions, and the same applied disturbances, the successor state, i.e. $x(t+1)$, under both controllers will also be identical. Thus,

$x(t) \in \Omega_{\rho'} \implies x(t+1) \in \Omega_{\rho'}$ holds for the closed-loop system under MPC. Therefore, $\Omega_{\rho'}$ is also the RPI set for the closed-loop system under MPC. Finally, since the two closed-loop systems are identical in $\Omega_{\rho'}$, i.e., the same control action is applied for any state in $\Omega_{\rho'}$, and $\mathbb{D}_x^* \subseteq \Omega_{\rho'}$, the mRPI set for the closed-loop system under MPC is also $\mathbb{D}_x^*$. □

3.3 Attack Detection Scheme

We now present the attack detection scheme that can detect attacks on closed-loop systems under MPC after long-term operation, i.e., after the state converges to the mRPI set. The detection scheme can be outlined as follows:

$$z(t) = \begin{cases} 0, & y(t) \in \mathbb{D}_y^* \\ 1, & y(t) \notin \mathbb{D}_y^* \end{cases} \tag{3.12}$$

where $z(t)$ represents the output of the detection scheme. A condition where $z(t) = 0$ signifies that the system outputs remain bounded within the set $\mathbb{D}_y^*$, indicating the scheme has not detected an attack. If $z(t) = 1$, the detection scheme raises the alarm since the system outputs are no longer bounded within the set $\mathbb{D}_y^*$, indicating abnormal operation is detected.

We prove the proposed detection scheme generates zero false alarms. The principle of ultimate boundedness of the state within the mRPI set and of outputs within the output terminal set serves as the basis for detecting cyberattacks. This is stated in the following theorem.

Theorem 3. *Consider the attack-free closed-loop system consisting of the system in Eq. 2.1 under the MPC with the problem defined in Eq. 2.2. If $x(0) \in \mathbb{D}_x^*$ and $\|f(t)\| < \epsilon_f^*$ for all $t \geq 0$, the detection scheme in Eq. 3.12 does not raise any alarms, i.e., $z(t) = 0$ of all $t \geq 0$.*

Proof. From Thm. 2, there exists $\epsilon_f > 0$ such that $\mathbb{D}_x^*$ is the mRPI set for the attack-free closed-loop system under MPC. Therefore, $x(t) \in \mathbb{D}_x^*$ for all $t \geq 0$ if $x(0) \in \mathbb{D}_y^*$ and the closed-loop system is attack-free. By construction of $\mathbb{D}_y^*$, $y(t) \in \mathbb{D}_y^*$ for all $t \geq 0$. Therefore, $z(t) = 0$ for all $t \geq 0$, and the detection scheme does not raise any alarm. □

In general, attack detection cannot be guaranteed because it depends on the system properties, control system, and detection scheme (e.g., [31, 15, 32, 33]). Nonetheless, a direct consequence of Thm. 3 is the following Corollary:

Corollary 2. *Consider the attack-free closed-loop system consisting of the system in Eq. 2.1 under the MPC with the problem defined in Eq. 2.2. If* $x(0) \in \mathbb{D}_x^*$, $\|f(t)\| < \epsilon_f^*$ *for all* $t \geq 0$, *and* $z(t) = 1$ *for some* $t \geq 0$, *the closed-loop system cannot be attack-free.*

Chapter 4

Application of Cyberattack Detection Scheme to Illustrative Systems

In this chapter, we demonstrate the proposed attack detection scheme on two illustrative systems. All the set computations in this chapter are performed using the MPT toolbox in MATLAB [36]. We use the algorithm presented in [37] with an error bound of $\epsilon = 5 \times 10^{-5}$ to compute the mRPI set.

4.1 Application to a Building Cooling System

In this section, we apply the proposed attack detection scheme to a building space conditioned by a heating, ventilation and air conditioning (HVAC) system to maintain a desired space temperature. External factors, including the ambient temperature and heat gains from solar radiation, affect the temperature inside the building, which are the disturbances

Table 4.1: Model parameters for the building HVAC system.

Parameter	**Value**
Indoor air thermal capacitance (C_{in})	$1.69 \times 10^7\,\text{J}\,^\circ\text{C}^{-1}$
Heat transfer coefficient (h_{amb})	$623\,\text{W}\,^\circ\text{C}^{-1}$
Temperature set point ($T_{\text{in,sp}}$)	$25\,^\circ\text{C}$
Ambient temperature range (T_{amb})	$[28, 38]\ ^\circ\text{C}$
Solar heat rate range ($\dot{Q}_{\text{solar}}$)	$[39, 41]$ kW
Nominal ambient temperature ($T_{\text{amb,nom}}$)	$33\,^\circ\text{C}$
Nominal solar heat rate ($\dot{Q}_{\text{solar,nom}}$)	$40\,\text{kW}$
Steady-state HVAC cooling rate ($\dot{Q}_{\text{HVAC,sp}}$)	$45.0\,\text{kW}$

of the system. The building space's thermal dynamics are given by:

$$C_{\text{in}}\frac{\mathrm{d}T_{\text{in}}}{\mathrm{d}t} = h_{\text{amb}}(T_{\text{amb}} - T_{\text{in}}) + \dot{Q}_{\text{solar}} - \dot{Q}_{\text{HVAC}} \tag{4.1}$$

where C_{in} is the thermal capacitance of the building space, T_{in} is the indoor building space temperature, h_{amb} is the heat transfer coefficient corresponding to the heat transfer between the ambient and the building space, T_{amb} is the temperature of the surroundings, $\dot{Q}_{\text{solar}}$ is the solar heat rate, and $\dot{Q}_{\text{HVAC}}$ is the manipulated rate of cooling rate of the HVAC system. Table 4.1 gives the parameter values.

The control objective is to maintain the temperature inside the facility at a desired set point $T_{\text{in,sp}}$ by manipulating the cooling rate of the HVAC systems. We convert the state variable T_{in}, the manipulated input $\dot{Q}_{\text{HVAC}}$, and the disturbance vector $[T_{\text{amb}}\ \dot{Q}_{\text{solar}}]^T$ to deviation variables such that $x = T_{\text{in}} - T_{\text{in,sp}}$, $u = \dot{Q}_{\text{HVAC}} - \dot{Q}_{\text{HVAC,sp}}$, and $w = [T_{\text{amb}} - T_{\text{amb,nom}}\ \dot{Q}_{\text{solar}} - \dot{Q}_{\text{solar,nom}}]^T$ where $\dot{Q}_{\text{HVAC,sp}}$ is the cooling rate to maintain the inside temperature at $T_{\text{in,sp}}$ when the ambient temperature (T_{amb}) is $33\,^\circ\text{C}$ and solar heat rate ($\dot{Q}_{\text{solar}}$) is $40\,\text{kW}$. We restrict the admissible values of the state and input to $-1.5\ \text{K} \leq x \leq 1.5\ \text{K}$ and $-50\ \text{kW} \leq u \leq 50\ \text{kW}$, respectively. The ambient temperature varies between $[-5, 5]^\circ\text{C}$ around $T_{\text{amb,nom}}$, whereas the solar heat rate varies between $[-1, 1]$ kW around $\dot{Q}_{\text{solar,nom}}$. Additionally, the measured output (T_{in}) is corrupted by bounded measurement

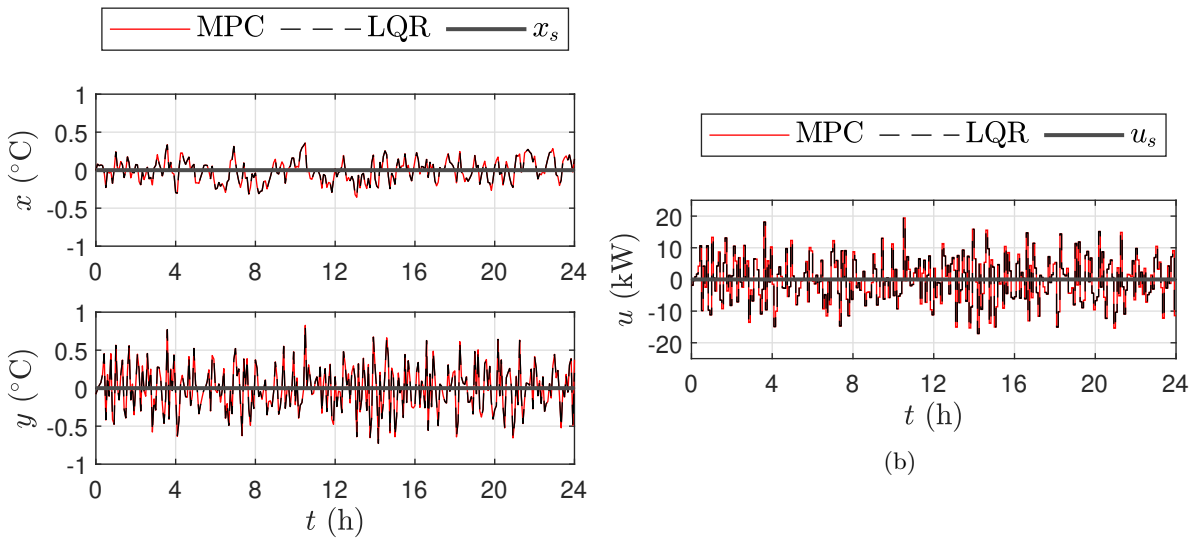


Figure 4.1: Profiles of (a) the actual state (x) and the falsified measured state (y) and (b) control input (u) of the building under the MPC and LQR controller under the attack-free operation. The profiles overlap each other.

noise where the bounded set is given by $[-0.5, 0.5]$°C. We model the disturbance (T_{amb}, $\dot{Q}_{\text{solar}}$) and measurement noise as random variables drawn from a uniform distribution in the interval specified by the admissible bounds set for them. The building temperature is measured every 5 minutes. We perform closed-loop simulations over a 24-hour period. We can represent the building's thermal dynamics, given by Eq. 4.1, as a continuous-time state-space model. We discretize this resulting continuous-time model by considering a zero-order hold of the inputs and disturbances with a sampling period of 5 minutes, obtaining the following discrete-time state-space model with matrices: $A = \begin{bmatrix} 0.989 \end{bmatrix}$, $B = \begin{bmatrix} -1.76 \times 10^{-5} \end{bmatrix}$, and $B_w = \begin{bmatrix} 1.10 \times 10^{-2} & 1.76 \times 10^{-5} \end{bmatrix}$.

4.1.1 Case 1: System in Absence of Attack

In this case, we consider an attack-free operation of the closed-loop system under an MPC. The MPC is tuned using the weighting matrices: $Q = \begin{bmatrix} 10^7 \end{bmatrix}, R = \begin{bmatrix} 10^{-2} \end{bmatrix}$, and $P = \begin{bmatrix} 2.32 \times 10^7 \end{bmatrix}$ where P is the solution to the discrete-time algebraic Riccatti equation.

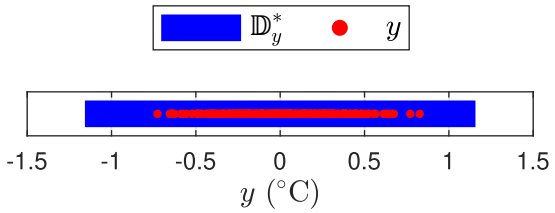
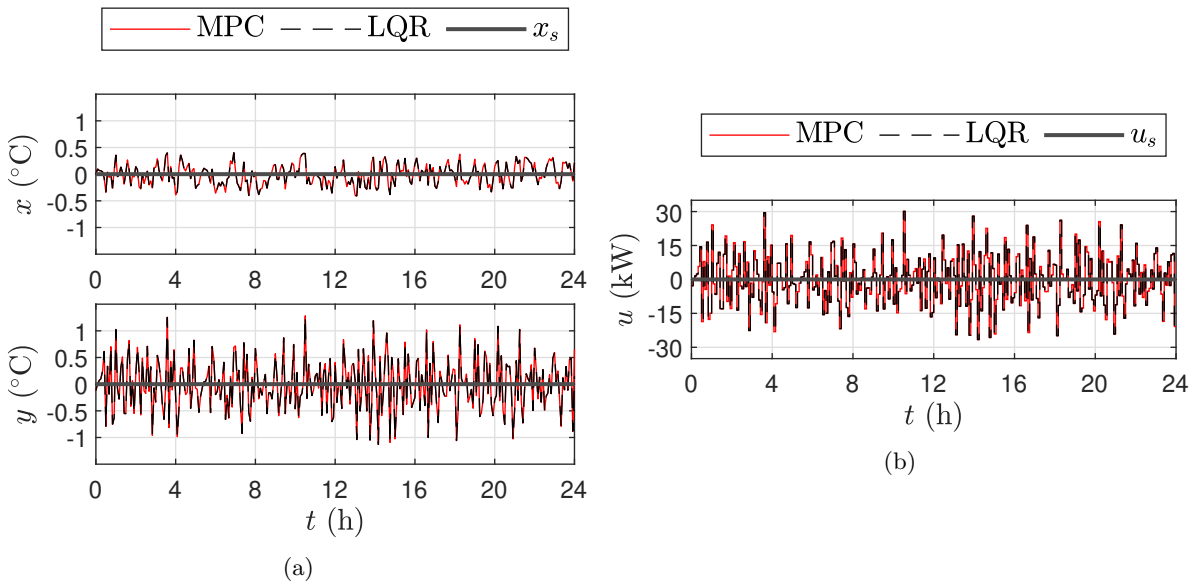


Figure 4.2: The values of the measured output for one of the attack-free closed-loop simulations under MPC. The output trajectory $y(t)$ evolves inside $\mathbb{D}_y^*$ for all times, indicating that no attack is detected.

We assess if the attack detection scheme presented in Eq. 3.12 correctly verifies that the system is not under attack and does not generate false alarms. First, we compare the closed-loop operation of the system under MPC, given by Eq. 2.2, to that under an LQR. The MPC is tuned with weighting matrices Q and R and a quadratic terminal cost where the terminal weighting matrix P is obtained by solving the algebraic Riccatti equation. The LQR is tuned using the same choices of Q and R. We initialize both closed-loop systems with an initial state of 0. The same realization of the disturbance and measurement noise is used in both closed-loop simulations. The eigenvalue of $A_{cl} = A - BK^*$ for the closed-loop system under LQR is $\lambda(A_{cl}) = 0.575$, which indicates that the closed-loop operation under LQR is stable. As shown in Figure 4.1b, the closed-loop input trajectories under both controllers MPC and LQR are identical, which is consistent with the results established in Lemma 1.

The mRPI set ($\mathbb{D}_x^*$) and output terminal set ($\mathbb{D}_y^*$) for the attack-free closed-loop system under LQR is computed using Eqs. 3.7 and 3.8. We estimate the mRPI set using the method presented in [37] and use it to compute the estimate of set $\mathbb{D}_y^*$. We perform 1000 closed-loop simulations for the system under MPC using different disturbance and measurement noise realizations. The output values of the system evolve within $\mathbb{D}_y^*$, i.e., $y(t) \in \mathbb{D}_y^*$, for all simulations. This follows from Theorem 2 as the mRPI set under LQR is the mRPI set for the attack-free closed-loop system under MPC. Therefore, no attack



(a) (b)

Figure 4.3: Profiles of (a) the actual state (x) and the falsified measured state (y) and (b) control input (u) of the building under the MPC and LQR controller under the attack. The profiles overlap each other.

is detected in any simulation, confirming that this detection approach does not raise false alarms, as expected from the findings of Theorem 3. The measured output values observed for one of the simulations and the set $\mathbb{D}_y^*$ are shown in Figure 4.2.

4.1.2 Case 2: System in Presence of Attack

In this case, we consider the operation of a closed-loop system under MPC in the presence of a multiplicative sensor-controller link false data injection attack with $y_a(t) = \Lambda y(t)$ where $y_a(t)$ is the altered measurement received by the controller and Λ is the multiplicative factor that multiplies the measured output. In this case study, we consider $\Lambda = 1.5$. We assess if the attack detection scheme presented in Eq. 3.12 detects that the system is under attack. Again, we consider the MPC and LQR to have the same tuning parameters as in the previous case. We initialize the system with an initial state of 0 for both closed-loop systems. The eigenvalue of $A_{cl} = A - BK^*\Lambda$ is $\lambda(A_{cl}) = 0.368$, indicating that the closed-loop operation under LQR and subjected to the multiplicative attack is still stable.

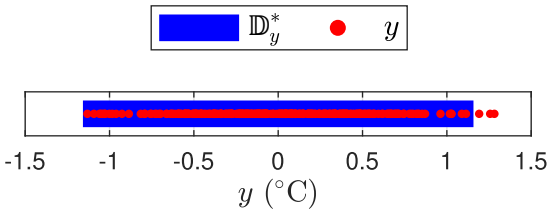


Figure 4.4: The values of the measured output for one of the closed-loop simulations under MPC and subject to a multiplicative attack on the sensor-controller link. The output trajectory $y(t)$ evolves outside $\mathbb{D}_y^*$, for some time, and the attack is detected.

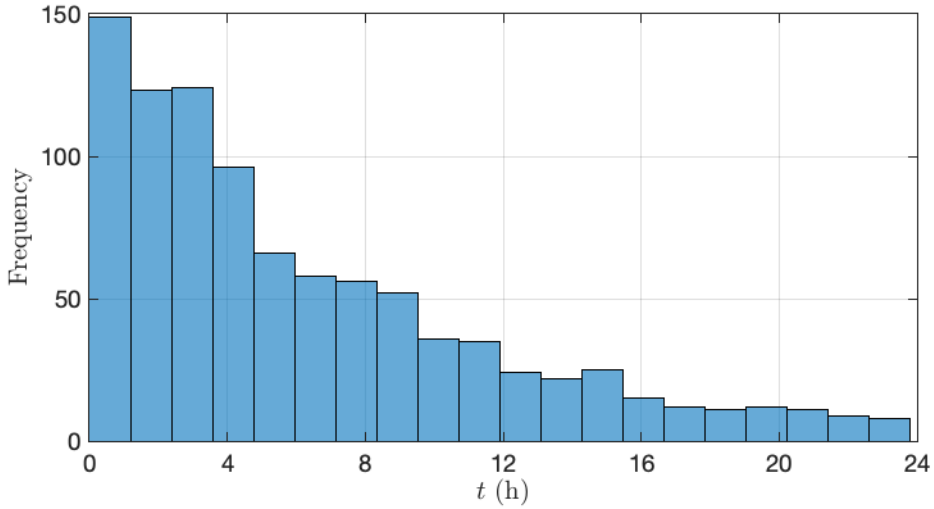


Figure 4.5: Distribution of attack detection times across the 944 simulations in which the attack was detected.

As shown in Figure 4.3b, the system's control input trajectories under MPC and under LQR are overlapping, indicating that they are equivalent under both controllers. As a result, system's state and output trajectories under MPC and under LQR are overlapping, as shown in Figure 4.3a.

We apply the detection scheme to 1000 closed-loop simulations with different disturbance and measurement noise realizations. In 944 of those simulations, at least one output value is outside $\mathbb{D}_y^*$, and an alarm is raised, demonstrating that the attack is detected in these 944 cases. The average time for detection across the 944 simulations is 6.28 hours. To assess the significance of this detection time, we compare it with the time constant of the system. The time constant of the system is approximately 7.53 hours. In the

Table 4.2: Model parameters for CSTR [38].

Parameter	**Value**
Feed concentration of the reactant (C_{A0})	4 kmol m^{-3}
Feed temperature (T_0)	300 K
Volumetric feed flow rate (F)	5 m^3 h^{-1}
Volume of reactor (V)	1 m^3
Heat of reaction (ΔH)	-1.15×10^4 kJ $kmol^{-1}$
Density of liquid (ρ_L)	1000 kg m^{-3}
Heat capacity of liquid (C_p)	0.231 kJ kg^{-1} K^{-1}
Activation energy (E)	5.0×10^4 kJ $kmol^{-1}$
Pre-exponential factor (k_0)	8.46×10^6 m^3 h^{-1} $kmol^{-1}$
Gas constant (R)	8.314 kJ $kmol^{-1}$ K^{-1}
Steady-state concentration of the reactant (C_{As})	1.22 kmol m^{-3}
Steady-state temperature (T_s)	438.2 K
Steady-state heat rate added/removed from reactor ($\dot{Q}_{\text{heat,s}}$)	0 kJ h^{-1}

remaining 56 simulations, the output values remains within $\mathbb{D}_y^*$ for the entire duration of the simulation, and hence, the attack goes undetected. Figure 4.4 demonstrates the result for one of the 1000 simulations. The output trajectory $y(t)$ evolves outside $\mathbb{D}_y^*$, for some time, indicating that the attack is detected. Figure 4.5 shows a histogram representing the distribution of detection times across the 944 simulations in which the attack was detected. The distribution indicates that in 637 simulations, detection occurred within one time constant of the attack onset; in 861 simulations, detection occurred within two time constants; and in 936 simulations, detection was achieved within three time constants—demonstrating the responsiveness of the proposed detection scheme.

4.2 Application to a Continuous Stirred Tank Reactor

In this section, we demonstrate the detection scheme's applicability to a nonlinear process. We consider a nonlinear chemical process consisting of a continuous stirred tank reactor (CSTR) where a second order, irreversible, exothermic reaction takes place of the form $A \longrightarrow B$. We derive the mass and energy balance of the CSTR using standard modeling

principles, and the resulting system of ordinary differential equations is as follows:

$$\begin{aligned}\frac{dC_A}{dt} &= \frac{F}{V}(C_{A0} + \Delta C_{A0} - C_A) - k_0 e^{\frac{-E}{RT}} C_A^2 \\ \frac{dT}{dt} &= \frac{F}{V}(T_0 + \Delta T_0 - T) - \frac{\Delta H k_0}{\rho_L C_p} e^{\frac{-E}{RT}} C_A^2 + \frac{\dot{Q}_{\text{heat}}}{\rho_L C_p V}\end{aligned} \tag{4.2}$$

where F is the volumetric flow rate, V is the volume of the tank, C_{A0} is the concentration of the reactant fed into the tank, T_0 is the temperature of the feed, C_A is the concentration of reactant, T is the temperature of the tank, ΔH is the heat of reaction, ρ_L and C_p are the density and the heat capacity of the liquid in the reactor respectively, k_0 is the pre-exponential factor, R is the universal gas constant and $\dot{Q}_{\text{heat}}$ is the heat rate added or removed from the reactor. The variables ΔC_{A0} and ΔT_0 represent bounded disturbance in the feed reactant concentration and temperature, respectively. The control objective is to operate the CSTR around the steady-state values for concentration (C_{As}) and temperature (T_s). Process disturbances ΔC_{A0} and ΔT_0 are modelled as bounded deviations from C_{As} and T_s, respectively. We provide the values for all these parameters in Table 4.2.

We define our state variables in the form of deviation variables such that $x = [x_1\ x_2]^T = [C_A - C_{As}\ T - T_s]^T$ and $u = \dot{Q}_{\text{heat}} - \dot{Q}_{\text{heat,s}}$. The admissible values of the state and input are restricted to: -0.02 kmol m^{-3} $\leq x_1 \leq$ 0.02 kmol m^{-3}, -1 K $\leq x_2 \leq$ 1 K, and -15000 kJ/hr $\leq u \leq$ 15000 kJ/hr respectively. The disturbance subject to the system can be given by $\Delta C_{A0} \in [-1, 1]$ kmol m^{-3} and $\Delta T_0 \in [-5, 5]$ K. Additionally, the outputs of the process are subject to measurement noise described by $[-0.01, 0.01]$ kmol m^{-3} and $[-0.2, 0.2]$ K. The disturbances and measurement noise are modeled as random variables drawn from a uniform distribution in the interval specified by the admissible bounds set for them. Given the small bounds chosen for the disturbances and noise, the linearized model may approximate the nonlinear process. The impact of the nonlinearities increases

with the value of disturbances and noise.

To solve the optimal control problem in MPC, we need the linearized state-space model of the system. The model is linearized around the steady-state point at which it is being operated. We discretize the resulting linearized continuous-time model with a sampling time of 10^{-2} h yielding a discrete-time linear state-space model of the form of Eq. 2.1 where the matrices are given by:

$$A = \begin{bmatrix} 0.736 & -0.0041 \\ 10.695 & 1.156 \end{bmatrix}, \; B = \begin{bmatrix} -9.071 \times 10^{-8} \\ 4.674 \times 10^{-5} \end{bmatrix}, \; B_w = \begin{bmatrix} 0.0433 & -1.048 \times 10^{-4} \\ 0.272 & 0.0540 \end{bmatrix}$$

We use the explicit Euler's method with a step size of 10^{-4} h to integrate the ordinary differential equations in Eq. 4.2 to perform closed-loop simulations.

4.2.1 Case 1: System in Absence of Attack

In this case, we consider an attack-free operation of the closed-loop system under MPC. The MPC is tuned with the following weighting matrices:

$$Q = \begin{bmatrix} 5 \times 10^{8} & 0 \\ 0 & 2 \times 10^{10} \end{bmatrix}, \; R = \begin{bmatrix} 1 \end{bmatrix}, \; P = \begin{bmatrix} 1.24 \times 10^{11} & 5.46 \times 10^{9} \\ 5.46 \times 10^{9} & 2.06 \times 10^{10} \end{bmatrix}$$

where P is the solution to the algebraic Riccatti equation. We assess if the attack detection scheme presented in Eq. 3.12 verifies that the CSTR model is not under attack. We compare the closed-loop response of the system under MPC, given by Eq. 2.2 to that under an LQR. Again, the MPC is tuned with weighting matrices Q and R and a quadratic terminal cost, where the terminal weighting matrix P is obtained by solving the algebraic Riccatti equation. The LQR is tuned using the same choices of Q and R. We initialize both systems with an initial state of 0 and use the same realization of disturbance and

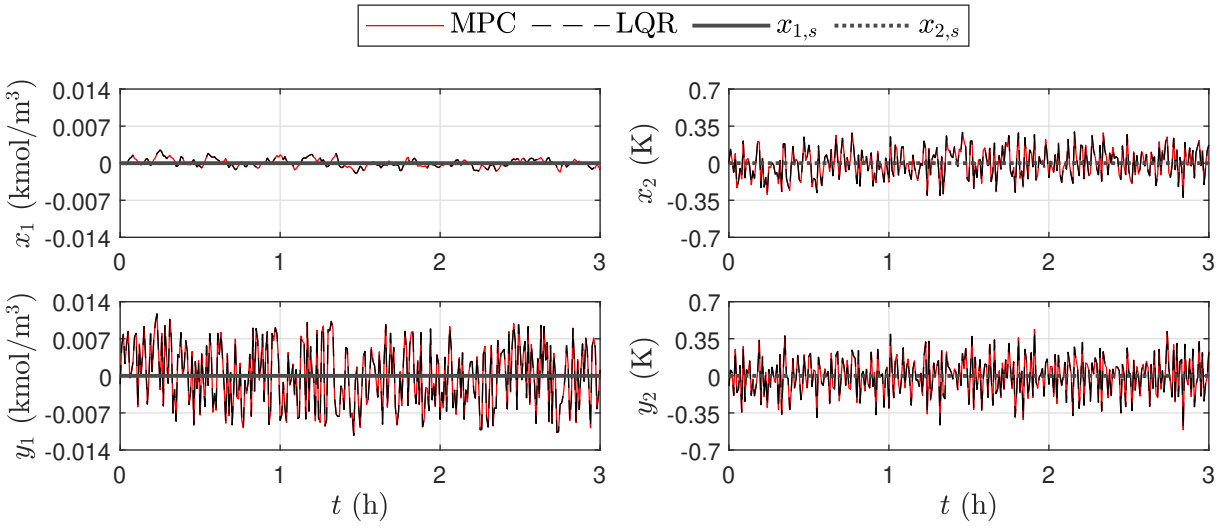


Figure 4.6: Profiles of the actual states (x_1, x_2) and the measured states (y_1, y_2) of the reactor under the MPC and LQR controller under the attack-free operation. The profiles overlap each other.

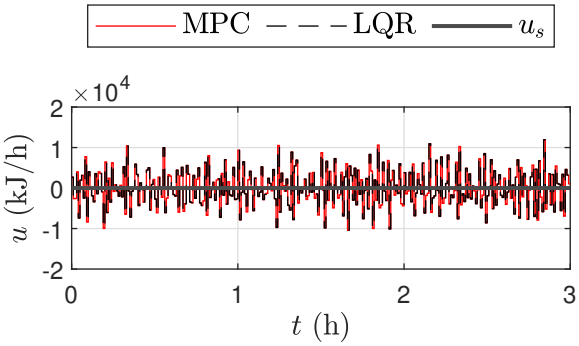


Figure 4.7: Profile of the control input (u) of the reactor under the MPC and LQR controller under attack-free operation. The profiles overlap with each other.

measurement noise for both closed-loop systems. The maximum magnitude of the eigenvalue of $A_{cl} = A - BK^*$ is given by $\max|\lambda(A_{cl})| = 0.852$, indicating that the attack-free closed-loop operation is stable.

As shown in Figure 4.7, the control input trajectories of the MPC and LQR overlap, indicating they are equivalent under both controllers. Consequently, the state and output trajectories of the system under MPC and LQR overlap, as shown in Figure 4.6. The mRPI set and the output terminal set for the system under LQR is approximated using Eqs. 3.7 and 3.8 and an estimate of the set is obtained using the method mentioned in [37]. We perform 1000 closed-loop simulations for the system under MPC using different

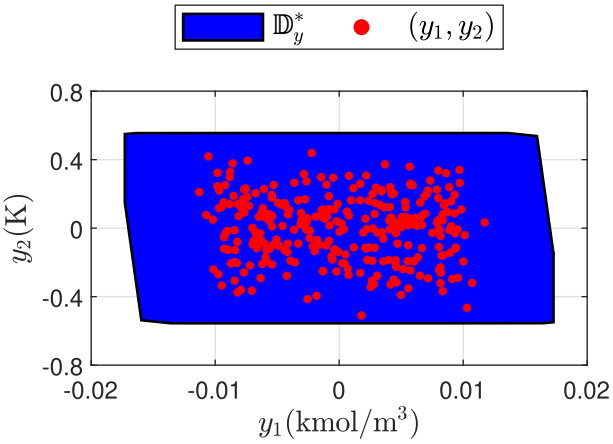


Figure 4.8: The values of the measured output for one of the attack-free closed-loop simulations under MPC. The output trajectory $y(t)$ evolves inside $\mathbb{D}_y^*$ for all times, indicating that no attack is detected.

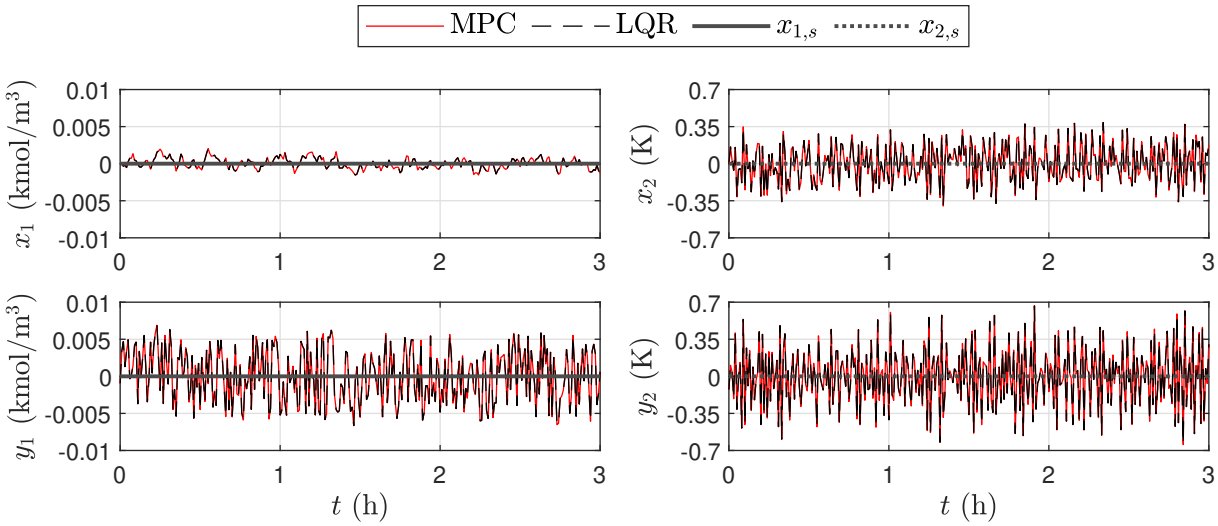


Figure 4.9: Profiles of the actual states (x_1, x_2) and the falsified measured states (y_1, y_2) of the reactor under the MPC and LQR controller under the attack. The profiles overlap with each other.

realizations of disturbances and measurement noise. The outputs of the process evolve within $\mathbb{D}_y^*$, i.e., $y(t) \in \mathbb{D}_y^*$, in all these simulations. Consequently, no attack is detected in the system, and no false alarms are raised. The measured output values observed for one of the simulations and the set $\mathbb{D}_y^*$ are shown in Figure 4.8.

4.2.2 Case 2: System in Presence of Attack

In this case, we consider the operation of a closed-loop system under MPC in the presence of a multiplicative attack where $y_a(t) = \Lambda y(t)$ and $\Lambda = \text{diag}(0.6,1.3)$. We assess if the

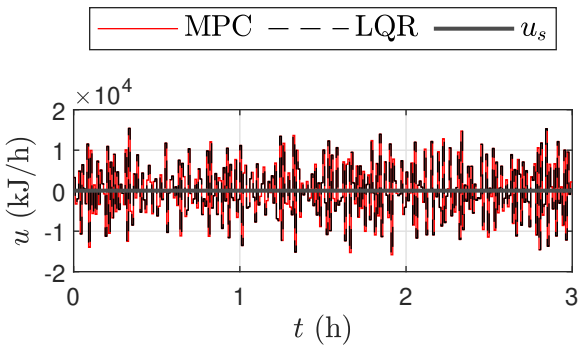


Figure 4.10: Profiles of the control input (u) of the reactor under the MPC and LQR controller under the attack. The profiles overlap with each other.

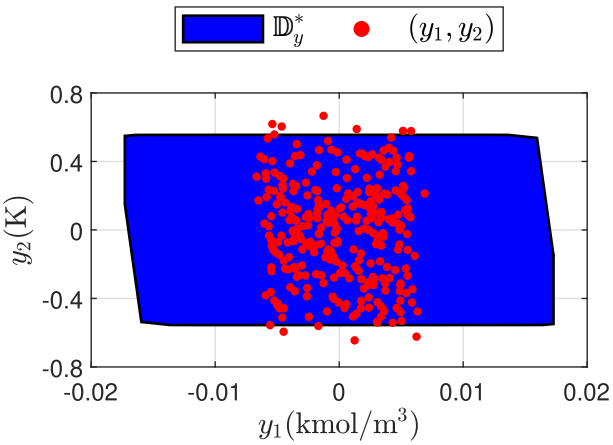


Figure 4.11: The values of the measured output for one of the closed-loop simulations under MPC and subject to a multiplicative attack on the sensor-controller link. The output trajectory $y(t)$ evolves outside $\mathbb{D}_y^*$ for some time, and the attack is detected.

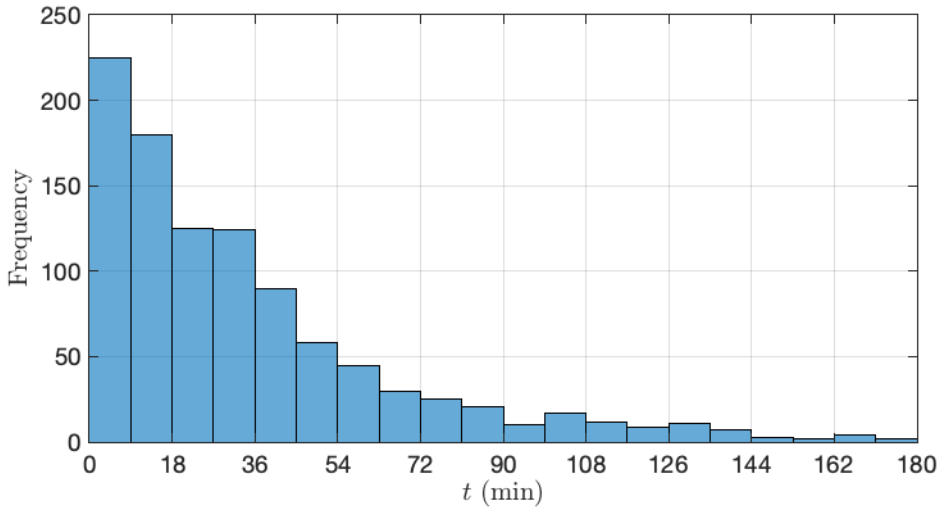


Figure 4.12: Distribution of attack detection times across the 1000 simulations in which the attack was detected.

attack detection scheme presented in Eq. 3.12 detects the attack. Again, we consider the MPC and LQR to have the same tuning parameters as in the previous case. We initialize both systems with an initial state of 0 and use the same realization of disturbance and measurement noise for both closed-loop systems. The maximum magnitude of the eigenvalue of the $A_{cl} = A - BK^*\Lambda$ matrix for the closed-loop system controlled by LQR under attack is given by $\max_i |\lambda_i(A_{cl})| = 0.744$, indicating that the closed-loop operation is stable under attack.

As shown in Figures 4.9 and 4.10, the state, output, and control input trajectories of the system under MPC and LQR overlap, indicating they are equivalent under both controllers. We apply the detection scheme to 1000 closed-loop simulations with different disturbance and measurement noise realization. In all of those simulations, the $y(t)$ evolves outside $\mathbb{D}_y^*$, and an alarm is raised, demonstrating that the attack is detected in all cases. Figure 4.11 demonstrates the result of one of the 1000 simulations. At least one of the output values evolve outside $\mathbb{D}_y^*$, indicating that the system is under an attack. The average time for detection across the 1000 simulations is 33.83 minutes. To assess the significance of this detection time, we compare it with the time constant of the system. The time constant of the system is approximately 11.55 minutes. Figure 4.12 shows a histogram representing the distribution of detection times across the 1000 simulations in which the attack was detected. The distribution indicates that in 287 simulations, detection occurred within one time constant of the attack onset; in 487 simulations, detection occurred within two time constants; and in 631 simulations, detection was achieved within three time constants—demonstrating the responsiveness of the proposed detection scheme.

Remark 1. *The detection rate observed in the case studies, i.e., the number of times*

the attack is and is not detected, and the average time of detection across the 1000 simulations, is valid for the specific systems, attacks, and scenarios considered and does not necessarily generalize to other systems and attacks, as the detection rate will depend on system properties, attack properties, and controller design/parameters.

Chapter 5

Conclusion

In this work, we presented a set-membership-based detection scheme for systems under MPC. Specifically, we established theoretical conditions to compute the terminal set for a pre-defined monitoring variable. We then utilized it to detect attacks using a set-membership-based detection that guarantees zero false alarms. We proved that, provided that the disturbances and measurement noise are sufficiently small, the mRPI set for a system under MPC, tuned with weighting matrices Q and R and a quadratic terminal cost where the terminal weighting matrix P is the solution to the algebraic Riccatti equation, is equivalent to the mRPI set under LQR, tuned using the same choice of Q and R. Consequently, we presented an attack detection scheme and derived conditions under which the detection scheme does not generate false alarms. The efficacy of the detection scheme was then demonstrated using an illustrative building space cooling system. Finally, the proposed detection scheme's applicability to nonlinear systems was then demonstrated using a nonlinear chemical process. Future research could potentially focus on classifying attacks based on their detectability properties with respect to the proposed detection scheme.

Bibliography

[1] M. Nankya, R. Chataut, R. Akl, Securing industrial control systems: components, cyber threats, and machine learning-driven defense strategies, Sensors 23 (2023) 8840. doi:10.3390/s23218840.

[2] H. S. Sánchez, D. Rotondo, T. Escobet, V. Puig, J. Quevedo, Bibliographical review on cyber attacks from a control oriented perspective, Annual Reviews in Control 48 (2019) 103–128. doi:10.1016/j.arcontrol.2019.08.002.

[3] R. Canonico, G. Sperlì, Industrial cyber-physical systems protection: A methodological review, Computers & Security 135 (2023) 103531. doi:10.1016/j.cose.2023.103531.

[4] N. Mims, Cyber-attack process, in: Computer and Information Security Handbook (Third Edition), Elsevier, Boston, 2017, pp. 1105–1116. doi:10.1016/B978-0-12-803843-7.00084-3.

[5] K. E. Hemsley, E. Fisher, History of industrial control system cyber incidents, Tech. rep., Idaho National Lab. (2018). doi:10.2172/1505628.

[6] H. Kayan, M. Nunes, O. Rana, P. Burnap, C. Perera, Cybersecurity of industrial cyber-physical systems: A review, ACM Computing Surveys 54 (2022) 1–35. doi:10.1145/3510410.

[7] W. Duo, M. Zhou, A. Abusorrah, A survey of cyber attacks on cyber physical systems: Recent advances and challenges, IEEE/CAA Journal of Automatica Sinica 9 (2022) 784–800. doi:10.1109/JAS.2022.105548.

[8] M. H. Rahman, T. Wuest, M. Shafae, Manufacturing cybersecurity threat attributes and countermeasures: Review, meta-taxonomy, and use cases of cyber-attack taxonomies, Journal of Manufacturing Systems 68 (2023) 196–208. doi:10.1016/j.jmsy.2023.03.009.

[9] M. Chamanbaz, F. Dabbene, R. Bouffanais, A physics-based attack detection technique in cyber-physical systems: A model predictive control co-design approach, in: Proceedings of the Australian & New Zealand Control Conference, 2019, pp. 18–23. doi:10.1109/ANZCC47194.2019.8945588.

[10] H. Oyama, D. Messina, K. K. Rangan, H. Durand, Lyapunov-based economic model predictive control for detecting and handling actuator and simultaneous sensor/actuator cyberattacks on process control systems, Frontiers in Chemical Engineering 4 (2022) 810129. doi:10.3389/fceng.2022.810129.

[11] S. Chen, Z. Wu, P. D. Christofides, Cyber-attack detection and resilient operation of nonlinear processes under economic model predictive control, Computers & Chemical Engineering 136 (2020) 106806. doi:10.1016/j.compchemeng.2020.106806.

[12] A. Zedan, N. H. El-Farra, A machine-learning approach for identification and mitigation of cyberattacks in networked process control systems, Chemical Engineering Research and Design 176 (2021) 102–115. doi:10.1109/JPROC.2015.2512235.

[13] G. Wu, Y. Wang, Z. Wu, Physics-informed machine learning in cyber-attack detec-

tion and resilient control of chemical processes, Chemical Engineering Research and Design 204 (2024) 544–555. doi:10.1016/j.cherd.2024.03.014.

[14] C. Murguia, J. Ruths, On model-based detectors for linear time-invariant stochastic systems under sensor attacks, IET Control Theory & Applications 13 (2019) 1051–1061. doi:10.1049/iet-cta.2018.5970.

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

[16] S. Narasimhan, N. H. El-Farra, M. J. Ellis, Active multiplicative cyberattack detection utilizing controller switching for process systems, Journal of Process Control 116 (2022) 64–79. doi:10.1016/j.jprocont.2022.05.014.

[17] S. Narasimhan, N. H. El-Farra, M. J. Ellis, A control-switching approach for cyberattack detection in process systems with minimal false alarms, AIChE Journal 68 (2022) e17875. doi:10.1002/aic.17875.

[18] J. H. Lee, Model predictive control: Review of the three decades of development, International Journal of Control, Automation and Systems 9 (2011) 415–424. doi:10.1007/s12555-011-0300-6.

[19] S. J. Qin, T. A. Badgwell, A survey of industrial model predictive control technology, Control engineering practice 11 (2003) 733–764. doi:10.1016/S0967-0661(02)00186-7.

[20] D. Q. Mayne, J. B. Rawlings, C. V. Rao, P. O. Scokaert, Constrained model predictive

control: Stability and optimality, Automatica 36 (2000) 789–814. doi:10.1016/S0005-1098(99)00214-9.

[21] G. Franzè, W. Lucia, F. Tedesco, Resilient model predictive control for constrained cyber-physical systems subject to severe attacks on the communication channels, IEEE Transactions on Automatic Control 67 (2021) 1822–1836. doi:10.1109/TAC.2021.3084237.

[22] D. Bertsekas, Dynamic programming and optimal control, 3rd Edition, Vol. I, Athena Scientific, 2005.

[23] R. Canonico, G. Sperlì, Industrial cyber-physical systems protection: A methodological review, Computers & Security 135 (2023) 103531. doi:10.1016/j.cose.2023.103531.

[24] S. Parker, Z. Wu, P. D. Christofides, Cybersecurity in process control, operations, and supply chain, Computers & Chemical Engineering (2023) 108169doi:10.1016/j.compchemeng.2023.108169.

[25] H. T. Reda, A. Anwar, A. Mahmood, Comprehensive survey and taxonomies of false data injection attacks in smart grids: attack models, targets, and impacts, Renewable and Sustainable Energy Reviews 163 (2022) 112423. doi:10.1016/j.rser.2022.112423.

[26] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, R. Candell, A survey of physics-based attack detection in cyber-physical systems, ACM Computing Surveys 51 (2018) 1–36. doi:10.1145/3203245.

[27] C. Murguia, J. Ruths, On reachable sets of hidden CPS sensor attacks, in: Pro-

ceedings of the Annual American Control Conference, 2018, pp. 178–184. doi:10.23919/ACC.2018.8431582.

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

[29] S. Narasimhan, N. H. El-Farra, M. J. Ellis, A reachable set-based scheme for the detection of false data injection cyberattacks on dynamic processes, Digital Chemical Engineering 7 (2023) 100100. doi:10.1016/j.dche.2023.100100.

[30] V. Kuntsevich, B. Pshenichnyi, Minimal invariant sets of dynamic systems with bounded disturbances, Cybernetics and Systems Analysis 32 (1996) 58–64. doi:10.1155/S1024123X98000866.

[31] W. Duo, M. Zhou, A. Abusorrah, A survey of cyber attacks on cyber physical systems: Recent advances and challenges, IEEE/CAA Journal of Automatica Sinica 9 (2022) 784–800. doi:10.1109/JAS.2022.105548.

[32] N. Mtukushe, A. K. Onaolapo, A. Aluko, D. G. Dorrell, Review of cyberattack implementation, detection, and mitigation methods in cyber-physical systems, Energies 16 (2023) 5206. doi:10.3390/en16135206.

[33] W. Xing, J. Shen, Security control of cyber–physical systems under cyber attacks: A survey, Sensors 24 (2024) 3815.

[34] H. K. Khalil, Nonlinear systems, Prentice-Hall, Upper Saddle River, NJ, 2002.

[35] P. D. Christofides, N. El-Farra, Control of nonlinear and hybrid process systems:

Designs for uncertainty, constraints and time-delays, Vol. 324, Springer Science & Business Media, 2005.

[36] M. Kvasnica, P. Grieder, M. Baotić, M. Morari, Multi-parametric toolbox (MPT), in: Proceedings of the 7th International Workshop on Hybrid Systems: Computation and Control, Philadelphia, PA, USA, 2004, pp. 448–462. doi:10.1007/978-3-540-24743-2_30.

[37] S. Rakovic, E. Kerrigan, K. Kouramas, D. Mayne, Invariant approximations of the minimal robust positively invariant set, IEEE Transactions on Automatic Control 50 (2005) 406–410. doi:10.1109/TAC.2005.843854.

[38] A. Alanqar, M. Ellis, P. D. Christofides, Economic model predictive control of nonlinear process systems using empirical models, AIChE Journal 61 (2015) 816–830. doi:10.1002/aic.14683.

Appendix A

Proposition 1 Proof

Proof. Since the eigenvalues of $A - BK$ lie within the unit circle, there exists a positive definite matrix $\hat{P} \in \mathbb{R}^{n_x \times n_x}$ such that $V(x) = x^T \hat{P} x$ is a Lyapunov function for the closed-loop system. Since the pair (x_s, u_s) is contained within the interior of $\mathbb{Z}$ and from the continuity of $-Kx$ with respect to x, there exist $\rho > 0$ such that:

$$\{(x, -Kx) | x \in \Omega_\rho\} \subset \mathbb{Z}$$

where Ω_ρ is a sublevel set of the Lyapunov function, and thus, it is a positive invariant set. □

Appendix B

Theorem 1 Proof

Proof. Consider the set Ω_ρ, a sublevel set of the Lyapunov function defined in the proof of Prop. 1 for the nominal system under the explicit stabilizing controller. For any $\rho' < \rho$, there exists $\epsilon_v^* > 0$ such that $\rho' + \epsilon_v^* = \rho$, i.e., $\Omega_{\rho'} := \{x \in \mathbb{R}^{n_x} | x^T \hat{P} x \leq \rho'\} \oplus \mathcal{B}_{\epsilon_v^*} = \Omega_\rho$. For any $\mathbb{V} \subseteq \mathcal{B}_{\epsilon_v^*}$, $-K(x+v) \in \mathbb{U}$ for all $x \in \Omega_{\rho'}$ and $v \in \mathbb{V}$. For any $x \in \Omega_{\rho'}$, the state constraint is also satisfied because $\Omega_{\rho'} \subset \Omega_\rho \subset \mathbb{X}$.

Consider the difference of the Lyapunov function at successive times:

$$V(x(t+1)) - V(x(t)) = x(t+1)^T \hat{P} x(t+1) - x(t) \hat{P} x(t)$$

Using Eq. 3.4, the above equation can be expressed as follows:

$$V(x(t+1)) - V(x(t)) = x(t)^T (A_{cl}^T \hat{P} A_{cl} - \hat{P}) x(t) + 2x(t)^T A_{cl}^T \hat{P} B_{cl} f(t) + f(t)^T B_{cl}^T \hat{P} B_{cl} f(t)$$

Given that $V(x) = x^T \hat{P} x$ is a Lyapunov function for the nominal closed-loop system, there exists a positive definite matrix $\hat{Q}$ such that $\hat{P}$ satisfies the Lyapunov equation $A_{cl} P A_{cl} - P = -\hat{Q}$. From the sub-multiplicative property and the triangle inequality, the difference between the Lyapunov function values at the two-time steps can be bounded from above by:

$$V(x(t+1)) - V(x(t)) \leq -x(t)^T \hat{Q} x(t) + 2\|A_{cl}^T \hat{P} B_{cl}\| \|x(t)\| \|f(t)\| + \|B_{cl}^T \hat{P} B_{cl}\| \|f(t)\|^2$$

Let $c_{f,1} = \|A_{cl}^T \hat{P} B_{cl}\|$ and $c_{f,2} = \|B_{cl}^T \hat{P} B_{cl}\|$. Since $\hat{Q}$ is positive definite, $x^T \hat{Q} x$ is a norm. Owing to the equivalence of norms, $c_x > 0$ exists such that $x^T \hat{Q} x \geq c_x \|x\|^2$ for all x.

Therefore,

$$V(x(t+1)) - V(x(t)) \leq -c_x\|x(t)\|^2 + 2c_{f,1}\|x(t)\|\|f(t)\| + c_{f,2}\|f(t)\|^2.$$

Since $\Omega_{\rho'}$ is compact, $R > 0$ exists such that $\|x(t)\| \leq R$ for all $x(t) \in \Omega_{\rho'}$. Thus,

$$V(x(t+1)) - V(x(t) \leq -c_x\|x(t)\|^2 + 2Rc_{f,1}\|f(t)\| + c_{f,2}\|f(t)\|^2 \tag{B.1}$$

for all $x(t) \in \Omega_{\rho'}$. Let $r' = \max_{\hat{r}} \hat{r}$ s.t. $\mathcal{B}_{\hat{r}} \subseteq \Omega_{\rho'}$ where $r' > 0$ since $\rho' > 0$. For any $r \in (0, r')$, there exists $\epsilon^*_{f,1} > 0$ defined by:

$$\epsilon^*_{f,1} := \frac{-Rc_{f,1} + \sqrt{R^2c^2_{f,1} + r^2c_xc_{f,2}}}{c_{f,2}} \tag{B.2}$$

If $\|f(t)\| < \epsilon^*_{f,1}$ for all $f(t) \in \mathbb{F}$, i.e., $\mathbb{W}\times\mathbb{V} =: \mathbb{F} \subset \mathcal{B}_{\epsilon^*_{f,1}}$, then the value of the Lyapunov function will decrease at the next time step, provided $x(t) \in \Omega_{\rho'} \setminus \mathcal{B}_r$. To show this, let $\epsilon_{f,1} := \max_{f\in\mathbb{F}}\|f\|$ (recall $\mathbb{F}$ is compact) and

$$\theta := \frac{2Rc_{f,1}\epsilon_{f,1} + c_{f,2}\epsilon^2_{f,1}}{r^2c_x} \tag{B.3}$$

Since $c_{f,1}$, $c_{f,2}$, c_x, R, and r are strictly positive numbers, $\theta > 0$. Moreover,

$$2Rc_{f,1}\epsilon_{f,1} + c_{f,2}\epsilon^2_{f,1} < 2Rc_{f,1}\epsilon^*_{f,1} + c_{f,2}{\epsilon^*_{f,1}}^2 \tag{B.4}$$

From Eq. B.2, we obtain $r^2c_x = 2Rc_{f,1}\epsilon^*_{f,1} + c_{f,2}{\epsilon^*_{f,1}}^2$ and

$$2Rc_{f,1}\epsilon_{f,1} + c_{f,2}\epsilon^2_{f,1} < r^2c_x \tag{B.5}$$

showing that $\theta < 1$. Thus, $\theta \in (0, 1)$. For any $x(t) \in \Omega_{\rho'} \setminus \mathcal{B}_r$,

$$V(x(t+1)) - V(x(t)) \leq -(1-\theta)c_x\|x(t)\|^2 - \theta c_x r^2 + 2Rc_{f,1}\epsilon_{f,1} + c_{f,2}\epsilon^2_{f,1} \tag{B.6}$$

From Eq. B.3, we obtain $\theta c_x r^2 = 2Rc_{f,1}\epsilon_{f,1} + c_{f,2}\epsilon^2_{f,1}$ and

$$V(x(t+1)) - V(x(t)) \leq -(1-\theta)c_x\|x(t)\|^2 \tag{B.7}$$

for any $\|f(t)\| \leq \epsilon_{f,1} < \epsilon^*_{f,1}$ and $x(t) \in \Omega_{\rho'} \setminus \mathcal{B}_r$. Since $\theta \in (0,1)$ and $\|x(t)\| > r$ for any $x(t) \in \Omega_{\rho'} \setminus \mathcal{B}_r$, the Lyapunov function value decreases at $t+1$.

Now, consider the case that $x(t) \in \mathcal{B}_r$. In this case, $V(x(t+1))$ can be greater than $V(x(t))$ owing to the perturbations applied to the system. Nonetheless, if the perturbations are sufficiently small, $x(t+1)$ will stay in $\Omega_{\rho'}$. To show this, consider the difference between the Lyapunov function values at two successive states when $x(t) \in \mathcal{B}_r$. From Eq. B.1,

$$V(x(t+1)) - V(x(t)) \leq 2Rc_{f,1}\|f(t)\| + c_{f,2}\|f(t)\|^2 \tag{B.8}$$

for all $x(t) \in \mathcal{B}_r$. For any $\rho_s \in (0, \rho')$ such that $\mathcal{B}_r \subset \Omega_{\rho_s} := \{x \in \mathbb{R}^{n_x} | x^T \hat{P} x \leq \rho_s\} \subset \Omega_{\rho'}$ (existence of ρ_s satisfying the conditions follows from $\mathcal{B}_r \subset \mathcal{B}_{\hat{r}} \subseteq \Omega_{\rho'}$), there exist $\epsilon^*_{f,2} > 0$ where:

$$\epsilon^*_{f,2} := \frac{-Rc_{f,1} + \sqrt{R^2 c^2_{f,1} + (\rho' - \rho_s)c_{f,2}}}{c_{f,2}} \tag{B.9}$$

From Eq. B.9, $\rho' - \rho_s = c_{f,2}(\epsilon^*_{f,2})^2 + 2Rc_{f,1}\epsilon^*_{f,2}$. From Eq. B.8,

$$\begin{aligned} V(x(t+1)) &\leq V(x(t)) + 2Rc_{f,1}\|f(t)\| + c_{f,2}\|f(t)\|^2 \\ &< \rho_s + 2Rc_{f,1}\epsilon_f + c_{f,2}\epsilon_f^2 = \rho' \end{aligned}$$

for all $\|f(t)\| < \epsilon^*_{f,2}$ and $x(t) \in \mathcal{B}_r \subset \Omega_{\rho_s}$. Even if the value of the Lyapunov function increases, $x(t+1) \in \Omega_{\rho'}$ when $x(t) \in \mathcal{B}_r$ and $\mathbb{F} \subset \mathcal{B}_{\epsilon^*_{f,2}}$.

Let $\epsilon^*_f := \min\{\epsilon^*_v, \epsilon^*_{f,1}, \epsilon^*_{f,2}\}$. Applying the arguments above, if $x(t) \in \Omega_{\rho'}$ and $\mathbb{F} \subseteq \mathcal{B}_{\epsilon_f}$ for any $\epsilon_f \in [0, \epsilon^*_f)$, then $x(t+1) \in \Omega_{\rho'}$. Hence, $\Omega_{\rho'}$ is an RPI set provided the perturbations $f(t)$ are small. Additionally, the state constraint is satisfied for all $x(t) \in \Omega_{\rho'}$ because $\Omega_{\rho'} \subseteq \mathbb{X}$ and the input constraint is satisfied because $-K(x(t)+v(t)) \in \mathbb{U}$ for all $x(t) \in \Omega_{\rho'}$ and $v(t) \in \mathbb{V} \subseteq \mathcal{B}_{\epsilon^*_v}$. □