

3-27-2021


## Integrated Cyberattack Detection and Handling for Nonlinear Systems with Evolving Process Dynamics under Lyapunov-Based Economic Model Predictive Control

Keshav Kasturi Rangan  
*Wayne State University*, keshav@wayne.edu

Henrique Oyama  
*Wayne State University*, gq6229@wayne.edu

Helen Durand  
*Wayne State University*, helen.durand@wayne.edu

Follow this and additional works at: [https://digitalcommons.wayne.edu/cems\\_eng\\_frp](https://digitalcommons.wayne.edu/cems_eng_frp)

 Part of the [Information Security Commons](#), [Systems and Communications Commons](#), and the [Systems Engineering Commons](#)

---

### Recommended Citation

Rangan, K. K., Oyemi, H., Durand, H., 2021. Integrated cyberattack detection and handling for nonlinear systems with evolving process dynamics under Lyapunov-based economic model predictive control. *Chemical Engineering Research and Design*, 170, 147-179. doi: [10.1016/j.cherd.2021.03.024](https://doi.org/10.1016/j.cherd.2021.03.024)

This Article is brought to you for free and open access by the Chemical Engineering and Materials Science at DigitalCommons@WayneState. It has been accepted for inclusion in Chemical Engineering and Materials Science Faculty Research Publications by an authorized administrator of DigitalCommons@WayneState.

# Integrated Cyberattack Detection and Handling for Nonlinear Systems with Evolving Process Dynamics under Lyapunov-based Economic Model Predictive Control

Keshav Kasturi Rangan<sup>a</sup>, Henrique Oyama<sup>a</sup>, Helen Durand<sup>\*,a</sup>

<sup>a</sup>*Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI 48202.*

---

## Abstract

Safety-critical processes are becoming increasingly automated and connected. While automation can increase efficiency, it brings new challenges associated with guaranteeing safety in the presence of uncertainty especially in the presence of control system cyberattacks. One of the challenges for developing control strategies with guaranteed safety and cybersecurity properties under sufficient conditions is the development of appropriate detection strategies that work with control laws to prevent undetected attacks that have immediate closed-loop stability consequences. Achieving this, in the presence of uncertainty brought about by plant/model mismatch and process dynamics that can change with time, requires a fundamental understanding of the characteristics of attacks that can be detected with reasonable detection mechanisms and characterizing and verifying system safety properties when cyberattacks and changing system behavior cannot be distinguished. Motivated by this, this paper discusses three cyberattack detection strategies for nonlinear processes whose dynamics change with time when these processes are operated under an optimization-based control strategy known as Lyapunov-based economic model predictive control (LEMPC) until the closed-loop state either leaves a characterizable region of state-space or an attack detection threshold related to state estimates or state predictions is exceeded. Following this, the closed-loop state is maintained within a larger region of operation under an updated cyberattack detection strategy for a characterizable time period. A Taylor series-based model is used for making state predictions to allow theoretical guarantees to be explicitly tied to the numerical approximation of the model used

---

\*Corresponding author: Tel: +1 (313) 577-3475; E-mail: helen.durand@wayne.edu.

within the LEMPC. A process example illustrates the Taylor series-based model concept.

---

## 1. Introduction

With the move toward smart manufacturing Davis et al. (2015) and Industry 4.0 Lezzi et al. (2018), there are increasing efforts to update production facilities to include greater integration of physical processes and sensor measurements with computer and communication networks to implement and update current automated systems with more advanced capabilities. Advances in automation of various data-gathering/analysis and control tasks has also, however, raised concerns regarding cyberattacks on industrial systems Ren et al., including control systems Tuptuk and Hailes (2018).

The potential for cybersecurity vulnerabilities in control systems motivates the design of methodologies that are capable of detecting an attack in order to maintain safe operating conditions. Research efforts have been made to identify game-theoretic frameworks for assessing security risks associated with cyberphysical systems (CPS's) Amin et al. (2013). Additionally, other perspectives of vulnerability identification and assessment (e.g., Ani et al. (2017)), and detection mechanisms and countermeasures to deal with cyber threats (e.g., Hoehn and Zhang (2016)) have been topics of interest. Detection of attacks on a water distribution network was addressed in Amin et al. (2012) where delay-differential observers, designed based on an analytically approximate model of the process, were used. Stealthy attacks are considered particularly problematic and are defined as attacks which are not detected by a given detection mechanism; Teixeira et al. (2012) develops methods for changing a system's dynamics to allow attacks to be detected.

Many works providing a means of combating cyberattacks on industrial control systems have focused on linear systems. For example, Pasqualetti et al. (2013) focuses for a class of linear systems on mathematically characterizing attack detectability and identifiability and the properties of attack monitors. Other examples include handling of delay-based attacks on control signals using a model-based maximum likelihood technique to affirm or refute the likely presence of an attack on a linear system under an optimization-based controller as discussed in Barboni et al. (2018), and model-based attack detector design and detectability analysis for stochastic actuator and sensor

attacks on a linear system with stochastic disturbances in Li et al. (2015).

Because chemical processes are often described by nonlinear dynamic models, recent efforts in cybersecurity for chemical process systems have focused on methods for detection and handling of attacks on nonlinear systems. These have included cyberattack mitigation techniques, discussed in Wu et al. (2018) which integrate a neural network (NN)-based detection method and a Lyapunov-based model predictive controller for a certain class of nonlinear systems and Durand (2018), in which several strategies such as randomization of control law selection are analyzed to clarify their inability to prevent cyberattacks on control systems from causing problems, as well as Durand and Wegener (2020); Oyama and Durand (2020), in which strategies for combining detection and model predictive control (MPC Qin and Badgwell (2003); Ellis et al. (2014a); Rawlings et al. (2012)) for nonlinear systems are devised that ensure that the closed-loop state does not leave a safe operating region before a certain time period passes after an undetected attack. Another recent work which has integrated detection and control for nonlinear systems in the presence of cyberattacks is Liu et al. (2016), which focuses on a class of discrete-time nonlinear systems with random sensor measurement attacks and develops a filter with a bound on error covariance over time.

Cyberattacks pose a challenge for ensuring safety of an automated system. Safety assurance for autonomous systems has received a good deal of attention, with techniques for guaranteeing safety ranging from barrier functions, as described, for example, in Xu et al. (2015), to reachability analysis, as described in works such as Xiang and Johnson (2018). Conditions for safety in the presence of changing dynamics have been developed in Durand (2020b). Our recent work Durand (2020a); Oyama et al. has begun an exploration into the topic of how to handle cyberattacks when changes in the dynamics may also occur. As demonstrated in Oyama et al., there may be situations in which a cyberattack detection mechanism could flag dynamics changes as attacks because the dynamics change could lead data to no longer appear “expected.” A two-tier strategy for cyberattack detection and handling was proposed in Oyama et al. in which a cyberattack detection strategy could be tuned to recognize attacks before a change in the dynamics, but then not definitively call detection of abnormality via this first strategy an attack or a model change. Subsequently, model re-identification could occur as long as a secondary detection strategy that should only

detect attacks if the dynamics have not changed significantly does not detect an abnormality. However, the simulation-based study of this concept in Oyama et al. indicated that this method may be difficult to tune in a way that does not leave vulnerabilities without theoretical analysis of whether the resulting tuning is guaranteed to eliminate such vulnerabilities. The first step in moving toward trying to address this issue is to develop theoretical conditions. This work addresses this by providing theoretical conditions for preventing a model change or an undetected attack from driving the closed-loop state out of a safe operating region before a certain amount of time passes after the attack or model change using a two-tier detection strategy, focused on the time period before model re-identification. Furthermore, while Oyama et al. only provides simulation studies for one integrated detection and control strategy, this work discusses how all three detection and control strategies from Oyama and Durand (2020) could be updated to account for attacks as well as changes in dynamics.

Motivated by this, cyberattack detection strategies from Oyama and Durand (2020) are extended, in this work, to examine their capabilities for detecting cyberattacks and allowing the attacks to be handled with safety guarantees for some period of time after the attack when the process dynamics can change over time. These strategies are examined when the controller utilized is an LEMPC that incorporates a truncated Taylor series version of the solution to an empirical model to allow connections between numerical error in the controller and one of the detection strategies to be explicitly correlated with the guarantees that are made. The detection strategies are based on triggering mechanisms when the state of the system breaches certain thresholds implemented based on the model developed from empirical data. However, as changes in the underlying process dynamics are considered, we elucidate the challenges encountered in differentiating cyberattacks from a change in the process dynamics, and our main contribution is to provide detection and control techniques with sufficient conditions under which falsified state measurements cannot cause safety problems within a certain timeframe even when model changes and attacks may both occur. This paper is an extension of Rangan and Durand (2020); Durand (2020a). It incorporates the Taylor series analysis in Rangan and Durand (2020) into a cyberattack analysis framework with changing dynamics, and updates the model change framework in Durand (2020a) to include analysis of how

to handle cyberattacks simultaneously.

## 2. Preliminaries

### 2.1. Notation

The vector Euclidean norm is denoted by  $|\cdot|$ . A class  $\mathcal{K}$  function  $\alpha : [0, a) \rightarrow [0, \infty)$  has  $\alpha(0) = 0$  and is strictly increasing.  $x^T$  denotes the transpose of a vector  $x$ . The notation “  $/$  ” signifies set subtraction such that  $x \in A/B := \{x \in R^n : x \in A, x \notin B\}$ . A level set of a positive definite function  $V$  is denoted by  $\Omega_\rho := \{x \in R^n : V(x) \leq \rho\}$ .  $R_+$  signifies the set of non-negative real numbers. It is assumed that a measurement of the process state is available to a controller at synchronous time instants separated by sampling periods of length  $\Delta$  (i.e., a state measurement is available to a controller at every  $t_k := k\Delta$ ,  $k = 0, 1, \dots$ ).

A function  $\bar{f}_s : \mathbf{I} \rightarrow \mathbb{R}$ , where  $\mathbf{I} \subset \mathbb{R}$  is an open set, is said to be real analytic on  $\mathbf{I}$  if, for any  $c \in \mathbf{I}$  there is a neighborhood  $\mathbf{J}$  of  $c$  in which the function can be expressed as a convergent Taylor series Krantz and Parks (2002):

$$\bar{f}_s(t) = \bar{f}_s(c) + \sum_{n=1}^{\infty} \left( \bar{f}_{s,deriv}^n(c) \frac{(t-c)^n}{n!} \right) \quad \forall t \in \mathbf{J} \quad (1)$$

where  $\bar{f}_{s,deriv}^n(c) = \frac{d^n \bar{f}_s}{dt^n}(c)$  and  $t \in \mathbb{R}$ . In other words, a function is said to be analytic if in the neighborhood of some point  $c$  within the domain  $\mathbf{J}$  the Taylor series converges to the function  $\bar{f}_s$ .

A function  $\bar{g}_m : \bar{V} \rightarrow \mathbb{R}$ , where  $\bar{V} \subset \mathbb{R}^b$  is an open set, is said to be real analytic on  $\mathcal{C}$  if, for any  $\alpha \in \bar{V}$  the function  $\bar{g}_m$  may be represented by a convergent power series in some neighborhood  $\mathcal{C}$  of  $\alpha$  Krantz and Parks (2002):

$$\bar{g}_m(y) = \sum_{\mu \in \Lambda(b)} \beta_\mu (y - \alpha)^\mu, \quad \forall y \in \mathcal{C} \quad (2)$$

where  $\mu = (\mu_1, \mu_2, \dots, \mu_b) \in \Lambda(b)$  is a multi-index (i.e., a  $b$ -tuple of non-negative integers) such that the following holds with  $y = (y_1, y_2, \dots, y_b) \in \mathbb{R}^b$ :

$$|\mu| = \mu_1 + \mu_2 + \dots + \mu_b \quad (3)$$

$$y^\mu = y_1^{\mu_1} y_2^{\mu_2} \dots y_b^{\mu_b} \quad (4)$$

$$\frac{\partial^\mu}{\partial y^\mu} = \frac{\partial^{\mu_1}}{\partial y_1^{\mu_1}} \frac{\partial^{\mu_2}}{\partial y_2^{\mu_2}} \cdots \frac{\partial^{\mu_b}}{\partial y_b^{\mu_b}} \quad (5)$$

$$\mu! = \mu_1! \mu_2! \cdots \mu_b! \quad (6)$$

$$\beta_\mu = \frac{1}{\mu!} \frac{\partial^\mu}{\partial y^\mu} \bar{g}_m(\alpha) \quad (7)$$

$diag(\cdot)$  denotes a diagonal matrix with the arguments of this function as the diagonal elements.

## 2.2. Class of Systems

This work considers nonlinear process systems of the form:

$$\dot{x}_{a,i} = f_i(x_{a,i}(t), u(t), w_i(t)) \quad (8)$$

where  $f_i$  is a locally Lipschitz nonlinear vector function of its arguments,  $x_{a,i} \in X \subset R^n$  is the state vector,  $u \in U \subset R^m$  is the input vector with  $u = [u_1, \dots, u_m]^T$ , and  $w_i \in W_i \subset R^z$  is the disturbance vector ( $W_i := \{w_i \in R^z : |w_i| \leq \theta, \theta > 0\}$ , for  $i = 1, 2, \dots$ ). The  $i$ -th model is used for  $t \in [t_{s,i}, t_{s,i+1})$ , where  $x_{a,i}(t_{s,i+1}) = x_{a,i+1}(t_{s,i+1})$  and  $t_{s,1} = t_0$ . It is considered that the origin is the equilibrium of the system of Eq. 8 (i.e.,  $f_i(0, 0, 0) = 0$  and  $f_i(x_{a,i,s}, u_{i,s}, 0) = 0$  for  $i > 1$  such that the steady-state of the models after they update when  $w_i = 0$  is 0 at  $x_{a,i} = x_{a,i,s}$ ,  $u = u_{i,s}$ ). When  $w_i \equiv 0$ , the system of Eq. 8 is termed the “nominal” system. Measurements are assumed to be continuously available but provided to a controller at every  $t_k = k\Delta$ ,  $k = 0, 1, \dots$ . It is not required for  $t_{s,i}$ ,  $i = 1, 2, \dots$ , to be an integer multiple of  $t_k$ . The deviation variable  $\bar{x}_{a,i}$  is defined as  $x_{a,i} - x_{a,i,s} = \bar{x}_{a,i}$ ,  $\bar{u}_i = u - u_{i,s}$ , and  $\bar{f}_i$  is  $f_i$  rewritten to have its origin at  $\bar{x}_{a,i} = 0$  and  $\bar{u}_i = 0$  with  $w_i = 0$ .  $U_i$  is the set  $U$  in deviation variable form from  $u_{i,s}$ , and  $X_i$  is  $X$  in deviation variable form from  $x_{a,i,s}$ .

It is assumed that the system of Eq. 8 is stabilizable in the sense that there exists an infinitely differentiable positive definite Lyapunov function  $V_i : R^n \rightarrow R_+$ , as well as class  $\mathcal{K}$  functions  $\alpha_{j,i}(\cdot)$ ,  $j = 1, \dots, 4$ , and a controller  $h_i(\bar{x}_{a,i}) = [h_{i,1}(\bar{x}_{a,i}) \dots h_{i,m}(\bar{x}_{a,i})]^T$  that asymptotically stabilizes the origin of the nominal closed-loop system of Eq. 8 such that:

$$\alpha_{1,i}(|\bar{x}_{a,i}|) \leq V_i(\bar{x}_{a,i}) \leq \alpha_{2,i}(|\bar{x}_{a,i}|) \quad (9a)$$

$$\frac{\partial V_i(\bar{x}_{a,i})}{\partial \bar{x}_{a,i}} \bar{f}_i(\bar{x}_{a,i}, h_i(\bar{x}_{a,i}), 0) \leq -\alpha_{3,i}(|\bar{x}_{a,i}|) \quad (9b)$$

$$\left| \frac{\partial V_i(\bar{x}_{a,i})}{\partial \bar{x}_{a,i}} \right| \leq \alpha_{4,i}(|\bar{x}_{a,i}|) \quad (9c)$$

$$h_i(\bar{x}_{a,i}) \in U_i \quad (9d)$$

for all  $\bar{x}_{a,i} \in D_i \subseteq R^n$  and  $i = 1, 2, \dots$ , where  $D_i$  is an open neighborhood of the origin of  $\bar{f}_i$ .  $\Omega_{\rho_i} \subset D_i$  denotes a level set of  $V_i$  and is referred to as the stability region of the system of Eq. 8 under the control action  $h_i(\bar{x}_{a,i})$ . It is assumed to be chosen such that it is contained within  $X_i$ . When  $u(t)$  is fixed/constant for the nominal ( $w_i(t) \equiv 0$ ) system of Eq. 8, the resulting function is considered to be analytic in  $\bar{x}_{a,i}$  on  $D_i$  and to have a solution  $\bar{x}_{a,i}(t)$  that is analytic in  $t$ .

From the Lipschitz continuity of  $\bar{f}_i$  and the boundedness of  $\bar{x}_i$ ,  $\bar{u}_i$ , and  $w_i$ , there exist positive constants  $M_i$ ,  $L_{x,i}$ ,  $L_{w,i}$ ,  $L'_{x,i}$ , and  $L'_{w,i}$  such that:

$$|\bar{f}_i(\bar{x}_i, \bar{u}_i, w_i) - \bar{f}_i(\bar{x}'_i, \bar{u}_i, 0)| \leq L_{x,i}|\bar{x}_i - \bar{x}'_i| + L_{w,i}|w_i| \quad (10a)$$

$$\left| \frac{\partial V_i(\bar{x}_i)}{\partial \bar{x}_i} \bar{f}_i(\bar{x}_i, \bar{u}_i, w_i) - \frac{\partial V_i(\bar{x}'_i)}{\partial \bar{x}'_i} \bar{f}_i(\bar{x}'_i, \bar{u}, 0) \right| \leq L'_{x,i}|\bar{x}_i - \bar{x}'_i| + L'_{w,i}|w_i| + L'_{u,i}|\bar{u}_i - \bar{u}| \quad (10b)$$

$$|\bar{f}_i(\bar{x}_i, \bar{u}, w_i)| \leq M_i \quad (10c)$$

$\forall \bar{x}_i, \bar{x}'_i \in \Omega_{\rho_i}$ ,  $\bar{u}_i, \bar{u} \in U_i$ , and  $w_i \in W_i$ .

Finally, the Lyapunov-based controller is assumed to be locally Lipschitz continuous such that the following inequalities hold:

$$|h_{i,j}(\bar{x}_i) - h_{i,j}(\bar{x}'_i)| \leq L_{h,i}|\bar{x}_i - \bar{x}'_i| \quad (11)$$

for a positive constant  $L_{h,i}$  for all  $\bar{x}_i, \bar{x}'_i \in \Omega_{\rho_i}$ ,  $L_{h,i} > 0$  and  $i = 1, 2, \dots$ , with  $j = 1, 2, \dots, m$ . We assume that there exists  $M_{i,N_1} > 0$  such that for any  $\bar{u} \in U_i$  and  $\bar{x} \in \Omega_{\rho_i}$ :

$$|\bar{f}_i^{N_1+1}(\bar{x}, \bar{u}, w_i)| \leq M_{i,N_1} \quad (12)$$

for all  $N_1 = 0, 1, 2, \dots$  and that  $M_{i,0} > 0$  bounds  $|\bar{f}_i(\bar{x}, \bar{u}, w_i)|$  for all  $|w_i| \leq \theta$ , where  $\bar{f}_i^n = \frac{d^n \bar{x}_{a,i}}{dt^n}$ .

### 2.3. Empirical Model

This work considers that the model of Eq. 8 is not available, and instead an empirical model with the following form may be available:

$$\dot{x}_{b,q}(t) = f_{NL,q}(x_{b,q}(t), u(t)) \quad (13)$$



where  $f_{NL,q}$  is a locally Lipschitz (and analytic in  $x_{b,q}$  for fixed  $u$  with a solution  $x_{b,q}(t)$  assumed to be analytic in  $t$ ) nonlinear vector function in  $x_{b,q} \in \mathbb{R}^n$  and in the input  $u \in \mathbb{R}^m$ . While  $f_{NL,1}(0,0) = 0$ , the steady-state of the updated models is at  $x_{b,q} = x_{b,q,s} = 0$  and  $u = u_{q,s}$  so that  $f_{NL,q}(x_{b,q,s}, u_{q,s}) = 0$  for  $q > 1$ . The index  $q = 1, 2, \dots$ , reflects the index for the empirical model used at a given time, which is not necessarily the same as  $i$  in Eq. 8 because the empirical model may not update at the same time as the process dynamics change. Eq. 13 is updated at the time  $t_{s,NL,q}$  and  $x_{b,q}(t_{s,NL,q}) = x_{b,q+1}(t_{s,NL,q})$ . The deviation variable  $\bar{x}_{b,q}$  is defined as  $x_{b,q} - x_{b,q,s} = \bar{x}_{b,q}$ ,  $\bar{u}_q = u - u_{q,s}$ , and  $\bar{f}_{NL,q}$  is  $f_{NL,q}$  rewritten to have its origin at  $\bar{x}_{b,q} = 0$ ,  $\bar{u}_q = 0$ , giving the following:

$$\dot{\bar{x}}_{b,q}(t) = \bar{f}_{NL,q}(\bar{x}_{b,q}(t), \bar{u}_q(t)) \quad (14)$$

$U_q$  is the set  $U$  in deviation variable form from  $u_{q,s}$ , and  $X_q$  is the set  $X$  in deviation variable form from  $x_{b,q,s}$ . We consider that there exist locally Lipschitz explicit stabilizing controllers  $h_{NL,q}(\bar{x}_{b,q})$  that can render the origin of the empirical models in Eq. 13 asymptotically stable in the sense that:

$$\hat{\alpha}_{1,q}(|\bar{x}_{b,q}|) \leq \hat{V}_q(\bar{x}_{b,q}) \leq \hat{\alpha}_{2,q}(|\bar{x}_{b,q}|) \quad (15a)$$

$$\frac{\partial \hat{V}_q(\bar{x}_{b,q})}{\partial \bar{x}_{b,q}} \bar{f}_{NL,q}(\bar{x}_{b,q}, h_{NL,q}(\bar{x}_{b,q})) \leq -\hat{\alpha}_{3,q}(|\bar{x}_{b,q}|) \quad (15b)$$

$$\left| \frac{\partial \hat{V}_q(\bar{x}_{b,q})}{\partial \bar{x}_{b,q}} \right| \leq \hat{\alpha}_{4,q}(|\bar{x}_{b,q}|) \quad (15c)$$

$$h_{NL,q}(\bar{x}_{b,q}) \in U_q \quad (15d)$$

for all  $\bar{x}_{b,q} \in D_{NL,q}$ , where  $D_{NL,q}$  is a neighborhood of the origin of  $\bar{f}_{NL,q}$  contained in  $X$ . The function  $\hat{V}_q : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is an infinitely differentiable Lyapunov function and is assumed to be the same as  $V_i$  for the underlying dynamics at the time  $\hat{V}_q$  is used (i.e., the  $V_i$  and  $\hat{V}_q$  are assumed to be the same at all times). The functions  $\hat{\alpha}_{\bar{i},q}$ ,  $\bar{i} = 1, 2, 3, 4$ , are class  $\mathcal{K}$  functions with  $q = 1, 2, \dots$ . The set  $\Omega_{\hat{\rho}_q} \subset D_{NL,q}$  is defined to be the stability region of the system of Eq. 13 under  $h_{NL,q}$ , and  $\Omega_{\hat{\rho}_{safe,q}}$  is a superset of  $\Omega_{\hat{\rho}_q}$  contained in both  $D_{NL,q}$  and  $X$ . Lipschitz continuity of  $f_{NL,q}$  and sufficient smoothness of  $\hat{V}_q$  imply that there exist  $M_{L,q} > 0$  and  $L_{L,q} > 0$  such that:

$$|\bar{f}_{NL,q}(x, u)| \leq M_{L,q} \quad (16a)$$

$$\left| \frac{\partial \hat{V}_q(x_1)}{\partial x} \bar{f}_{NL,q}(x_1, u) - \frac{\partial \hat{V}_q(x_2)}{\partial x} \bar{f}_{NL,q}(x_2, u) \right| \leq L_{L,q} |x_1 - x_2| \quad (16b)$$

$\forall x, x_1, x_2 \in \Omega_{\hat{\rho}_q}, u \in U_q$ , and  $q = 1, 2, \dots$

We assume that  $x_{b,q,s}$  and  $x_{a,i,s}$  do not change over time and that  $x_{b,q,s} = x_{a,i,s}$  even after the empirical and process models change, though the steady-state inputs required to maintain the closed-loop state at these conditions change as the models update. It is assumed that for any  $i$ -th process model which describes that dynamics when the  $q$ -th empirical model is used,  $\Omega_{\hat{\rho}_{safe,q}} \in \Omega_{\rho_i}$ .

We consider that:

$$|\bar{f}_{NL,q}^n(\bar{x}_{b,q}, \bar{u}_q) - \bar{f}_{NL,q}^n(\bar{x}'_{b,q}, \bar{u}'_q)| \leq L_{x,n,q} |\bar{x}_{b,q} - \bar{x}'_{b,q}| \quad (17a)$$

$$\left| \frac{\partial \hat{V}_q(\bar{x}_{b,q})}{\partial \bar{x}_{b,q}} \bar{f}_{NL,q}^n(\bar{x}_{b,q}, \bar{u}_q) - \frac{\partial \hat{V}_q(\bar{x}'_{b,q})}{\partial \bar{x}_{b,q}} \bar{f}_{NL,q}^n(\bar{x}'_{b,q}, \bar{u}'_q) \right| \leq L'_{x,n,q} |\bar{x}_{b,q} - \bar{x}'_{b,q}| + L'_{u,n,q} |\bar{u}_q - \bar{u}'_q| \quad (17b)$$

for all  $\bar{x}_{b,q}, \bar{x}'_{b,q} \in \Omega_{\hat{\rho}_q}$  and  $\bar{u}_q, \bar{u}'_q \in U_q$ , where  $L_{x,n,q}$ ,  $L'_{x,n,q}$ ,  $L_{u,n,q}$ , and  $L'_{u,n,q}$  are positive constants.

We consider that  $h_{NL,q}$  satisfies:

$$|h_{NL,q}(x) - h_{NL,q}(x')| \leq L_{h,NL} |x - x'| \quad (18)$$

for all  $x, x' \in \Omega_{\hat{\rho}_{safe,q}}$  with  $L_{h,NL} > 0$ .

#### 2.4. Observability assumption

We assume that there are  $M$  sets of measurements  $y_p \in R^{q_p}$ ,  $p = 1, \dots, M$ , available continuously, as follows:

$$y_p(t) = k_{p,i}(\bar{x}_{a,i}(t)) + v_p(t) \quad (19)$$

where  $y_p$  represents the measurement vector in deviation variable form,  $k_{p,i}$  is a vector-valued function that enables  $y_p$  to be written in deviation form from the  $i$ -th steady-state, and  $v_p$  represents bounded measurement noise (i.e.,  $v_p \in V_p := \{v_p \in R^{q_p} : |v_p| \leq \theta_{v,p}, \theta_{v,p} > 0\}$ ). We consider that a deterministic observer exists for each of the  $M$  sets of measurements when the  $q$ -th empirical model is used with the form:

$$\dot{z}_{q,p} = F_{p,q}(\epsilon_{pq}, z_{q,p}, y_{p,q}) \quad (20)$$

where  $z_{q,p}$  is the state estimate from the  $p$ -th observer,  $p = 1, \dots, M$ ,  $F_{p,q}$  is a vector-valued function, and  $\epsilon_{pq} > 0$ . When a controller  $h_{NL,q}(z_{q,p})$  with Eq. 20 is used to control the closed-loop system of

Eq. 8 and no change in the underlying dynamics occurs, we make the following assumptions which are similar to Ellis et al. (2014b); Lao et al. (2015), where  $M_{err,i,q} > 0$  is defined by:

$$|\bar{f}_i(x, \bar{u}_i, 0) - \bar{f}_{NL,q}(x, \bar{u}_q)| \leq M_{err,i,q} \quad (21)$$

for all  $x \in \Omega_{\hat{\rho}_{safe,q}}$  and all  $\bar{u}_i = \bar{u}_q + u_{q,s} - u_{i,s}$  in the input bounds.

**Assumption 1.** *There exist positive constants  $M_{err,i,q}^*$ ,  $\theta^*$ ,  $\theta_{v,p}^*$ , such that for each pair of  $\{M_{err,i,q}, \theta, \theta_{v,p}\}$  with  $M_{err,i,q} \leq M_{err,i,q}^*$ ,  $\theta \leq \theta^*$ , and  $\theta_{v,p} \leq \theta_{v,p}^*$ , there exist  $0 < \hat{\rho}_{1,p,q} < \hat{\rho}_q$ ,  $e_{m0pq} > 0$  and  $\epsilon_{Lpq}^* > 0$ ,  $\epsilon_{Upq}^* > 0$  such that if  $x(0) \in \Omega_{\hat{\rho}_{1,p,q}}$ ,  $|z_{q,p}(0) - \bar{x}_{a,i}(0)| \leq e_{m0pq}$  and  $\epsilon_{pq} \in (\epsilon_{Lpq}^*, \epsilon_{Upq}^*)$ , the trajectories of the closed-loop system are bounded in  $\Omega_{\hat{\rho}_q}$ ,  $\forall t \geq 0$  before a change in the process dynamics.*

**Assumption 2.** *There exists  $e_{pq}^* > 0$  such that for each  $e_{pq} \geq e_{pq}^*$ , there exist  $t_{bpq}(\epsilon_{pq})$  such that  $|z_{q,p}(t) - \bar{x}_{a,i}(t)| \leq e_{pq}$ ,  $\forall t \geq t_{bpq}(\epsilon_{pq})$  before a change in the process dynamics.*

**Remark 1.** *We assume that multiple observers exist that are capable of making state estimates with a bound on their accuracy.*

## 2.5. Taylor Series Error Bounds

There exists an upper bound on the error when truncating the Taylor series representation of the function  $\bar{f}_i(t)$  in Eq. 8 to  $N_1 + 1$  terms that is captured in the following proposition.

**Proposition 1.** *Stewart (2003) The error  $E_i(t)$  from truncating the Taylor series representation of  $\bar{f}_i(t)$  in Eq. 1 to  $N_1 + 1$  terms is given by:*

$$E_i(t) = \bar{f}_s(t) - \bar{f}_s(c) - \sum_{n=1}^{N_1} \bar{f}_{s,deriv}^n(c) \frac{(t-c)^n}{n!} \quad (22)$$

If  $|\bar{f}_{s,deriv}^{N_1+1}| \leq \bar{M}_{N_1,s}$  for  $|t-c| \leq d$ , then for  $|t-c| \leq d$ :

$$|E_i(t)| \leq \frac{\bar{M}_{N_1,s} |t-c|^{N_1+1}}{(N_1+1)!} \quad (23)$$

## 2.6. Lyapunov-based Economic Model Predictive Control with Empirical Models

In this work, we use the optimization-based control design known as LEMPC as described in Heidarinejad et al. (2012), to control the process described in Eq. 8. This formulation can be developed using the empirical model of Eq. 13 Alanqar et al. (2015b,a); Giuliani and Durand (2018) and can be represented in the following form:

$$\min_{\bar{u}_q(t) \in S(\Delta)} \int_{t_k}^{t_{k+N}} [L_e(\bar{x}_{b,q}(\tau), \bar{u}_q(\tau))] d\tau \quad (24a)$$

$$\text{s.t.} \quad \dot{\bar{x}}_{b,q} = \bar{f}_{NL,q}(\bar{x}_{b,q}(t), \bar{u}_q(t)) \quad (24b)$$

$$\bar{x}_{b,q}(t_k) = x(t_k) \quad (24c)$$

$$\bar{x}_{b,q}(t) \in X_q, \forall t \in [t_k, t_{k+N}) \quad (24d)$$

$$\bar{u}_q(t) \in U_q, \forall t \in [t_k, t_{k+N}) \quad (24e)$$

$$\hat{V}_q(\bar{x}_{b,q}(t)) \leq \hat{\rho}'_{e,q}, \quad \forall t \in [t_k, t_{k+N}) \quad \text{if } x(t_k) \in \Omega_{\hat{\rho}'_{e,q}} \quad (24f)$$

$$\frac{\partial \hat{V}_q(x(t_k))}{\partial x}(\bar{f}_{NL,q}(x(t_k), \bar{u}_q(t_k))) \leq \frac{\partial \hat{V}_q(x(t_k))}{\partial x}(\bar{f}_{NL,q}(x(t_k), h_{NL,q}(x(t_k))))$$

$$\text{if } x(t_k) \notin \Omega_{\hat{\rho}'_{e,q}} \quad (24g)$$

where  $L_e(\cdot, \cdot)$  represents a general scalar-valued stage cost of the LEMPC that is minimized in Eq. 24.  $\bar{u}_q$  is a piecewise-constant input trajectory with period  $\Delta$ , which is indicated by the notation  $\bar{u}_q(t) \in S(\Delta)$ . The prediction horizon is denoted by  $N$ . Eq. 24b represents the nominal process model, with predicted state  $\bar{x}_{b,q}$  for the  $q$ -th model.  $x(t_k)$  in Eq. 24c sets the predicted state of the empirical model at  $t_k$  equal to the measured state. Eqs. 24e and 24d are the input and state constraints, respectively. The set  $\Omega_{\hat{\rho}'_{e,q}}$  is selected as a subset of  $\Omega_{\hat{\rho}_q}$  that causes the closed-loop state to be maintained within  $\Omega_{\hat{\rho}_q}$  over time when the system of Eq. 8 is operated under the controller of Eq. 24. The constraints of the LEMPC guarantee recursive feasibility. We assume that there exists  $M_{q,k,N_1} > 0$  such that for all  $\bar{u}_q(t_k) \in U_q$  and  $x(t_k) \in \Omega_{\hat{\rho}_{safe,q}}$ :

$$|\bar{f}_{NL,q}^{N_1+1}(x(t_k), \bar{u}_q(t_k))| \leq M_{q,k,N_1} \quad (25)$$

for all  $N_1 = 0, 1, 2, \dots$ , where  $\bar{f}_{NL,q}^n = \frac{d^n \bar{x}_{b,q}}{dt^n}$ .

### 3. Run-Time Cyberattack Resilience Verification

Theoretical guarantees regarding the ability to maintain the closed-loop state in a known operating region for a certain amount of time after a cyberattack under certain detection strategies have been previously developed for systems of the form in Eq. 8 when the underlying dynamics do not change over time in Oyama and Durand (2020). New challenges arise in using these previously proposed cyberattack detection methods, to be further discussed below, when the process dynamics can change over time. This section will focus on developing cyberattack detection strategies that can guarantee that when coupled with certain control strategies, the closed-loop state does not leave a predefined region of operation for a defined amount of time after an undetected attack even when the process dynamics change with time.

#### 3.1. Run-time verification in the absence of attacks

In Durand (2020b) and Durand (2020a), a method for guaranteeing that the closed-loop state of the system of Eq. 8 under the LEMPC of Eq. 24 does not exit a known operating region for a defined amount of time after the underlying process dynamics change, but in the absence of an attack, was developed. In this strategy, which we consider to be an (admittedly conservative and potentially difficult to practically impose, but nonetheless theoretically valuable) method for verifying safety at run-time, a region  $\Omega_{\hat{\rho}_{safe,q}}$  (a superset of  $\Omega_{\hat{\rho}_q}$ ) which the closed-loop state should not leave after a change in the underlying dynamics is defined, and  $\Omega_{\hat{\rho}_q}$  is defined such that the closed-loop state should not leave  $\Omega_{\hat{\rho}_q}$  before the dynamics change. If the closed-loop state leaves  $\Omega_{\hat{\rho}_q}$ , this can signal a change in the underlying dynamics. As a result, if the closed-loop state leaves  $\Omega_{\hat{\rho}_q}$  (a sampling time at which this occurs is denoted by  $t_{d,q}$ ),  $h_{NL,q}$  is used as the controller for ease of use until a model re-identification can be performed and used to update the model incorporated within the LEMPC at a sampling time  $t_{ID,q}$ . Once the model is re-identified, the parameters/functions utilized in the design of the LEMPC and the Lyapunov-based controller are updated (specifically, the Lyapunov function  $\hat{V}_{q+1}$ , the Lyapunov-based controller  $h_{NL,q+1}$ , and the stability region  $\Omega_{\hat{\rho}_{q+1}}$  are used for developing the controller for the process under the updated process model). A worst-case bound is placed on the number of sampling periods  $t_{h,q}$  available between  $t_{d,q}$  and  $t_{ID,q}$  before model re-

identification must be performed and the LEMPC and Lyapunov-based controller updated before the closed-loop state will leave  $\Omega_{\hat{\rho}_{safe,q}}$ .

Combining the above strategy for operating an LEMPC with safety guarantees in the presence of changing dynamic models with guarantees on cyberattack detection from Oyama and Durand (2020) raises new issues that will be discussed and handled via updated implementation strategies in the subsequent sections. Furthermore, in these sections, we will explore a modified version of the LEMPC of Eq. 24 as follows:

$$\min_{\bar{u}_q(t) \in S(\Delta)} \int_{t_k}^{t_{k+N}} [L_e(\tilde{x}_{b,q}(\tau), \bar{u}_q(\tau))] d\tau \quad (26a)$$

$$\begin{aligned} \text{s.t. } \tilde{x}_{b,q}(t) &= \tilde{x}_{b,q}(t_j) + \sum_{n=1}^{N_1} \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_j), \bar{u}_q(t_j)) \frac{(t-t_j)^n}{n!} \\ &\forall t \in [t_j, t_{j+1}), j = k, \dots, k+N-1 \end{aligned} \quad (26b)$$

$$\tilde{x}_{b,q}(t_k) = x(t_k) \quad (26c)$$

$$\tilde{x}_{b,q}(t) \in X_q, \forall t \in [t_k, t_{k+N}) \quad (26d)$$

$$\bar{u}_q(t) \in U_q, \forall t \in [t_k, t_{k+N}) \quad (26e)$$

$$\hat{V}_q(\tilde{x}_{b,q}(t)) \leq \hat{\rho}_{e,q}, \forall t \in [t_k, t_{k+N}), \text{ if } \tilde{x}_{b,q}(t_k) \in \Omega_{\hat{\rho}_{e,q}} \quad (26f)$$

$$\begin{aligned} \frac{\partial \hat{V}_q(x(t_k))}{\partial x} \bar{f}_{NL,q}(x(t_k), \bar{u}_q(t_k)) &\leq \frac{\partial \hat{V}_q(x(t_k))}{\partial x} \bar{f}_{NL,q}(x(t_k), h_{NL,q}(x(t_k))) \\ &\text{if } \bar{x}_{b,q}(t_k) \notin \Omega_{\hat{\rho}_{e,q}} \end{aligned} \quad (26g)$$

The formulation in Eq. 26, introduced in Rangan and Durand (2020), is similar to that in Eq. 24, but it uses a different upper bound on  $\hat{V}_q$  in Eqs. 26f-26g, and it uses a truncated Taylor series approximation of the solution of Eq. 13 to make the state predictions  $\tilde{x}_{b,q}$  that appear in the objective function and constraints. This formulation assumes that the model of Eq. 13 is known. This allows the constraint of Eq. 26g to be written in terms of  $\bar{f}_{NL,q}$  explicitly, as in Eq. 24g. The closeness of this formulation to that in Eq. 24 allows the results in this work to be applicable to the LEMPC of Eq. 24 as well (i.e., if  $N_1 = \infty$ , they are the same). However, we select the formulation in Eq. 26 in this work because it allows an explicit relationship to be developed essentially between how a numerical method (in this case, a truncated Taylor series) impacts the guarantees to be developed compared to noise and disturbances. This allows a clear relationship, in cyberattack

detection strategies to be presented, to be developed between the impacts of noise, plant/model mismatch, and numerical error in allowing the cyberattack-resilience guarantees to be developed.

For this formulation, as noted previously, we assume that Eqs. 8 and Eq. 14, for fixed inputs in the input bounds, are analytic in the state and have a solution that is analytic in  $t$  such that the model of Eq. 14 throughout the prediction horizon of  $N$  sampling periods can be written as a set of  $N$  equations with fixed values of the inputs. We further consider that  $\Delta$  is sufficiently small (to be denoted by  $\Delta \leq \Delta_{ub,q}$ ) such that at any  $t_j$  at which a model with a fixed input begins to represent Eq. 14, there is a neighborhood of  $t_j$  including  $t_{j+1}$ , where  $j = k, \dots, k + N - 1$ , on which the solution of that model can be represented as a convergent Taylor series. These assumptions lead to the form of the approximate solution of Eq. 14 represented by Eq. 26b.

### 3.2. Run-Time Cybersecurity Verification with Changing Process Dynamics

In Oyama and Durand (2020), guarantees that closed-loop stability can be maintained after an attack for at least some period of time were developed for nonlinear systems, of the form in Eq. 8, in the scenario where the dynamics of the system do not change with time. These guarantees, however, are based on the availability of detection strategies rooted in stability guarantees under Lyapunov-based EMPC with a constraint in the form like that in Eq. 24g activated (i.e., that the time derivative of the Lyapunov function decreases when the constraint is activated and the closed-loop state is outside of a neighborhood of the origin; this strategy is referred to as Detection Strategy 1), on state predictions being sufficiently accurate (Detection Strategy 2), or on state estimates being sufficiently accurate (Detection Strategy 3). When the underlying dynamics change, it would not be expected that state predictions and state estimates would necessarily continue to be accurate, and in addition, it is not necessarily true that the Lyapunov function would decrease for an LEMPC with a constraint of the form of Eq. 24g. Therefore, from the perspective of these detection strategies, cyberattacks and changes in the underlying dynamics may be difficult to distinguish.

This section addresses this by presenting modified versions of Detection Strategies 2 and 3 from Oyama and Durand (2020) designed to allow the closed-loop state to remain within  $\Omega_{\hat{\rho}_q, safe}$  for a characterizable amount of time after it is detected to have left  $\Omega_{\hat{\rho}_q}$ . This modification holds even when the reason that the closed-loop state has left  $\Omega_{\hat{\rho}_q}$  cannot be definitively characterized

as being the result of a cyberattack versus a change in the underlying process dynamics. We also briefly discuss Detection Strategy 1 from Oyama and Durand (2020) for the case where changes in the underlying dynamics cannot be differentiated from cyberattacks. The results from Oyama and Durand (2020) were obtained without explicitly accounting for numerical error when solving a process model as suggested in, for example, Eq. 26b, but can be extended to such a case. We highlight that the theoretical results in the subsequent sections consider sufficiently small bounded measurement noise and plant/model mismatch.

### 3.2.1. Detection Strategy 1: Randomized LEMPC Changes to Probe for Cyberattacks

This detection strategy in Oyama and Durand (2020) takes advantage of the closed-loop stability properties of LEMPC to probe for cyberattacks. Specifically, when the constraint of the form of Eq. 24g is activated in the LEMPC under sufficient conditions, the Lyapunov function should decrease over the subsequent sampling period as long as the state measurement at the beginning of the sampling period is not within a neighborhood of the origin. Detection Strategy 1 takes advantage of this by operating a process under an LEMPC designed based on the original ( $j = 1$ ) steady-state for the majority of the operation, but at random times develops alternative steady-states (i.e.,  $j$ -th steady-states with  $j > 1$ ) with stability regions containing the state measurement at  $t_k$ . At these random times, it switches from using the  $j = 1$  LEMPC (or 1-LEMPC) to using that LEMPC designed around the new steady-state (i.e., the  $j$ -LEMPC,  $j > 1$ , has the process model, Lyapunov function, and Lyapunov-based controller adjusted to be with respect to the  $j$ -th steady-state), but with a constraint of the form of that in Eq. 24g activated regardless of the position of  $x(t_k)$  within the new stability region (i.e., a constraint similar to that in Eq. 24f is not activated). After a sampling period, the  $j$ -LEMPC formulation is switched back to the 1-LEMPC formulation. In the absence of a change in the process dynamics, the value of the Lyapunov function will decrease over the sampling period following the activation of the  $j$ -LEMPC, so that a lack of decrease in the Lyapunov function in the state measurement data could therefore signal a potential cyberattack on the state measurements.

When the process dynamics may change, however, there become two possible reasons that the value of the Lyapunov function may not decrease over the sampling period following the activation



of the  $j$ -LEMPC: 1) the underlying dynamics of the process of Eq. 8 have changed (i.e., Eq. 14 is no longer a sufficiently accurate approximation of the actual process dynamics to enable the  $j$ -LEMPC to decrease  $\hat{V}_q$ ) or 2) a cyberattack on the sensor measurements has occurred. These two cases might not be distinguishable using this detection mechanism alone (i.e., when the value of the Lyapunov function fails to decrease over a sampling period following activation of the  $j$ -LEMPC, this detection method alone might not reveal whether the reason is due to a cyberattack on the sensor measurements or due to a change in the underlying process dynamics).

### 3.2.2. Detection Strategy 2: Cyberattack-Mitigating State Feedback LEMPC

The second detection strategy from Oyama and Durand (2020) to be explored uses the difference between state measurements and state predictions to flag cyberattacks on the process sensors. Specifically, a threshold  $\nu_q$  is selected *a priori* to upper bound the error between state predictions at  $t_k$  made from a measurement at  $t_{k-1}$  and state measurements (denoted by  $x(t_k)$ ). Though Oyama and Durand (2020) does not use the empirical models or the approximate Taylor series-based solution of Eq. 24b, state predictions at  $t_k$  made from the measurements at  $t_{k-1}$ , in this section, will be denoted by  $\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1})$  to introduce notation that will be subsequently used when the LEMPC of Eq. 26 is used and will capture the intent of the work in Oyama and Durand (2020). If  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_q$  at a sampling time, an attack is detected. If  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_q$  at a sampling time, the LEMPC of Eq. 24 is used to control the process for the subsequent sampling period. If the parameters of the control law (e.g.,  $\Delta$  and  $\hat{\rho}_{e,q}$ ) are selected in a sufficiently conservative fashion, then there is at least a sampling period after an undetected attack occurs during which the closed-loop state does not leave  $\Omega_{\hat{\rho}_q}$  when the process dynamics do not change over time.

In this strategy, when the process dynamics do not change over time, the value of  $\nu_q$  is designed to ensure that there is no way that disturbances (or plant/model mismatch caused by the use of the empirical model) or noise could cause  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)|$  to be greater than  $\nu_q$ , making the detection method capable of flagging attacks only. However, when the process dynamics are allowed to change over time, there become two reasons that  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)|$  could exceed  $\nu_q$ : 1) the dynamics of the process have changed such that Eq. 14 is no longer adequate for making

accurate state predictions or 2) a cyberattack has occurred on the process sensors. It may not be possible to differentiate between these two cases using this detection strategy as it depends on state predictions. Furthermore, when the state measurement leaves  $\Omega_{\hat{\rho}_q}$ , it may not be possible to know whether this has occurred due to an attack on the sensors or due to a change in the process dynamics. This necessitates the need for an updated implementation strategy and value of  $\nu_q$  for guaranteeing that the closed-loop state remains within  $\Omega_{\hat{\rho}_{safe,q}}$  for a defined amount of time after the closed-loop state leaves  $\Omega_{\hat{\rho}_q}$  or after  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_q$  when it is not known whether the cause of the mismatch between the state prediction and measurement arises from an attack or a change in the dynamics.

To achieve this, we will utilize two stages of monitoring for cyberattacks and model changes. The first stage will utilize a detection strategy based on an initial upper bound on  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)|$ , denoted by  $\nu_{s,q}$ . This bound will be designed such that, if there were no model changes,  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_{s,q}$  would signify a cyberattack with certainty according to the method in Oyama and Durand (2020). However, when model changes are allowed, it is uncertain whether  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_{s,q}$  signifies a cyberattack; therefore, we will develop a second bound  $\nu_{l,q}$  ( $\nu_{l,q} \geq \nu_{s,q}$ ) where, if  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{l,q}$  after a model change but no cyberattack is detected via this updated detection mechanism, the closed-loop state should not leave  $\Omega_{\hat{\rho}_{safe,q}}$  within a sampling period after the attack occurs if it is not detected. Initially, the process is operated within  $\Omega_{\hat{\rho}_q}$ , and using the cyberattack detection mechanism based on  $\nu_{s,q}$ . Either a measurement outside of  $\Omega_{\hat{\rho}_q}$  or a measurement which causes  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_{s,q}$  triggers activation of the second cyberattack detection method based on  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{l,q}$ . Because either of those conditions which trigger the activation of the new cyberattack detection mechanism could signify that the underlying process dynamics changed, a new model will be re-identified within  $t_{h,q}$  sampling periods after either of the detection conditions is triggered if no attack is detected within  $t_{h,q}$  sampling times after  $t_{d,q}$ . However, if an attack is undetected, an auxiliary detection mechanism may be needed to prevent the attack from causing the closed-loop state to leave  $\Omega_{\hat{\rho}_{safe,q}}$  after a sampling period following the attack.

The implementation strategy just described for this detection method is as follows:

1. At  $t_0$ , the  $i = 1$  model (Eq. 8) describes the dynamics of the process. The  $q = 1$  empirical model (Eq. 14) is used to design the LEMPC of Eq. 26. An index  $i_{hx}$  is set to 0. An index  $\zeta$  is set to 0. Go to Step 2.
2. Check if  $x(t_k) \notin \Omega_{\hat{\rho}_q}$ , but  $\zeta = 0$ . If so, set  $\zeta = 1$  and  $t_{d,q} = t_k$ . Go to Step 3.
3. Check the value of  $e_{dif} = |\tilde{x}_{b,q}(t_k|t_{k-1}) - x(t_k)|$ . If  $\zeta = 0$  and  $e_{dif} > v_{s,q}$ , set  $\zeta = 1$  and  $t_{d,q} = t_k$  and check if  $e_{dif} > v_{l,q}$ . If  $e_{dif} > v_{l,q}$ , consider that a cyberattack on the sensors is occurring and initiate a backup strategy (e.g., redundant sensors or an emergency shut-down mode). If  $i_{hx} = 1$ , go to Step 3a. Else, if  $\zeta = 1$ , go to Step 3b, or if  $\zeta = 0$ , go to Step 3c.
  - (a) If  $x(t_k) \in \Omega_{\hat{\rho}_{q+1}}$ , operate the process under the LEMPC of Eq. 24 with  $q \leftarrow q + 1$ , set  $i_{hx} = 0$  and  $\zeta = 0$ . Else, apply  $h_{NL,q+1}(x(t_k))$  to the process. Go to Step 2.  $t_k \leftarrow t_{k+1}$ .
  - (b) If  $(t_{k+1} - t_{d,q}) < t_{h,q}$ , gather on-line data to develop an improved process model as well as updated functions  $\hat{V}_{q+1}$  and  $h_{NL,q+1}$ , and an updated stability region  $\Omega_{\hat{\rho}_{q+1}}$ , for the new empirical model, but do not yet update the LEMPC and control the process using the prior LEMPC. Else, if  $(t_{k+1} - t_{d,q}) \geq t_{h,q}$ , set  $i_{hx} = 1$  and apply  $h_{NL,q+1}(x(t_k))$ . Go to Step 2.  $t_k \leftarrow t_{k+1}$ .
  - (c) Operate the process under the LEMPC of Eq. 26 that was used at the prior sampling time. Go to Step 2.  $t_k \leftarrow t_{k+1}$ .

**Remark 2.** We assume  $t_{s,i+1}$  and  $t_{s,i+2}$  are separated by a sufficient length of time such that  $t_k > t_{s,i+2}$  always occurs after the closed-loop state has entered  $\Omega_{\hat{\rho}_{q+1}}$  when there is no attack.

**Remark 3.** An alternative to the implementation strategy described (which holds also for Detection Strategy 1) would be to design the original stability region so conservatively that the closed-loop state will not exit that region under a model change but only in the case of an attack. This loses the capability, however, to know if the model needs to be re-identified due to changing underlying dynamics via the method focused on whether the closed-loop state leaves  $\Omega_{\hat{\rho}_q}$ .

*3.2.2.1. Detection Strategy 2: Cyberattack-Mitigating State Feedback LEMPC: Stability and Feasibility Analysis* In this section, it will be demonstrated that in the presence of either a change in the underlying dynamics or an undetected cyberattack, the implementation strategy for Detection

Strategy 2 above maintains the closed-loop state within  $\Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q}$  sampling periods after a model change and at least one sampling period after an undetected cyberattack (providing some time available for auxiliary detection mechanisms to attempt to detect an attack that bypasses this detection strategy). This strategy achieves these goals in the presence of bounded process noise and disturbances, and with the attack potentially impacting all state measurements. To develop this proof, a number of propositions are presented. The first bounds the value of  $\hat{V}_q$  at any point in  $\Omega_{\hat{\rho}_q}$ , and the second bounds the difference between the closed-loop state of the system of Eq. 8 and that of Eq. 14 over time.

**Proposition 2.** *P. Mhaskar and Christofides (2013) Consider the Lyapunov function  $\hat{V}_q(\cdot)$ . There exists a quadratic function  $f_{V,q}(\cdot)$  such that:*

$$\hat{V}_q(x) \leq \hat{V}_q(x') + f_{V,q}(|x - x'|) \quad (27)$$

for all  $x, x' \in \Omega_{\hat{\rho}_{safe,q}}$  with

$$f_{V,q}(s) := \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_{safe,q}))s + M_{v,q}s^2 \quad (28)$$

where  $M_{v,q}$  is a positive constant.

**Proposition 3.** *Consider the systems*

$$\dot{\bar{x}}_{a,i} = \bar{f}_i(\bar{x}_{a,i}(t), \bar{u}_i(t), w_i(t)) \quad (29a)$$

$$\dot{\bar{x}}_{b,q} = \bar{f}_{NL,q}(\bar{x}_{b,q}(t), \bar{u}_q(t)) \quad (29b)$$

with initial states  $|\bar{x}_{a,i}(t_0) - \bar{x}_{b,q}(t_0)| \leq \delta$ , with  $\bar{x}_{a,i}(t_0)$  and  $\bar{x}_{b,q}(t_0)$  contained within  $\Omega_{\hat{\rho}_{safe,q}}$ , with  $t_0 = 0$ ,  $\bar{u}_i = \bar{u}_q + u_{q,s} - u_{i,s}$  contained within the input bounds, and  $w_i \in W_i$ . If  $\bar{x}_{a,i}(t)$  and  $\bar{x}_{b,q}(t)$  remain within  $\Omega_{\hat{\rho}_{safe,q}}$  for  $t \in [0, T]$ , then:

$$|\bar{x}_{a,i}(t) - \bar{x}_{b,q}(t)| \leq \left( \left( \delta + \frac{L_{w,i}\theta + M_{err,i,q}}{L_{x,i}} \right) e^{L_{x,i}t} \right) - \frac{L_{w,i}\theta + M_{err,i,q}}{L_{x,i}} \quad (30)$$

**Proof 1.** *The proof follows that in Giuliani and Durand (2018) and Durand (2020b) by taking the integral of Eqs. 29a and 29b, subtracting them, taking the norm with application of the triangle*

inequality, adding and subtracting  $\bar{f}_i(\bar{x}_{b,q}(s), \bar{u}_i(s), 0)$  on the right-hand side, and applying Eq. 21, Eq. 10a and the bound on  $w_i$  to give:

$$\begin{aligned} |\bar{x}_{a,i}(t) - \bar{x}_{b,q}(t)| &\leq |\bar{x}_{a,i}(t_0) - \bar{x}_{b,q}(t_0)| + \int_0^t |\bar{f}_i(\bar{x}_{a,i}(s), \bar{u}_i(s), w_i(s)) - \bar{f}_{NL,q}(\bar{x}_{b,q}(s), \bar{u}_q(s))| ds \\ &\leq \delta + (L_{w,i}\theta + M_{err,i,q})t + \int_0^t L_{x,i}|\bar{x}_{a,i}(s) - \bar{x}_{b,q}(s)| ds \end{aligned} \quad (31)$$

Using the Gronwall-Bellman inequality Khalil (2002), Eq. 30 is obtained.

The next proposition establishes a bound on the difference between the deviation form of the state of the system of Eq. 8 and the state of the nominal ( $w_i \equiv 0$ ) system.

**Proposition 4.** *Heidarinejad et al. (2012) Consider the systems*

$$\dot{\bar{x}}_{a,i} = \bar{f}_i(\bar{x}_{a,i}(t), \bar{u}_i(t), w_i(t)) \quad (32a)$$

$$\dot{\hat{x}}_{a,i} = \bar{f}_i(\bar{x}_{a,i}(t), \bar{u}_i(t), 0) \quad (32b)$$

with initial states  $\bar{x}_{a,i}(t_0) = \hat{x}_{a,i}(t_0)$  and contained within  $\Omega_{\hat{\rho}_{safe,q}}$ , with  $t_0 = 0$ , and  $w_i \in W_i$ . If  $\bar{x}_{a,i}(t)$  and  $\hat{x}_{a,i}(t)$  remain within  $\Omega_{\hat{\rho}_q}$  for  $t \in [0, T]$ , then:

$$|\bar{x}_{a,i}(t) - \hat{x}_{a,i}(t)| \leq \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}t} - 1) \quad (33)$$

The next proposition characterizes the error between the solution of the model of Eq. 14 and the approximate solution given by the following equation:

$$\tilde{\tilde{x}}_{b,q}(t) = \tilde{\tilde{x}}_{b,q}(t_j) + \sum_{n=1}^{N_1} \left( \bar{f}_{NL,q}^n(\tilde{\tilde{x}}_{b,q}(t_j), \bar{u}_q(t_j)) \frac{(t - t_j)^n}{n!} \right) \quad (34)$$

$\forall t \in [t_j, t_{j+1})$ ,  $j = k, \dots, k + N - 1$ , when  $\tilde{\tilde{x}}_{b,q}(t_k)$  is the state measurement at  $t_k$ .

**Proposition 5.** *Consider the solution of the system of Eq. 14 and  $\tilde{\tilde{x}}_{b,q}(t)$  from the model of Eq. 34.*

*There exists an upper bound on the error  $\bar{E}_{q,j}(t_j)$ ,  $j = k, \dots, k + N - 1$ , between  $\bar{x}_{b,q}(t)$  and  $\tilde{\tilde{x}}_{b,q}(t)$  throughout a sampling period beginning at  $t_j$ ,  $j = k, \dots, k + N - 1$ , in the interval  $[t_k, t_{k+N})$  under a sample-and-hold input policy defined by  $\bar{u}_q(t_j) \in U_q$ ,  $\forall t \in [t_j, t_{j+1})$ ,  $j = k, \dots, k + N - 1$ , where  $\bar{x}_{b,q}(t_k) = \tilde{\tilde{x}}_{b,q}(t_k)$  and  $\Delta < \Delta_{ub,q}$ , where the expression for the error is defined recursively by:*

$$|\bar{E}_{q,k}(t)| \leq \frac{M_{q,k,N_1} \Delta^{N_1+1}}{(N_1 + 1)!} := \bar{E}_{q,k}(t_k), \text{ for } t \in [t_k, t_{k+1}) \quad (35)$$

$$|\bar{E}_{q,k+p}(t)| \leq \bar{E}_{q,k+p-1}(t_{k+p-1}) + \frac{M_{q,k+p,N_1} \Delta^{N_1+1}}{(N_1+1)!} + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} L_{x,n,q} \bar{E}_{q,k+p-1}(t_{k+p-1}) := \bar{E}_{q,k+p}(t_{k+p}),$$

for  $t \in [t_{k+p}, t_{k+p+1}]$ ,  $p = 1, \dots, N-1$

(36)

**Proof 2.** This proof follows the proof of Proposition 4 in Rangan and Durand (2020). From Proposition 1,  $\bar{x}_{b,q}(t)$  can be represented as follows for  $t \in [t_j, t_{j+1}]$ ,  $j = k, \dots, k+N-1$ :

$$\bar{x}_{b,q}(t) = \bar{x}_{b,q}(t_j) + \sum_{n=1}^{\infty} \bar{f}_{NL,q}^n(\bar{x}_{b,q}(t_j), \bar{u}_q(t_j)) \frac{(t-t_j)^n}{n!}$$
(37)

$$= \bar{x}_{b,q}(t_j) + \sum_{n=1}^{N_1} \bar{f}_{NL,q}^n(\bar{x}_{b,q}(t_j), \bar{u}_q(t_j)) \frac{(t-t_j)^n}{n!} + E_{q,j}(t)$$
(38)

$\forall t \in [t_j, t_{j+1}]$ ,  $j = k, \dots, k+N-1$ , where  $E_{q,j}(t)$  represents the Taylor series error from truncating the Taylor series representation of the solution of Eq. 14 to  $N_1+1$  terms. Defining  $\bar{E}_{q,j}(t) = \bar{x}_{b,q}(t) - \tilde{x}_{b,q}(t)$  for  $t \in [t_j, t_{j+1}]$ , then for  $t \in [t_k, t_{k+1}]$ ,  $\tilde{x}_{b,q}(t) = \bar{x}_{b,q}(t) - \bar{E}_{q,k}(t) = \bar{x}_{b,q}(t) - E_{q,k}(t)$ , and

$$|\bar{E}_{q,k}(t)| = |\bar{x}_{b,q}(t) - \tilde{x}_{b,q}(t)| \leq \frac{M_{q,k,N_1} (\Delta)^{N_1+1}}{(N_1+1)!} := \bar{E}_{q,k}(t_k)$$
(39)

from Eq. 23, with  $M_{q,k,N_1} > 0$ . For  $t \in [t_{k+1}, t_{k+2}]$ , the following equations hold:

$$\tilde{x}_{b,q}(t) = \tilde{x}_{b,q}(t_{k+1}) + \sum_{n=1}^{N_1} \left( \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_{k+1}), \bar{u}_q(t_{k+1})) \frac{(t-t_{k+1})^n}{n!} \right)$$
(40)

$$\bar{x}_{b,q}(t) = \bar{x}_{b,q}(t_{k+1}) + \sum_{n=1}^{N_1} \left( \bar{f}_{NL,q}^n(\bar{x}_{b,q}(t_{k+1}), \bar{u}_q(t_{k+1})) \frac{(t-t_{k+1})^n}{n!} \right) + E_{q,k+1}(t)$$
(41)

Taking the Euclidean norm of Eq. 41 minus Eq. 40 and applying the triangle inequality and Eqs. 23, 39 and 17a gives:

$$\begin{aligned} |\bar{x}_{b,q}(t) - \tilde{x}_{b,q}(t)| &\leq |\bar{x}_{b,q}(t_{k+1}) - \tilde{x}_{b,q}(t_{k+1})| + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} \left| \bar{f}_{NL,q}^n(\bar{x}_{b,q}(t_{k+1}), \bar{u}_q(t_{k+1})) - \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_{k+1}), \bar{u}_q(t_{k+1})) \right| \\ &\quad + |E_{q,k+1}(t)| \\ &\leq \frac{M_{q,k,N_1} (\Delta)^{N_1+1}}{(N_1+1)!} + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} L_{x,n,q} |\bar{x}_{b,q}(t_{k+1}) - \tilde{x}_{b,q}(t_{k+1})| + \frac{M_{q,k+1,N_1} (\Delta)^{N_1+1}}{(N_1+1)!} \\ &\leq \frac{M_{q,k,N_1} (\Delta)^{N_1+1}}{(N_1+1)!} + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} L_{x,n,q} \frac{M_{q,k,N_1} (\Delta)^{N_1+1}}{(N_1+1)!} + \frac{M_{q,k+1,N_1} (\Delta)^{N_1+1}}{(N_1+1)!} := \bar{E}_{q,k+1}(t_{k+1}) \end{aligned}$$
(42)

for  $t \in [t_{k+1}, t_{k+2}]$ , where  $\bar{x}_{b,q}(t_{k+1})$  and  $\tilde{x}_{b,q}(t_{k+1}) \in \Omega_{\hat{\rho}_q}$ , and  $M_{q,k+1,N_1} > 0$ .

Continuing to follow this procedure for subsequent sampling periods gives the general form of the error bound for  $t \in [t_{k+p}, t_{k+p+1}]$ , where  $p = 1, \dots, N-1$ , as follows, with  $M_{q,k+p,N_1} > 0$ :

$$|\bar{E}_{q,k+p}(t)| \leq \bar{E}_{q,k+p-1}(t_{k+p-1}) + \frac{M_{q,k+p,N_1} \Delta^{N_1+1}}{(N_1+1)!} + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} L_{x,n,q} \bar{E}_{q,k+p-1}(t_{k+p-1}) \quad (43)$$

From Proposition 5, decreasing  $N_1$  and  $\Delta$  decreases the error between  $\bar{x}_{b,q}$  and  $\tilde{x}_{b,q}$  throughout a sampling period. For a given  $N_1$  and  $\Delta$ , the error at the end of the prediction horizon will be less when the prediction horizon includes less sampling periods. Furthermore,  $\bar{E}_{q,j}$ ,  $j = k+1, \dots, k+N-1$ , in Proposition 5 incorporates error both from truncation of the Taylor series solution and from using the approximate value  $\tilde{x}_{b,q}$  at each sampling time in the approximation of the dynamics at a subsequent time.  $E_{q,j}$ , in contrast, only reflects truncation error. The following proposition bounds the difference between the state trajectory of Eq. 34 and the solution of the nominal ( $w_i \equiv 0$ ) system of Eq. 8.

**Proposition 6.** *Consider the following systems:*

$$\hat{x}_{a,i}(t) = \hat{x}_{a,i}(t_0) + \sum_{n=1}^{\infty} \left( \bar{f}_i^n(\hat{x}_{a,i}(t_0), \bar{u}_i(t_0), 0) \frac{(t-t_0)^n}{n!} \right) \quad (44)$$

$$\tilde{x}_{b,q}(t) = \tilde{x}_{b,q}(t_0) + \sum_{n=1}^{N_1} \left( \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_0), \bar{u}_q(t_0)) \frac{(t-t_0)^n}{n!} \right) \quad (45)$$

with initial states  $|\hat{x}_{a,i}(t_0) - \tilde{x}_{b,q}(t_0)| \leq \delta$  and  $\hat{x}_{a,i}(t_0), \tilde{x}_{b,q}(t_0) \in \Omega_{\hat{\rho}_q}$  with  $t_0 = 0$ ,  $\bar{u}_i(t_0) \in U_i$ ,  $\bar{u}_q(t_0) \in U_q$ ,  $\bar{u}_i(t_0) = \bar{u}_q(t_0) + u_{q,s} - u_{i,s}$ , and  $w_i \in W_i$ . Also consider that for all  $n \geq 1$ :

$$|\bar{f}_i^n(\bar{x}, u, 0) - \bar{f}_{NL,q}^n(\bar{x}', u')| \leq M_{deriv,i,q} \quad (46)$$

for all  $|\bar{x} - \bar{x}'| \leq \delta$  and contained in  $\Omega_{\hat{\rho}_{safe,q}}$ , and for all  $u = u' + u_{q,s} - u_{i,s}$  in the input bounds, and Eq. 12 holds. If  $\hat{x}_{a,i}(t), \tilde{x}_{b,q}(t) \in \Omega_{\hat{\rho}_q}$ , then for  $t \in [0, t_{s,i+1}]$ ,

$$|\hat{x}_{a,i}(t) - \tilde{x}_{b,q}(t)| \leq \delta + \sum_{n=1}^{N_1} \frac{(t-t_0)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1} (t-t_0)^{N_1+1}}{(N_1+1)!} \quad (47)$$

**Proof 3.** Taking the Euclidean norm of Eq. 44 minus Eq. 45 and applying Eqs. 23, 12 and 46 and the triangle inequality and taking the norm on both sides:

$$\begin{aligned}
|\hat{x}_{a,i}(t) - \tilde{x}_{b,q}(t)| &\leq |\hat{x}_{a,i}(t_0) - \tilde{x}_{b,q}(t_0)| + \\
&\left| \sum_{n=1}^{\infty} (\bar{f}_i^n(\hat{x}_{a,i}(t_0), \bar{u}_i(t_0), 0)) \frac{(t-t_0)^n}{n!} - \sum_{n=1}^{N_1} \left( \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_0), \bar{u}_q(t_0)) \frac{(t-t_0)^n}{n!} \right) \right| \\
&\leq \delta + \sum_{n=1}^{N_1} \frac{(t-t_0)^n}{n!} |\bar{f}_i^n(\hat{x}_{a,i}(t_0), \bar{u}_i(t_0), 0) - \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_0), \bar{u}_q(t_0))| + \frac{M_{i,N_1}(t-t_0)^{N_1+1}}{(N_1+1)!} \\
&\leq \delta + \sum_{n=1}^{N_1} \frac{(t-t_0)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(t-t_0)^{N_1+1}}{(N_1+1)!}
\end{aligned} \tag{48}$$

The following proposition provides the conditions under which  $h_{NL,q}$  implemented in sample-and-hold can maintain the closed-loop state of the system of Eq. 34 within  $\Omega_{\hat{\rho}_q}$  when it is initialized within that region.

**Proposition 7.** Consider the model of Eq. 34 under the Lyapunov-based controller  $h_{NL,q}$  implemented in a sample-and-hold fashion from  $t_k$  to  $t_{k+N}$  that satisfies the requirements of Eqs. 15a-15d and 18. If  $\tilde{x}_{b,q}(t_k) \in \Omega_{\hat{\rho}_q}$ ,  $0 < \Delta < \Delta_{ub,q}$  and

$$\begin{aligned}
L_{v,q} \left[ M_{L,q} \Delta + \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^n}{n!} \right] M_{L,q} + \sum_{n=2}^{N_1} L_{v,q} \left[ M_{L,q} \Delta + \sum_{\bar{n}=2}^{N_1} M_{q,k,\bar{n}-1} \frac{\Delta^{\bar{n}}}{\bar{n}!} \right] M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \\
- \hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,q})) + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q)) \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \leq -\epsilon_{w,q}/\Delta
\end{aligned} \tag{49}$$

$$\hat{\rho}'_{\min,q} = \max\{\hat{V}_q(\tilde{x}_{b,q}(t+\Delta)) : \hat{V}_q(\tilde{x}_{b,q}(t)) \leq \hat{\rho}_{s,q}\} \tag{50}$$

for  $j = k, \dots, k+N-1$ , where  $\epsilon_{w,q} > 0$ ,  $L_{v,q} > 0$ , and  $\hat{\rho}_q > \hat{\rho}'_{\min,q} > \hat{\rho}_{s,q}$ , then  $\tilde{x}_{b,q}(t) \in \Omega_{\hat{\rho}_q}$  for  $t \in [t_k, t_{k+N})$ .

**Proof 4.** Eq. 15b gives:

$$\frac{\partial \hat{V}_q(\tilde{x}_{b,q}(t_j))}{\partial \tilde{x}_{b,q}} \bar{f}_{NL,q}(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j))) \leq -\hat{\alpha}_{3,q}(|\tilde{x}_{b,q}(t_j)|) \tag{51}$$



From Eq. 34, for  $t \in [t_j, t_{j+1})$ :

$$\begin{aligned} \frac{d\hat{V}_q(\tilde{x}_{b,q}(t))}{dt} &= \frac{\partial\hat{V}_q(\tilde{x}_{b,q}(t))}{\partial\tilde{x}_{b,q}} \left( \sum_{n=1}^{N_1} (\bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j)))) \frac{(t-t_j)^{n-1}}{(n-1)!} \right) \\ &= \frac{\partial\hat{V}_q(\tilde{x}_{b,q}(t))}{\partial\tilde{x}_{b,q}} \left( \bar{f}_{NL,q}(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j))) + \sum_{n=2}^{N_1} (\bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j)))) \frac{(t-t_j)^{n-1}}{(n-1)!} \right) \end{aligned} \quad (52)$$

With  $\frac{\partial\hat{V}_q(\tilde{x}_{b,q}(t_j))}{\partial\tilde{x}_{b,q}} \left( \bar{f}_{NL,q}(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j))) + \sum_{n=2}^{N_1} (\bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j)))) \frac{(t-t_j)^{n-1}}{(n-1)!} \right)$

added and subtracted from the right-hand side of Eq. 52, and Eqs. 15 and 51, we obtain:

$$\begin{aligned} \frac{d\hat{V}_q(\tilde{x}_{b,q}(t))}{dt} &\leq \left| \frac{\partial\hat{V}_q(\tilde{x}_{b,q}(t))}{\partial\tilde{x}_{b,q}} - \frac{\partial\hat{V}_q(\tilde{x}_{b,q}(t_j))}{\partial\tilde{x}_{b,q}} \right| \left| \bar{f}_{NL,q}(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j))) \right| \\ &\quad + \sum_{n=2}^{N_1} \left| \frac{\partial\hat{V}_q(\tilde{x}_{b,q}(t))}{\partial\tilde{x}_{b,q}} - \frac{\partial\hat{V}_q(\tilde{x}_{b,q}(t_j))}{\partial\tilde{x}_{b,q}} \right| \left| \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j))) \right| \frac{\Delta^{n-1}}{(n-1)!} \\ &\quad - \hat{\alpha}_{3,q}(|\tilde{x}_{b,q}(t_j)|) + \sum_{n=2}^{N_1} \hat{\alpha}_{4,q}(|\tilde{x}_{b,q}(t_j)|) \left| \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_j), h_{NL,q}(\tilde{x}_{b,q}(t_j))) \right| \frac{\Delta^{n-1}}{(n-1)!} \end{aligned} \quad (53)$$

Because  $\hat{V}_q$  is infinitely differentiable, there exists  $L_{v,q} > 0$  such that:

$$\left| \frac{\partial\hat{V}_q(x')}{\partial x} - \frac{\partial\hat{V}_q(x'')}{\partial x} \right| \leq L_{v,q} |x' - x''| \quad (54)$$

for all  $x, x'' \in \Omega_{\hat{\rho}_{s_a f_e, q}}$ . Using Eqs. 54, 51, 16a, 25, and 15c:

$$\begin{aligned} \frac{d\hat{V}_q(\tilde{x}_{b,q}(t))}{dt} &\leq L_{v,q} |\tilde{x}_{b,q}(t) - \tilde{x}_{b,q}(t_j)| M_{L,q} + \sum_{n=2}^{N_1} L_{v,q} |\tilde{x}_{b,q}(t) - \tilde{x}_{b,q}(t_j)| M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \\ &\quad - \hat{\alpha}_{3,q}(|\tilde{x}_{b,q}(t_j)|) + \hat{\alpha}_{4,q}(|\tilde{x}_{b,q}(t_j)|) \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \end{aligned} \quad (55)$$

Using the definition of  $\tilde{x}_{b,q}(t)$  from Eq. 34 and Eqs. 16a and 25 gives:

$$|\tilde{x}_{b,q}(t) - \tilde{x}_{b,q}(t_j)| = \left| \sum_{n=1}^{N_1} \bar{f}_{NL,q}^n(\tilde{x}_{b,q}(t_j), h_{NL,q}(t_j)) \frac{(t-t_j)^n}{n!} \right| \leq M_{L,q} \Delta + \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^n}{n!} \quad (56)$$

Combining Eq. 56 with Eq. 55 gives:

$$\begin{aligned} \frac{d\hat{V}_q(\tilde{x}_{b,q}(t))}{dt} &\leq L_{v,q} \left[ M_{L,q} \Delta + \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^n}{n!} \right] M_{L,q} + \sum_{n=2}^{N_1} L_{v,q} \left[ M_{L,q} \Delta + \sum_{\bar{n}=2}^{N_1} M_{q,k,\bar{n}-1} \frac{\Delta^{\bar{n}}}{\bar{n}!} \right] M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \\ &\quad - \hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(|\tilde{x}_{b,q}(t_j)|)) + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(|\tilde{x}_{b,q}(t_j)|)) \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \end{aligned} \quad (57)$$

If  $\tilde{x}_{b,q}(t_j) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{s,q}}$ , then for  $t \in [t_j, t_{j+1}]$ :

$$\begin{aligned} \frac{d\hat{V}_q(\tilde{x}_{b,q}(t))}{dt} \leq & L_{v,q} \left[ M_{L,q}\Delta + \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^n}{n!} \right] M_{L,q} + \sum_{n=2}^{N_1} L_{v,q} \left[ M_{L,q}\Delta + \sum_{\bar{n}=2}^{N_1} M_{q,k,\bar{n}-1} \frac{\Delta^{\bar{n}}}{\bar{n}!} \right] M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \\ & - \hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,q})) + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q)) \sum_{n=2}^{N_1} M_{q,k,n-1} \frac{\Delta^{n-1}}{(n-1)!} \end{aligned} \quad (58)$$

If Eq. 49 holds, then  $\hat{V}_q(\tilde{x}_{b,q}(t)) < \hat{V}_q(\tilde{x}_{b,q}(t_j))$ ,  $\forall t \in (t_j, t_{j+1}]$ , so that  $\tilde{x}_{b,q}(t) \in \Omega_{\hat{\rho}_q}$ . If instead  $\tilde{x}_{b,q}(t_j) \in \Omega_{\hat{\rho}_{s,q}}$ , then  $\tilde{x}_{b,q}(t) \in \Omega_{\hat{\rho}'_{\min,q}}$ ,  $\forall t \in [t_j, t_{j+1}]$ , from Eq. 50. When  $\hat{\rho}_q > \hat{\rho}'_{\min,q}$  as required by the theorem,  $\tilde{x}_{b,q}(t_{j+1}) \in \Omega_{\hat{\rho}_q}$  in this case also. Therefore, if  $\tilde{x}_{b,q}(t_k) \in \Omega_{\hat{\rho}_q}$ , then  $\tilde{x}_{b,q}(t) \in \Omega_{\hat{\rho}_q}$  for  $t \in [t_k, t_{k+N})$ .

The following theorem guarantees that in the presence of bounded measurement noise and disturbances, the implementation strategy of Section 3.2.2: 1) maintains the closed-loop state within  $\Omega_{\hat{\rho}_q}$  before an attack or model change occurs; 2) maintains the closed-loop state in  $\Omega_{\hat{\rho}_{safe,q}}$  for at least one sampling period after an attack; and 3) maintains the closed-loop state in  $\Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q}$  sampling periods after  $t_{d,q}$  if no attack occurs.

**Theorem 1.** Consider the system of Eq. 8 in closed-loop, under the implementation strategy of Section 3.2.2 based on a controller  $h_{NL,q}(\cdot)$  that satisfies Eqs. 15a-15d and 18, as well as the requirements of Proposition 7. Let  $\epsilon'_{w,q,i} > 0$ ,  $\bar{\epsilon}'_{w,q,i+1} > 0$ ,  $0 < \Delta < \Delta_{ub,q}$ ,  $N \geq 1$ ,  $\hat{\rho}_{samp2,i+1,q} = \hat{\rho}_q + f_{V,q}(\theta_v) + \bar{\epsilon}'_{w,q,i+1} > 0$ ,  $\Omega_{\hat{\rho}_{samp,q}} \subset \Omega_{\hat{\rho}_q} \subset \Omega_{\hat{\rho}_{safe,q}} \subset X_q$ ,  $q \geq 1$ ,  $\hat{\rho}_{samp,q} > \hat{\rho}_{e,q} > \hat{\rho}'_{\min,q} > \hat{\rho}_{s,q} > 0$ ,  $\hat{\rho}_{e,q} > \hat{\rho}_{\min,i,q} > \hat{\rho}_{s,q} > 0$ , and  $\hat{\rho}_{e,q} > \hat{\rho}_{\min,i+1,q} > \hat{\rho}_{s,q} > 0$ . If the following equations are satisfied:

$$- \hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,q})) + L'_{x,i}(\theta_v + M_{i,0}\Delta) + L'_{w,i}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q))M_{deriv,i,q} \leq -\epsilon'_{w,q,i}/\Delta \quad (59)$$

$$- \hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,q})) + L'_{x,i+1}(\theta_v + M_{i+1,0}\Delta) + L'_{w,i+1}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q))M_{deriv,i+1,q} \leq \bar{\epsilon}'_{w,q,i+1}/\Delta \quad (60)$$

$$\hat{\rho}_{e,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) \right) \leq \hat{\rho}_{samp,q} \quad (61)$$

$$\hat{\rho}_{e,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) \right) + f_{V,q}(\theta_v) \leq \hat{\rho}_q \quad (62)$$

$$\hat{\rho}_{\min,i,q} := \max\{\hat{V}_q(\bar{x}_{a,i}(t + \Delta)) : \hat{V}_q(\bar{x}_{a,i}(t)) \leq \hat{\rho}_{s,q}\} \quad (63)$$

$$\hat{\rho}_{smp,q} + f_{V,q}(\theta_v) \leq \hat{\rho}_q \quad (64)$$

$$2\theta_v + \frac{L_{w,i}\theta}{L_{x,i}}(e^{L_{x,i}\Delta} - 1) + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} \leq \nu_{s,q} \quad (65)$$

$$2\theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}\Delta^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) \leq \nu_{l,q} \quad (66)$$

$$\hat{\rho}_{smp2,i+1,q} + \frac{\bar{e}'_{w,q,i+1}(t_{h,q} - \Delta)}{\Delta} := \hat{\rho}_{far} \quad (67)$$

$$\begin{aligned} & \hat{\rho}_{smp,q} + f_{V,q} \left( \frac{L_{w,i}\theta}{L_{x,i}}(e^{L_{x,i}\Delta} - 1) + \theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} + \nu_{l,q} \right) \\ & + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) + f_{V,q} \left( \delta + \frac{L_{w,i}\theta}{L_{x,i}}(e^{L_{x,i}\Delta} - 1) + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} \right) \leq \hat{\rho}_{safe,q} \end{aligned} \quad (68)$$

$$\begin{aligned} & \hat{\rho}_{far} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) + \nu_{l,q} \right) \\ & + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) + f_{V,q} \left( \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) \right) \\ & + \delta + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} \leq \hat{\rho}_{safe,q} \end{aligned} \quad (69)$$

$$\begin{aligned} & \hat{\rho}_{smp,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i}\theta}{L_{x,i}}(e^{L_{x,i}\Delta} - 1) + \nu_{l,q} \right) + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) \\ & + f_{V,q} \left( \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) + \delta + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} \right) \leq \hat{\rho}_{safe,q} \end{aligned} \quad (70)$$

$$\begin{aligned} & \hat{\rho}_{smp,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} \right) \\ & + \frac{L_{w,i}\theta}{L_{x,i}}(e^{L_{x,i}\Delta} - 1) + \nu_{l,q} + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) + f_{V,q} \left( M_{i+1,0}\Delta + \frac{L_{w,i}\theta}{L_{x,i}}(e^{L_{x,i}\Delta} - 1) \right) \\ & + \delta + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \leq \hat{\rho}_{safe,q} \end{aligned} \quad (71)$$

with  $\nu_{l,q} \geq \nu_{s,q}$ ,  $x(t_0) \in \Omega_{\hat{\rho}_{e,q}}$ ,  $\bar{x}_{a,i}(t_0) \in \Omega_{\hat{\rho}_{e,q}}$ , and  $|\bar{x}_{a,i}(t_k) - x(t_k)| \leq \delta$ ,  $k = 0, 1, \dots$ , then the closed-loop state is contained in  $\Omega_{\hat{\rho}_{smp,q}}$  and the state measurement is in  $\Omega_{\hat{\rho}_q}$  for all  $t \geq 0$  until the dynamics of the process change at  $t_{s,i+1}$  or there is a cyberattack on the sensors at  $t_A$ . Furthermore,

$\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{safe,q}}$  for at least one sampling period after  $t_A$  and  $\bar{x}_{a,i+1}(t) \in \Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q} = \text{floor}\left(\frac{\hat{\rho}_{safe,q} - \hat{\rho}_{samp2,i+1,q}}{\epsilon_{w,q,i+1}}\right)$  sampling periods after  $t_{d,q}$  if  $t_A > t_{ID,q}$ .

*Proof.* The proof consists of five parts. In the first part, recursive feasibility at every sampling time in which the LEMPC of Eq. 26 is used under the implementation strategy is demonstrated. In the second part, it is demonstrated that the closed-loop state and state measurement are maintained within  $\Omega_{\hat{\rho}_q}$  before any model change or cyberattack occurs. In the third part, it is demonstrated that after  $t_{d,q}$ , if only a model change occurs that is detected via either the closed-loop state measurement leaving  $\Omega_{\hat{\rho}_q}$  or if it is detected via  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > v_{s,q}$ , the closed-loop state stays within  $\Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q}$  sampling times and no attack will be flagged by the updated detection mechanism ( $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > v_{l,q}$ ) before  $t_{ID,q}$ . In the fourth part, it is demonstrated that if there is no change in the dynamics but there is an undetected attack (whether it occurs when  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| < v_{s,q}$  is checked or if it occurs at  $t_{d,q}$  when  $v_{l,q} < |\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| < v_{s,q}$ ), there is at least one sampling period before the closed-loop state leaves  $\Omega_{\hat{\rho}_{safe,q}}$ . In the fifth part, it is demonstrated that if there is a change in the dynamic model as well as an undetected cyberattack, then the closed-loop state is maintained within  $\Omega_{\hat{\rho}_{safe,q}}$  for at least one sampling period after the attack.

*Part 1.*  $h_{NL,q}$  implemented in sample-and-hold is a feasible input policy for the LEMPC of Eq. 26 whenever the LEMPC of Eq. 26 is used according to the implementation strategy in Section 3.2.2. Specifically,  $h_{NL,q}$  in sample-and-hold maintains  $\tilde{\tilde{x}}_{b,q}$  in  $\Omega_{\hat{\rho}_q} \subset X_q$  according to Proposition 7 (i.e., Eq. 26f is met) and meets the constraints of Eqs. 26d-26e. In addition, it trivially satisfies Eq. 26g.

*Part 2.* In this part, we demonstrate that before any attack or change in the underlying dynamics, the closed-loop state is maintained within  $\Omega_{\hat{\rho}_{samp,q}} \subset \Omega_{\hat{\rho}_q}$  and the state measurement is maintained within  $\Omega_{\hat{\rho}_q}$ . In this case, either  $x(t_k) \in \Omega_{\hat{\rho}_{e,q}}$  so that the constraint of Eq. 26f is activated, or  $x(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,q}}$  so that the constraint of Eq. 26g is activated.

Consider first the case that  $x(t_k) \in \Omega_{\hat{\rho}_{e,q}}$ . Eq. 26f ensures that  $\tilde{\tilde{x}}_{b,q}(t)$  is maintained within  $\Omega_{\hat{\rho}_{e,q}}$  throughout the prediction horizon, so we must demonstrate that  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{samp,q}}$  and  $x(t_{k+1}) \in \Omega_{\hat{\rho}_q}$  for  $t \in [t_k, t_{k+1})$ . From Proposition 2, and defining  $\theta_v$  to be the measurement noise associated with

full state feedback:

$$\begin{aligned}
\hat{V}_q(\bar{x}_{a,i}(t)) &\leq \hat{V}_q(\tilde{x}_{b,q}(t)) + f_{V,q}(|\tilde{x}_{b,q}(t) - \bar{x}_{a,i}(t)|) \\
&\leq \hat{V}_q(\tilde{x}_{b,q}(t)) + f_{V,q}(|\tilde{x}_{b,q}(t) - \hat{x}_{a,i}(t)| + |\hat{x}_{a,i}(t) - \bar{x}_{a,i}(t)|) \\
&\leq \hat{\rho}_{e,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) \right)
\end{aligned} \tag{72}$$

for  $t \in [t_k, t_{k+1})$  if  $\bar{x}_{a,i}(t)$  and  $\tilde{x}_{b,q}(t) \in \Omega_{\hat{\rho}_q}$ , where the second inequality follows from Eqs. 26f, 47, and 33. If Eq. 61 holds, then  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{samp,q}}$  for  $t \in [t_k, t_{k+1})$  when  $x(t_k) \in \Omega_{\hat{\rho}_{e,q}}$ .

To ensure that  $x(t_{k+1}) \in \Omega_{\hat{\rho}_q}$ , Eq. 72 and Proposition 2 give:

$$\begin{aligned}
\hat{V}_q(x(t_{k+1})) &\leq \hat{V}_q(\bar{x}_{a,i}(t_{k+1})) + f_{V,q}(|x(t_{k+1}) - \bar{x}_{a,i}(t_{k+1})|) \\
&\leq \hat{\rho}_{e,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) \right) + f_{V,q}(\theta_v)
\end{aligned} \tag{73}$$

When Eq. 62 holds, Eq. 73 gives that  $x(t_{k+1}) \in \Omega_{\hat{\rho}_q}$  when  $x(t_k) \in \Omega_{\hat{\rho}_{e,q}}$ .

Next, we evaluate the case that  $x(t_k) \in \Omega_{\hat{\rho}}/\Omega_{\hat{\rho}_{e,q}}$  (i.e., Eq. 26g is activated). The time derivative of the Lyapunov function along the state trajectory of the system of Eq. 8 can be written as follows:

$$\dot{\hat{V}}_q(\bar{x}_{a,i}(t)) = \frac{\partial \hat{V}_q(\bar{x}_{a,i}(t))}{\partial x} \bar{f}_i(\bar{x}_{a,i}(t), \bar{u}_i(t_k), w_i(t)) \tag{74}$$

for  $t \in [t_k, t_{k+1})$ . Adding and subtracting  $\frac{\partial \hat{V}_q(\bar{x}_{b,q}(t_k))}{\partial x} \bar{f}_{NL,q}(\bar{x}_{b,q}(t_k), \bar{u}_q(t_k))$  and  $\frac{\partial \hat{V}_q(\bar{x}_{b,q}(t_k))}{\partial x} \bar{f}_i(\bar{x}_{b,q}(t_k), \bar{u}_i(t_k), 0)$  to/from the above equation (where  $\bar{u}_i(t_k) = \bar{u}_q(t_k) + u_{q,s} - u_{i,s}$ ), and using Eqs. 26g, 15b, 10b, 46, 15c, and the bound on  $w_i$ , we obtain that:

$$\begin{aligned}
\dot{\hat{V}}_q(\bar{x}_{a,i}(t)) &\leq -\hat{\alpha}_{3,q}(|\bar{x}_{b,q}(t_k)|) + \frac{\partial \hat{V}_q(\bar{x}_{a,i}(t))}{\partial x} \bar{f}_i(\bar{x}_{a,i}(t), \bar{u}_i(t_k), w_i(t)) - \frac{\partial \hat{V}_q(\bar{x}_{b,q}(t_k))}{\partial x} \bar{f}_i(\bar{x}_{b,q}(t_k), \bar{u}_i(t_k), 0) \\
&\quad + \frac{\partial \hat{V}_q(\bar{x}_{b,q}(t_k))}{\partial x} \bar{f}_i(\bar{x}_{b,q}(t_k), \bar{u}_i(t_k), 0) - \frac{\partial \hat{V}_q(\bar{x}_{b,q}(t_k))}{\partial x} \bar{f}_{NL,q}(\bar{x}_{b,q}(t_k), \bar{u}_q(t_k)) \\
&\leq -\hat{\alpha}_{3,q}(|\bar{x}_{b,q}(t_k)|) + L'_{x,i} |\bar{x}_{a,i}(t) - \bar{x}_{b,q}(t_k)| + L'_{w,i} \theta + \left| \frac{\partial \hat{V}_q(\bar{x}_{b,q}(t_k))}{\partial x} \right| M_{deriv,i,q} \\
&\leq -\hat{\alpha}_{3,q}(|\bar{x}_{b,q}(t_k)|) + L'_{x,i} (|\bar{x}_{a,i}(t) - \bar{x}_{a,i}(t_k)| + |\bar{x}_{a,i}(t_k) - \bar{x}_{b,q}(t_k)|) + L'_{w,i} \theta + \hat{\alpha}_{4,q}(|\bar{x}_{b,q}(t_k)|) M_{deriv,i,q} \\
&\leq -\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,q})) + L'_{x,i} (\theta_v + M_{i,0}\Delta) + L'_{w,i} \theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q)) M_{deriv,i,q}
\end{aligned} \tag{75}$$

for all  $\bar{x}_{b,q}(t_k) = x(t_k) \in \Omega_{\hat{\rho}}/\Omega_{\hat{\rho}_{s,q}}$ . If the condition of Eq. 59 is satisfied:

$$\hat{V}_q(\bar{x}_{a,i}(t)) \leq \hat{V}_q(\bar{x}_{a,i}(t_k)) - \frac{\epsilon'_{w,q,i}(t-t_k)}{\Delta}, \quad t \in [t_k, t_{k+1}) \tag{76}$$

Thus, when  $x(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,q}}$ , then  $\hat{V}_q(\bar{x}_{a,i}(t))$  decreases over the subsequent sampling period. Since  $\hat{\rho}_{e,q} > \hat{\rho}_{s,q}$ ,  $x(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,q}}$  only if  $x(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{s,q}}$ .  $\bar{x}_{a,i}(t_k)$  is guaranteed to be within  $\Omega_{\hat{\rho}_{samp,q}}$  when the conditions of the theorem are satisfied, as demonstrated below. In addition, we can demonstrate that  $x(t_{k+1}) \in \Omega_{\hat{\rho}_q}$  using Proposition 2:

$$\begin{aligned}\hat{V}_q(x(t_{k+1})) &\leq \hat{V}_q(\bar{x}_{a,i}(t_{k+1})) + f_{V,q}(|x(t_{k+1}) - \bar{x}_{a,i}(t_{k+1})|) \\ &\leq \hat{V}_q(\bar{x}_{a,i}(t_k)) + f_{V,q}(\theta_v) \\ &\leq \hat{\rho}_{samp,q} + f_{V,q}(\theta_v)\end{aligned}\tag{77}$$

When Eq. 64 holds,  $x(t_{k+1}) \in \Omega_{\hat{\rho}_q}$  when  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{samp,q}}$ .

The results above require that  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{samp,q}}$  whenever  $x(t_k) \in \Omega_{\hat{\rho}_q}$ . To demonstrate that this always holds under the proposed implementation strategy, we note that initially,  $\bar{x}_{a,i}(t_0)$  and  $x(t_0) \in \Omega_{\hat{\rho}_{e,q}}$  as assumed in the theorem. As a result, from  $t_0$  to  $t_1$ , Eqs. 73 and 62 guarantee that  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{samp,q}}$  for  $t \in [t_k, t_{k+1})$  and  $x(t_{k+1}) \in \Omega_{\hat{\rho}_q}$ . At the next sampling time, one of four things happens: 1)  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{samp,q}}/\Omega_{\hat{\rho}_{s,q}}$  and  $x(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,q}}$ ; 2)  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{samp,q}}/\Omega_{\hat{\rho}_{s,q}}$  and  $x(t_k) \in \Omega_{\hat{\rho}_{e,q}}$ ; 3)  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{s,q}}$  and  $x(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,q}}$ ; or 4)  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{s,q}}$  and  $x(t_k) \in \Omega_{\hat{\rho}_{e,q}}$ . In the first case, Eq. 76 demonstrates that  $\hat{V}_q(\bar{x}_{a,i}(t))$  decreases over the subsequent sampling period so that since  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{samp,q}}$ ,  $\bar{x}_{a,i}(t_{k+1}) \in \Omega_{\hat{\rho}_{samp,q}}$  as well. In the second case, Eqs. 72 and 61 guarantee that  $\bar{x}_{a,i}(t_{k+1}) \in \Omega_{\hat{\rho}_{samp,q}}$ . In the third case and the fourth case, Eq. 63 and the assumption that  $\hat{\rho}_{\min,i,q} < \hat{\rho}_{e,q}$  guarantees that  $\bar{x}_{a,i}(t_{k+1}) \in \Omega_{\hat{\rho}_{samp,q}}$ . Therefore, applying this recursively,  $\bar{x}_{a,i}(t)$  is always maintained in  $\Omega_{\hat{\rho}_{samp,q}}$  and the state measurement is always maintained within  $\Omega_{\hat{\rho}_q}$  under the proposed implementation strategy in the absence of a cyberattack or a model change.

In addition, it remains to be demonstrated that when there is no attack or model change, the condition  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{s,q}$  always holds (i.e., there will be no false alarms) if  $\nu_{s,q}$  is selected to satisfy Eq. 65. To demonstrate this, we note that Eq. 47 and Proposition 4 give:

$$\begin{aligned}|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| &\leq |\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - \hat{\tilde{x}}_{a,i}(t_k|t_{k-1}) + \hat{\tilde{x}}_{a,i}(t_k|t_{k-1}) - \bar{x}_{a,i}(t_k) + \bar{x}_{a,i}(t_k) - x(t_k)| \\ &\leq 2\theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{L_{w,i}\theta}{L_{x,i}}(e^{L_{x,i}\Delta} - 1) + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!}\end{aligned}\tag{78}$$

If Eq. 65 holds, then at all times,  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{s,q}$  such that there are no false alarms with this detection threshold.

*Part 3.* In this part, we demonstrate that after a model change occurs, if there is no attack, the closed-loop state stays in  $\Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q}$  sampling periods after  $t_{d,q}$  and no attack is detected after  $t_{d,q}$  until  $t_{ID,q}$  (i.e., there are no false alarms under the proposed implementation strategy).

Until  $t_{s,i+1}$ ,  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{samp,q}}$  and the state measurement is maintained within  $\Omega_{\hat{\rho}_q}$  under the implementation strategy of Section 3.2.2 as proven in Part 2. After  $t_{s,i+1}$  and until  $t_{d,q}$ , either: 1) the state measurement is maintained within  $\Omega_{\hat{\rho}_q}$  but  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_{s,q}$  at  $t_{d,q}$ ; 2) the state measurement is outside  $\Omega_{\hat{\rho}_q}$  at  $t_{d,q}$  but still  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{s,q}$ ; or 3) both  $x(t_k) \notin \Omega_{\hat{\rho}_q}$  and  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_{s,q}$  at  $t_{d,q}$ . In any of these cases, the upper bound on  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)|$  is changed to  $\nu_{l,q}$  and re-checked, and the worst-case value of  $\hat{V}_q(x(t_{d,q}))$  is determined from Eq. 60 and a similar procedure to that in Eq. 75 by assuming that the model change can occur at  $t_{k-1}$  with  $\hat{V}_q(\bar{x}_{a,i}(t_{k-1})) = \hat{\rho}_{samp,q}$  such that an equation similar to that in Eq. 75 by using the  $i+1$  model holds for the entire sampling period (i.e.,  $\hat{V}_q$  is increasing for the entire sampling period according to Eq. 60), giving

$$\hat{V}_q(\bar{x}_{a,i+1}(t)) \leq \bar{e}'_{w,q,i+1}/\Delta \quad (79)$$

which gives that  $\hat{V}_q(\bar{x}_{a,i+1}(t)) \leq \hat{V}_q(\bar{x}_{a,i}(t_{k-1})) + \frac{\bar{e}'_{w,q,i+1}(t-t_{k-1})}{\Delta}$ ,  $\forall t \in [t_{k-1}, t_k]$ . Because there is no detection of the model change before  $t_{d,q}$ ,  $\hat{V}_q(x(t_{d,q} - \Delta)) \leq \hat{\rho}_q$ . From Proposition 2:

$$\begin{aligned} \hat{V}_q(\bar{x}_{a,i+1}(t_{d,q} - \Delta)) &\leq \hat{V}_q(x(t_{d,q} - \Delta)) + f_{V,q}(|\bar{x}_{a,i+1}(t_{d,q} - \Delta) - x(t_{d,q} - \Delta)|) \\ &\leq \hat{\rho}_q + f_{V,q}(\theta_v) \end{aligned} \quad (80)$$

This gives that the worst-case value of  $\hat{V}_q(\bar{x}_{a,i+1}(t_{d,q}))$  is  $\hat{\rho}_{samp2,i+1,q} := \hat{\rho}_q + f_{V,q}(\theta_v) + \bar{e}'_{w,q,i+1}$ . In this case, Eq. 79 continues to hold even under  $h_{NL,q}$  (which is triggered to be used after  $t_{d,q}$  according to the implementation strategy in Section 3.2.2) such that there are  $\text{floor}(\frac{\hat{\rho}_{safe,q} - \hat{\rho}_{samp2,i+1,q}}{\bar{e}'_{w,q,i+1}})$  sampling periods before the closed-loop state leaves  $\Omega_{\hat{\rho}_{safe,q}}$  as required.

Second, we must demonstrate that if there is no attack and the attack detection strategy is updated at  $t_{d,q}$  to become  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{l,q}$ , then no cyberattack will be flagged after the underlying process dynamics changed by determining an upper bound on  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)|$  in the absence of an attack and presence of a model change and setting  $\nu_{l,q}$  larger than that bound as

follows:

$$\begin{aligned}
|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| &\leq |\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - \hat{\tilde{x}}_{a,i+1}(t_k|t_{k-1}) + \hat{\tilde{x}}_{a,i+1}(t_k|t_{k-1}) - \bar{x}_{a,i+1}(t_k)| + |\bar{x}_{a,i+1}(t_k) - x(t_k)| \\
&\leq 2\theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}\Delta^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i+1}\theta}{L_{x,i+1}} (e^{L_{x,i+1}\Delta} - 1)
\end{aligned} \tag{81}$$

where the second inequality uses the bound on the measurement noise and Eqs. 33 and 47. When Eq. 66 holds, then  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{l,q}$  at all times after the change in the dynamics is detected so that there are no false detections.

*Part 4.* In this part, we demonstrate that if there is no model change but an undetected attack occurs at  $t_A$ , the closed-loop state is maintained within  $\Omega_{\hat{\rho}_{safe,q}}$  for at least one sampling period after  $t_A$ . If an attack is undetected, one of several cases has occurred: 1)  $x(t_k) \in \Omega_{\hat{\rho}_{safe,q}}/\Omega_{\hat{\rho}_q}$  but  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{s,q}$ ; 2)  $x(t_k) \in \Omega_{\hat{\rho}_q}$  and  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{l,q}$  but  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_{s,q}$ ; 3)  $x(t_k) \in \Omega_{\hat{\rho}_{safe,q}}/\Omega_{\hat{\rho}_q}$  and  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{l,q}$  but  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| > \nu_{s,q}$ ; or 4)  $x(t_k) \in \Omega_{\hat{\rho}_q}$  and  $|\tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)| \leq \nu_{s,q}$ . In each case, however, because there was no model change,  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{samp,q}}$  according to the proof in Part 2 (i.e., no model change and no attack before  $t_k$ ). However, in some of these cases, the implementation strategy of Section 3.2.2 dictates that  $h_{NL,q}$  be used starting at  $t_k$  given the above conditions, and in some of these cases, the LEMPC of Eq. 26 continues to be used.

Propositions 4 and 6 give:

$$\begin{aligned}
|\bar{x}_{a,i}(t_k) - \tilde{\tilde{x}}_{b,q}(t_k|t_{k-1})| &\leq |\bar{x}_{a,i}(t_k) - \hat{\tilde{x}}_{a,i}(t_k|t_{k-1}) + \hat{\tilde{x}}_{a,i}(t_k|t_{k-1}) - \tilde{\tilde{x}}_{b,q}(t_k|t_{k-1})| \\
&\leq \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!}
\end{aligned} \tag{82}$$

regardless of the input used (i.e.,  $h_{NL,q}$  or the LEMPC of Eq. 26 can be used, depending on which is utilized according to the implementation strategy in Section 3.2.2), so that if an attack is not flagged at  $t_k$ :

$$\begin{aligned}
|\bar{x}_{a,i}(t_k) - \tilde{\tilde{x}}_{b,q}(t_k|t_k)| &\leq |\bar{x}_{a,i}(t_k) - \tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) + \tilde{\tilde{x}}_{b,q}(t_k|t_{k-1}) - x(t_k|t_k)| \\
&\leq \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} + \nu_{l,q}
\end{aligned} \tag{83}$$



where  $\nu_{l,q} \geq \nu_{s,q}$  according to the statement of the theorem. From Proposition 2,

$$\begin{aligned} \hat{V}_q(\tilde{x}_{b,q}(t_k|t_k)) &\leq \hat{V}_q(\bar{x}_{a,i}(t_k)) + f_{V,q}(|\bar{x}_{a,i}(t_k) - \tilde{x}_{b,q}(t_k|t_k)|) \\ &\leq \hat{\rho}_{smp,q} + f_{V,q} \left( \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} + \nu_{l,q} \right) \end{aligned} \quad (84)$$

Defining  $M_{\max,q}$  to be the maximum value of  $|f_{NL,q}^n(\tilde{x}_{b,q}, \bar{u}_q)|$  for all  $n = 1, \dots, N_1$ ,  $\tilde{x}_{b,q} \in \Omega_{\hat{\rho}_{safe,q}}$ , and  $\bar{u}_q \in U_q$ :

$$\begin{aligned} \hat{V}_q(\tilde{x}_{b,q}(t_{k+1}|t_k)) &\leq \hat{V}_q(\tilde{x}_{b,q}(t_k|t_k)) + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) \\ &\leq \hat{\rho}_{smp,q} + f_{V,q} \left( \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} + \nu_{l,q} \right) \\ &\quad + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) \end{aligned} \quad (85)$$

$$\begin{aligned} \hat{V}_q(\bar{x}_{a,i}(t_{k+1})) &\leq \hat{V}_q(\tilde{x}_{b,q}(t_{k+1}|t_k)) + f_{V,q}(|\bar{x}_{a,i}(t_{k+1}) - \hat{x}_{a,i}(t_{k+1}|t_k) + \hat{x}_{a,i}(t_{k+1}|t_k) - \tilde{x}_{b,q}(t_{k+1}|t_k)|) \\ &\leq \hat{\rho}_{smp,q} + f_{V,q} \left( \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \theta_v + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} + \nu_{l,q} \right) \\ &\quad + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) + f_{V,q} \left( \delta + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \sum_{n=1}^{N_1} \frac{\Delta^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}\Delta^{N_1+1}}{(N_1+1)!} \right) \end{aligned} \quad (86)$$

When Eq. 68 holds,  $\bar{x}_{a,i}(t_{k+1}) \in \Omega_{\hat{\rho}_{safe,q}}$  though there is an undetected attack at  $t_k$ .

*Part 5.* In this case, we consider that there is both a model change and an undetected attack. The attack may occur first, or the model change may occur first, or both may occur at the same time. However, regardless of which occurs first or the order in which they occur, the worst-case condition is that the closed-loop state at the time at which the attack occurs is as close to the boundary of  $\Omega_{\hat{\rho}_{safe,q}}$  as it can be. When a model change occurs at  $t_{d,q}$  and an attack occurs subsequently, the closed-loop state remains within  $\Omega_{\hat{\rho}_{safe,q}}$  before the attack occurs for  $t_{h,q}$  time units according to the proof of Part 3. However, an attack could then occur at any subsequent sampling time before  $t_{ID,q}$ . In a worst case, it occurs at  $t_{ID,q} - \Delta$  when it may be possible that the closed-loop state is near the boundary of  $\Omega_{\hat{\rho}_{safe,q}}$ . Therefore, it is necessary to ensure that the closed-loop state at  $t_{ID,q} - \Delta$  is within a region from which, even if an undetected cyberattack occurs at that time, the closed-loop

state is still within  $\Omega_{\hat{\rho}_{safe,q}}$  at  $t_{ID,q}$ . From Part 3, the farthest that the state could be at  $t_{ID,q} - \Delta$  is given by:

$$\hat{V}_q(\bar{x}_{a,i+1}(t_{ID,q} - \Delta)) \leq \hat{V}_q(\bar{x}_{a,i+1}(t_{d,q})) + \frac{\bar{e}'_{w,q,i+1}(t_{h,q} - \Delta)}{\Delta} \quad (87)$$

where  $\hat{V}_q(\bar{x}_{a,i+1}(t_{d,q})) \leq \hat{\rho}_{smp2,i+1,q}$  from Part 3. If Eq. 67 holds, then the largest possible value of  $\hat{V}_q(\bar{x}_{a,i+1}(t_{ID,q} - \Delta))$  is  $\hat{\rho}_{far}$ . If  $\bar{x}_{a,i+1}(t_k) \in \Omega_{\hat{\rho}_{far}}$ , then the model change already occurred and if there has not yet been an attack, it is only necessary to demonstrate that  $\bar{x}_{a,i+1}(t_{k+1}) \in \Omega_{\hat{\rho}_{safe,q}}$  if  $\bar{x}_{a,i+1}(t_k) \in \Omega_{\hat{\rho}_{far}}$  and an attack occurs at  $t_k$ . Using similar steps as for Eq. 82 gives:

$$|\bar{x}_{a,i+1}(t_k) - \tilde{x}_{b,q}(t_k|t_{k-1})| \leq \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) \quad (88)$$

regardless of the input used. Then if an attack is not flagged at  $t_k$ , similar steps as in Eqs. 83-86 give:

$$|\bar{x}_{a,i+1}(t_k) - \tilde{x}_{b,q}(t_k|t_k)| \leq \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) + \nu_{l,q} \quad (89)$$

$$\hat{V}_q(\tilde{x}_{b,q}(t_k|t_k)) \leq \hat{\rho}_{far} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) + \nu_{l,q} \right) \quad (90)$$

$$\begin{aligned} \hat{V}_q(\bar{x}_{a,i+1}(t_{k+1})) &\leq \hat{\rho}_{far} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) + \nu_{l,q} \right) \\ &\quad + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) + f_{V,q} \left( \frac{L_{w,i+1}\theta}{L_{x,i+1}}(e^{L_{x,i+1}\Delta} - 1) + \delta \right) \\ &\quad + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} \end{aligned} \quad (91)$$

If Eq. 69 holds, then the closed-loop state is still within  $\Omega_{\hat{\rho}_{safe,q}}$  after a sampling period if an attack occurs after the model change.

If an attack occurs but the model did not change before the attack, the model can change at the same time as the attack or in the sampling period following it. In this case, the worst-case possible value of  $\hat{V}_q(\bar{x}_{a,i}(t_k))$  is  $\hat{\rho}_{smp,q}$  according to the proof of Part 2. We consider first the case that the model and attack both occur for the first time at  $t_k$ . In this case, following similar steps to those

in Eqs. 88-91 and using Eqs. 85 and 89 gives that Eqs. 84 and 85 hold and:

$$\begin{aligned}
\hat{V}_q(\bar{x}_{a,i+1}(t_{k+1})) &\leq \hat{\rho}_{smp,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \nu_{l,q} \right) \\
&\quad + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) + f_{V,q} \left( \frac{L_{w,i+1}\theta}{L_{x,i+1}} (e^{L_{x,i+1}\Delta} - 1) + \delta + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i+1,q} \right. \\
&\quad \left. + \frac{M_{i+1,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} \right)
\end{aligned} \tag{92}$$

If Eq. 70 holds, then  $\bar{x}_{a,i+1}(t_{k+1}) \in \Omega_{\hat{\rho}_{safe,q}}$  after a sampling period if an attack occurs at the same time as the model change.

We consider second the case that the attack occurs first, at  $t_k$ , and the model change occurs at some  $t_{s,i+1} \in [t_k, t_{k+1})$  (if the model change does not occur in the sampling period following the attack, it does not occur in the timeframe over which we guarantee that the closed-loop state remains in  $\Omega_{\hat{\rho}_{safe,q}}$  after an attack). Again Eq. 83 holds and using Eq. 85 and similar steps to those used in deriving Eq. 89, with Eqs. 84-85 and 92 gives,

$$\begin{aligned}
\hat{V}_q(\bar{x}_{a,i+1}(t_{k+1})) &\leq \hat{V}_q(\tilde{x}_{b,q}(t_{k+1}|t_k)) + f_{V,q}(|\bar{x}_{a,i+1}(t_{k+1}) - \bar{x}_{a,i+1}(t_{s,i+1}) \\
&\quad + \bar{x}_{a,i+1}(t_{s,i+1}) - \hat{x}_{a,i}(t_{s,i+1}|t_k) + \hat{x}_{a,i}(t_{s,i+1}|t_k) - \tilde{x}_{b,q}(t_{s,i+1}|t_k) + \tilde{x}_{b,q}(t_{s,i+1}|t_k) - \tilde{x}_{b,q}(t_{k+1}|t_k)|) \\
&\leq \hat{\rho}_{smp,q} + f_{V,q} \left( \theta_v + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} \right. \\
&\quad \left. + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \nu_{l,q} \right) + f_{V,q} \left( \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right) \\
&\quad + f_{V,q} \left( M_{i+1,0}\Delta + \frac{L_{w,i}\theta}{L_{x,i}} (e^{L_{x,i}\Delta} - 1) + \delta + \sum_{n=1}^{N_1} \frac{(\Delta)^n}{n!} M_{deriv,i,q} + \frac{M_{i,N_1}(\Delta)^{N_1+1}}{(N_1+1)!} \right. \\
&\quad \left. + \sum_{n=1}^{N_1} \left( M_{\max,q} \frac{\Delta^n}{n!} \right) \right)
\end{aligned} \tag{93}$$

If Eq. 71 holds, then  $\bar{x}_{a,i+1}(t_{k+1}) \in \Omega_{\hat{\rho}_{safe,q}}$  after a sampling period if an attack occurs before the model change.  $\square$

**Remark 4.** *As seen in the proof of Theorem 1, the use of the truncated Taylor series model solution in the LEMPC allows impacts of numerical approximations of the solution to the empirical model to be accounted for not only in the closed-loop stability guarantees in the absence of an attack, but also*

in the guarantees that can be made with the proposed detection strategy based on predictions from this approximated model. This could allow tradeoffs between numerical error (and therefore computation time) and stability conditions (e.g., stability region sizes) to be assessed from a verification perspective.

**Remark 5.** *The proposed method is focused on the case when it is not straightforward to immediately re-identify the model after a change in the dynamic model occurs (i.e., some additional data since the model change is needed first). If the model was updated immediately at  $t_{d,q}$ , then it is only important to guarantee that  $\bar{x}_{a,i+1}(t_{d,q})$  is still within  $\Omega_{\hat{\rho}_{saf,q}}$  (i.e., at  $\hat{\rho}_{samp2,i+1,q}$ ) and then to use  $h_{NL,q+1}$  after the model update at  $t_{d,q}$  until the closed-loop state enters  $\Omega_{\hat{\rho}_{q+1}}$ . Even with slowly changing dynamics (i.e.,  $M_{err,i+1,q}$  and  $M_{deriv,i+1,q}$  are small), the condition in Eq. 60 could result in  $\bar{\epsilon}_{w,q,i+1}$  being only slightly positive, at least for a short time after the model change, since it was negative previously (Eq. 59), and the proposed method could still be used to flag when the model has changed sufficiently such that a model update is needed. If it is desired to re-identify the model at  $t_{d,q}$ , the potential increase of the detection bounds for the cyberattack detection method may no longer pose a significant benefit, but breaching of the initial bound could still signify either a model change or a cyberattack. The proposed strategy provides insights into how model changes and cyberattacks are related, and the ways that time-varying process dynamics could impact the benefits of cyberattack detection strategies and potentially allow stealthy cyberattacks to be developed that fly under the radar of detection strategies that are inconclusive regarding whether the detection conditions were breached due to model updates or sensor attacks, providing falsified data that over time is re-coding the controller through model re-identification using falsified data.*

**Remark 6.** *While the numerical approximation used in this paper is the Taylor series approximation method, it can be substituted with other numerical methods for which bounds on the error between the predicted state with the numerical method and the actual trajectory of the system are available to be used in place of Eq. 23. This may be preferable to, for example, suggesting many potential terms that may be in the solution of a differential equation and hoping to find those which provide the best approximation to the solution Brunton et al. (2016a), as that does not guarantee that the*

correct terms are guessed to allow an error bound to be developed as above.

**Remark 7.** *Ultimate boundedness of the closed-loop state in  $\Omega_{\hat{\rho}_{\min,i,q}}$  in the absence of an attack or model change could also be obtained by the techniques in Theorem 1 if the constraint of Eq. 24g is repeatedly applied until the closed-loop state enters a region  $\Omega_{\hat{\rho}_{s,q}}$ , with  $\hat{\rho}_{s,q}$  defined in Eq. 63, given the proof of Part 2 of Theorem 1.*

**Remark 8.** *The works Alanqar et al. (2015a,b) utilize empirical models in LEMPC as well. The major difference is that those works assume the empirical dynamic model is used and do not explicitly account for the manner in which error in finding the solution of that dynamic model impacts the closed-loop stability results outside of the plant/model mismatch associated with modeling error.*

### 3.2.3. Detection Strategy 3: Cyberattack-Resilient Output Feedback LEMPC

Detection Strategy 3 from Oyama and Durand (2020) uses multiple redundant state estimators in a detection strategy that, compared to Detection Strategy 2, has the benefit of guaranteeing that the closed-loop state remains within  $\Omega_{\hat{\rho}_q}$  when there is no model change even if undetected attacks occur, but the disadvantage that the guarantees are made with restrictions on the number of sensors that can be attacked compared to Detection Strategy 2. For Detection Strategy 3, the LEMPC of Eq. 26 no longer uses a state measurement at  $t_k$ , but instead uses one of the redundant state estimates (denoted as  $z_{q,1}$ ), and also switches  $\Omega_{\hat{\rho}_{e,q}}$  with the stability region  $\Omega_{\hat{\rho}_{e,1,q}}$  that corresponds to the subset of  $\Omega_{\hat{\rho}_q}$  used with the 1-th observer. With slight abuse of notation, we will consider that references in this section to the LEMPC of Eq. 26 imply that  $x(t_k)$  in Eq. 26c is replaced by  $z_{q,1}(t_k)$ . Cyberattacks are detected by comparing  $|z_{q,r}(t_k) - z_{q,l}(t_k)|$  with an upper bound  $\epsilon_{\max,q} := \max\{e_{rq}^* + e_{lq}^*\}$ ,  $r = 1, \dots, M$ ,  $l = 1, \dots, M$ . Up to  $M - 1$  state estimates can be impacted by the sensor attack, and the attack is assumed to occur after  $t_{bpq}$ ,  $p = 1, \dots, M$ . If  $|z_{q,r}(t_k) - z_{q,l}(t_k)| \leq \epsilon_{\max,q}$  at a sampling time, the LEMPC of Eq. 24 is used to control the process for the subsequent sampling period.

When the process dynamics are allowed to change over time,  $|z_{q,r}(t_k) - z_{q,l}(t_k)|$  could exceed  $\epsilon_{\max,q}$  either because the process dynamics changed or because a cyberattack occurred on the process sensors (the closed-loop state may also be detected to leave  $\Omega_{\hat{\rho}_q}$  for either reason as well). As for

Detection Strategies 1 and 2, these cases may not be able to be distinguished from the sensor data because the estimates are derived from Eq. 20, which may have been developed based on a process model. This necessitates the need for an updated implementation strategy and value of  $\epsilon_{\max,q}$  for guaranteeing that the closed-loop state remains within  $\Omega_{\hat{\rho}_{safe,q}}$  for a characterizable time period after the closed-loop state leaves  $\Omega_{\hat{\rho}_q}$  or after  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q}$  when it cannot be known whether the cause of the mismatch between the different state estimators arises from an attack or a change in the dynamics. To handle this, we propose two methods that could be used to allow for a known amount of time before the closed-loop state leaves  $\Omega_{\hat{\rho}_{safe,q}}$  after an attack or a change in the underlying dynamics while still allowing model changes to trigger re-identification.

*3.2.3.1. Detection Strategy 3, Method 1: Implementation Strategy* The first method to be explored for Detection Strategy 3 will, similar to the method proposed in Section 3.2.2, utilize two stages of monitoring for cyberattacks and model changes. The first stage will utilize a bound  $\epsilon_{\max,q,s}$  on  $|z_{q,r}(t_k) - z_{q,l}(t_k)|$  designed such that, if there were no model changes, the difference between the two estimated states would signify a cyberattack with certainty according to the method in Oyama and Durand (2020). After  $x(t_{d,q}) \notin \Omega_{\hat{\rho}_q}$  or  $|z_{q,r}(t_{d,q}) - z_{q,l}(t_{d,q})| > \epsilon_{\max,q,s}$  for some  $r = 1, \dots, M$ ,  $l = 1, \dots, M$ , a second bound  $\epsilon_{\max,q,l}$  will be used where, if  $|z_{q,r}(t_k) - z_{q,l}(t_k)| \leq \epsilon_{\max,q,l}$  after a model change but no cyberattack is detected via the updated detection mechanism, the closed-loop state should not leave  $\Omega_{\hat{\rho}_{safe,q}}$  before  $t_{h,q}$  time units pass after  $t_{d,q}$ . The goal of this is to ensure that if  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,s}$  because of a model change or cyberattack, subsequent cyberattacks before  $t_{ID,q}$  cannot cause the closed-loop state to leave  $\Omega_{\hat{\rho}_{safe,q}}$  within a sampling period.

To set  $\epsilon_{\max,q,s}$  and  $\epsilon_{\max,q,l}$ , we note that the bounds in Assumption 2 imply that, as demonstrated in Oyama and Durand (2020), the following holds for  $t < t_{s,i+1}$ :

$$|z_{q,r}(t) - z_{q,l}(t)| \leq \max\{e_{r_q}^* + e_{l_q}^*\} := \epsilon_{\max,q} \quad (94)$$

for all  $r \neq l$ ,  $r = 1, \dots, M$ ,  $l = 1, \dots, M$ , as long as  $t \geq t_{\max} := \max\{t_{b1q}, \dots, t_{bMq}\}$ . However, for  $t \geq t_{s,i+1}$ :

$$|z_{q,r}(t) - z_{q,l}(t)| \leq |z_{q,r}(t) - \bar{x}_{a,i+1}(t) + \bar{x}_{a,i+1}(t) - z_{q,l}(t)| \quad (95)$$

However, despite that the norm of the difference between  $z_{q,r}(t)$  and  $\bar{x}_{a,i}(t)$  is assumed to be within a given bound by Assumption 2, after a model change, the state estimate may become inaccurate. Therefore, we make the following assumption.

**Assumption 3.** *There exists  $e_{p,q,i+1} > 0$  such that when  $|z_{q,p}(t_{s,i+1}) - \bar{x}_{a,i+1}(t_{s,i+1})| \leq e_{pq}^*$ ,  $p = 1, \dots, M$ ,  $|z_{q,p}(t) - \bar{x}_{a,i+1}(t)| \leq e_{p,q,i+1}$  for  $t_{s,i+1} \leq t \leq t_{d,q} + t_{h,q}\Delta$ .*

Using this assumption and Eq. 95, we conclude that for  $t \geq t_{s,i+1}$ :

$$|z_{q,r}(t) - z_{q,l}(t)| \leq \max\{e_{r,q,i+1} + e_{l,q,i+1}\} := \epsilon_{\max,i+1} \quad (96)$$

for  $r = 1, \dots, M$  and  $l = 1, \dots, M$ . From this, if  $\epsilon_{\max,q} \leq \epsilon_{\max,q,s}$  and  $\epsilon_{\max,i+1} \leq \epsilon_{\max,q,l}$ , with  $\epsilon_{\max,q,s} \leq \epsilon_{\max,q,l}$ , then for  $t \in [t_{\max}, t_{s,i+1})$ , there will not be any false alarm with the detection strategy based on the selected value of  $\epsilon_{\max,q,s}$ , and for  $t \in [t_{s,i+1}, t_{d,q} + t_{h,q}\Delta)$ , there will again not be a false alarm whether the process dynamics changed or not.

The implementation strategy in this case follows that in Section 3.2.2 with the difference being that in Step 3, instead of setting  $e_{dif} = |\tilde{\bar{x}}_{b,q}(t_k|t_{k-1}) - x(t_k)|$ , it is set to  $|z_{q,r}(t_k) - z_{q,l}(t_k)|$ , for  $r = 1, \dots, M$  and  $l = 1, \dots, M$ , and  $v_{s,q}$  and  $v_{l,q}$  are replaced by  $\epsilon_{\max,q,s}$  and  $\epsilon_{\max,q,l}$ , respectively.

**Remark 9.** *For ease of presentation, we do not consider impacts of numerical error (e.g., a truncated Taylor series) in solving Eq. 20.*

**3.2.3.2. Detection Strategy 3, Method 1: Stability and Feasibility Analysis** We first present a proposition which bounds the worst-case error between the state estimate and closed-loop state before and after a model change, considering Assumptions 2 and 3.

**Proposition 8.** *Consider the system of Eq. 8 under the implementation strategy of Section 3.2.3 where  $M > 1$  state estimators develop independent estimates of the process state and at least one of these estimators is not impacted by false state measurements being provided to the estimators (and the attacks do not begin until after  $t_{\max}$ ). If a false sensor measurement cyberattack is not flagged at  $t_k$  according to the implementation strategy, then the worst-case difference between  $z_{q,1}$  and the actual state  $\bar{x}_{a,i}(t_k)$  is given by:*

$$|z_{q,1}(t_k) - \bar{x}_{a,i}(t_k)| \leq \epsilon_{M,i,q}^* := \epsilon_{\max,q} + \max\{e_{pq}^*\}, \quad p = 1, \dots, M \quad (97)$$

for  $t_k < t_{s,i+1}$ , and the worst-case difference between  $z_{q,1}$  and the actual state  $\bar{x}_{a,i+1}(t_k)$  is given by:

$$|z_{q,1}(t_k) - \bar{x}_{a,i+1}(t_k)| \leq \epsilon_{M,i+1,q}^* := \epsilon_{\max,i+1} + \max\{e_{p,q,i+1}\}, \quad p = 1, \dots, M \quad (98)$$

for  $t_{s,i+1} \leq t \leq t_{d,q} + t_{h,q}\Delta$ .

*Proof.* The proof consists of three parts. In Part 1, it is demonstrated that the bound in Eq. 97 holds when  $t_k < t_{s,i+1}$  whether or not  $z_{q,1}$  is impacted by a sensor attack. In Part 2, it is demonstrated that if  $t_k \geq t_{s,i+1}$  and  $z_{q,1}$  is not impacted by an attack, then Eq. 98 holds. In Part 3, it is demonstrated that if  $t_k \geq t_{s,i+1}$  and  $z_{q,1}$  is impacted by an attack, then Eq. 98 holds.

Part 1 was proven in Oyama and Durand (2020). Part 2 follows from Eq. 96. Specifically, when  $z_{q,1}$  is not impacted by an attack, Assumption 3 gives:

$$|z_{q,1}(t_k) - \bar{x}_{a,i+1}(t_k)| \leq e_{p,q,i+1} \leq \epsilon_{\max,i+1} + \max\{e_{p,q,i+1}\} := \epsilon_{M,i+1,q}^* \quad (99)$$

for  $p = 1, \dots, M$ , satisfying Eq. 98. Part 3 uses a similar technique to develop the following upper bound on  $z_{q,1}$  when it is experiencing an attack and at least one of the other state estimators (with state estimate denoted by  $z_{q,2}$ ) is not:

$$\begin{aligned} |z_{q,1}(t_k) - \bar{x}_{a,i+1}(t_k)| &= |z_{q,1}(t_k) - z_{q,2}(t_k) + z_{q,2}(t_k) - \bar{x}_{a,i+1}(t_k)| \\ &\leq |z_{q,1}(t_k) - z_{q,2}(t_k)| + |z_{q,2}(t_k) - \bar{x}_{a,i+1}(t_k)| \\ &\leq \epsilon_{\max,i+1} + \max\{e_{p,q,i+1}\} := \epsilon_{M,i+1,q}^*, \quad p = 1, \dots, M \end{aligned} \quad (100)$$

□

The following theorem guarantees that in the presence of bounded measurement noise and disturbances, the implementation strategy of Section 3.2.3: 1) maintains the closed-loop state within  $\Omega_{\hat{\rho}_q}$  before an attack or model change occurs; 2) maintains the closed-loop state in  $\Omega_{\hat{\rho}_{safe,q}}$  for at least one sampling period after an attack; and 3) maintains the closed-loop state in  $\Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q}$  sampling periods after  $t_{d,q}$  if no attack occurs.

**Theorem 2.** *Consider the system of Eq. 8 in closed-loop under the implementation strategy of Section 3.2.3 based on a controller  $h_{NL,q}(\cdot)$  that satisfies Eqs. 15a-15d and 18, as well as the requirements of Proposition 7, and based on an observer and controller pair satisfying Assumptions 1-3 and*



formulated with respect to the  $p = 1$  measurement vector. Let  $\theta \leq \theta^*$ ,  $\theta_{v,p} \leq \theta_{v,p}^*$ ,  $M_{err,i,q} \leq M_{err,i,q}^*$ ,  $\epsilon_{pq} \in (\epsilon_{Lpq}^*, \epsilon_{Upq}^*)$ ,  $\hat{\rho}'_{smp2,i+1,q} = \hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^*) + \bar{\epsilon}'_{w,q,i+1} > 0$ , and  $|z_{q,p}(t_0) - \bar{x}_{a,i}(t_0)| \leq e_{m0pq}$ . Also, let  $\epsilon'_{w,q,i} > 0$ ,  $\epsilon''_{w,q,i} > 0$ ,  $\bar{\epsilon}'_{w,q,i+1} > 0$ ,  $0 < \Delta < \Delta_{ub,q}$ ,  $\Omega_{\hat{\rho}_{smp,q}} \subset \Omega_{\hat{\rho}_q} \subset \Omega_{\hat{\rho}_{safe,q}} \subset X_q$ ,  $\hat{\rho}_q > \hat{\rho}_{1,1,q} > \hat{\rho}_{smp,q} > \hat{\rho}_{e,1,q} > \hat{\rho}_{min,1,i,q} > \hat{\rho}_{s,1,q} > 0$ ,  $\hat{\rho}_{e,1,q} > \hat{\rho}_{min,1,i+1,q} > \hat{\rho}_{s,1,q} > 0$ , and  $\hat{\rho}_{e,1,q} > \hat{\rho}'_{min,q} > \hat{\rho}_{s,1,q}$  satisfy:

$$\hat{\rho}_{e,1,q} + M_{i,0} \max\{\Delta, t_{z1}\} \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q)) \leq \hat{\rho}_q \quad (101)$$

$$\hat{\rho}_{min,1,i,q} := \max\{\hat{V}_q(\bar{x}_{a,i}(t + \Delta)) \quad : \quad \hat{V}_q(\bar{x}_{a,i}(t)) \leq \hat{\rho}_{s,1,q}\} \quad (102)$$

$$\hat{\rho}_{e,1,q} + f_{V,q}(\epsilon_{M,i,q}^* + M_{i,0}\Delta) \leq \hat{\rho}_{smp,q} \quad (103)$$

$$\hat{\rho}_{e,1,q} + f_{V,q}(\epsilon_{M,i,q}^* + M_{i,0}\Delta) + f_{V,q}(\epsilon_{M,i,q}^*) \leq \hat{\rho}_q \quad (104)$$

$$-\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i}(\epsilon_{M,i,q}^* + M_{i,0}\Delta) + L'_{w,i}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q))M_{deriv,i,q} \leq -\epsilon'_{w,q,i}/\Delta \quad (105)$$

$$\hat{\rho}_{smp,q} + f_{V,q}(\epsilon_{M,i,q}^*) \leq \hat{\rho}_q \quad (106)$$

$$-\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i+1}(\epsilon_{M,i+1,q}^* + M_{i+1,0}\Delta) + L'_{w,i+1}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q))M_{deriv,i+1,q} \leq \bar{\epsilon}'_{w,q,i+1}/\Delta \quad (107)$$

$$\hat{\rho}_{smp,q} + f_{V,q}(M_{i,0}, \Delta) \leq \hat{\rho}_{safe,q} \quad (108)$$

$$\hat{\rho}_{far} + \bar{\epsilon}'_{w,q,i+1} \leq \hat{\rho}_{safe,q} \quad (109)$$

$$\hat{\rho}_{e,1,q} + f_{V,q}(M_{i,0}\Delta + \epsilon_{M,i,q}^* + \epsilon_{max,q,s}) \leq \hat{\rho}_{smp,q} \quad (110)$$

$$\hat{\rho}_{e,1,q} + f_{V,q}(M_{i,0}\Delta + \epsilon_{M,i,q}^* + \epsilon_{max,q,s}) + f_{V,q}(\epsilon_{max,q,s} + \epsilon_{M,i,q}^*) \leq \hat{\rho}_q \quad (111)$$

$$-\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i}(M_{i,0}\Delta + \epsilon_{M,i,q}^* + \epsilon_{max,q,s}) + L'_{w,i}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q))M_{deriv,i,q} \leq -\epsilon''_{w,q,i}/\Delta \quad (112)$$

$$\hat{\rho}_{smp,q} + f_{V,q}(\epsilon_{max,q,s} + \epsilon_{M,i,q}^*) \leq \hat{\rho}_q \quad (113)$$

where  $t_{z1}$  is the first sampling time after  $t_{b1q}$ ,  $\epsilon_{max,q} \leq \epsilon_{max,q,s}$  and  $\epsilon_{max,i+1} \leq \epsilon_{max,q,l}$ , with  $\epsilon_{max,q,s} \leq \epsilon_{max,q,l}$ , and  $\epsilon_{M,i,q}^* \leq \epsilon_{M,i,q+1}^*$ . Then, if  $\bar{x}_{a,i}(t_0) \in \Omega_{\hat{\rho}_{e,1,q}}$  and  $z_{q,1}(t_0) \in \Omega_{\hat{\rho}_{e,1,q}}$ , then  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{smp,q}}$  and  $z_{q,1}(t) \in \Omega_{\hat{\rho}_q}$  before an attack or a change in the model occur if  $t_{max} < t_{s,i+1}$  and  $t_{max} < t_A$ . Furthermore,  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{safe,q}}$  for at least one sampling period after  $t_A$  and  $\bar{x}_{a,i+1}(t) \in \Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q} = \text{floor}\left(\frac{\hat{\rho}_{safe,q} - \hat{\rho}'_{smp2,i+1,q}}{\bar{\epsilon}'_{w,q,i+1}}\right)$  sampling periods after  $t_{d,q}$  if  $t_A > t_{ID,q}$ .

*Proof.* The proof consists of six parts. In the first part, recursive feasibility at every sampling time in which the LEMPC of Eq. 26 is used under the implementation strategy is demonstrated. In the second part, it is demonstrated that the closed-loop state trajectory is contained in  $\Omega_{\hat{\rho}_q}$  for  $t \in [t_0, \max\{\Delta, t_{b1q}\})$ . In the third part, it is shown that the closed-loop state is maintained in  $\Omega_{\hat{\rho}_{smp,q}}$  and the state measurement is maintained within  $\Omega_{\hat{\rho}_q}$  before any model change or cyberattack occurs. In the fourth part, it is demonstrated that after  $t_{d,q}$ , if only a model change occurs that is detected via either the closed-loop state measurement leaving  $\Omega_{\hat{\rho}_q}$  or if it is detected via  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,s}$ ,  $r = 1, \dots, M$ ,  $l = 1, \dots, M$ , the closed-loop state and state measurement then stay within  $\Omega_{\hat{\rho}_{safe,q}}$  for at least  $t_{h,q}$  sampling times and no attack will be flagged by the updated detection mechanism ( $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,l}$ ). In the fifth part, it is demonstrated that after there is no change in the model but there is an undetected attack (whether it occurs when  $|z_{q,r}(t_k) - z_{q,l}(t_k)| < \epsilon_{\max,q,s}$  is checked or if it occurs at  $t_{d,q}$  when  $\epsilon_{\max,q,s} < |z_{q,r}(t_{d,q}) - z_{q,l}(t_{d,q})| \leq \epsilon_{\max,q,l}$ ), there is at least one sampling period before the closed-loop state leaves  $\Omega_{\hat{\rho}_{safe,q}}$ . In the sixth part, it is demonstrated that if there is a change in the dynamic model as well as an undetected cyberattack at  $t_{d,q}$  that lead to either the state measurement being outside  $\Omega_{\hat{\rho}_q}$  at  $t_{d,q}$ ,  $\epsilon_{\max,q,s} < |z_{q,r}(t) - z_{q,l}(t)| \leq \epsilon_{\max,q,l}$  at  $t_{d,q}$ , or both at  $t_{d,q}$ , then the closed-loop state is maintained within  $\Omega_{\hat{\rho}_{safe,q}}$  for at least one sampling period.

*Part 1.*  $h_{NL,q}$  implemented in sample-and-hold is a feasible input policy for the LEMPC of Eq. 26 whenever the LEMPC of Eq. 26 is used according to the implementation strategy in Section 3.2.3 by the same proof as for Part 1 for Theorem 1.

*Part 2.* For  $t \in [t_0, \max\{\Delta, t_{z1}\})$ , when no attack or model change occurs in that time interval as stated in the conditions of the theorem, the steps in Ellis et al. (2014b); Oyama and Durand (2020) for demonstrating boundedness of the closed-loop state in  $\Omega_{\hat{\rho}_q}$  follow. Specifically, integrating the time derivative of the Lyapunov function along the trajectory of Eq. 8 with  $\bar{x}_{a,i}(t_0) \in \Omega_{\hat{\rho}_{e,1,q}}$  and Eqs. 9a and 9c gives:

$$\hat{V}_q(\bar{x}_{a,i}(t)) \leq \hat{\rho}_{e,1,q} + M_{i,0} \max\{\Delta, t_{z1}\} \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q)) \quad (114)$$

for all  $t \in [t_0, \max\{\Delta, t_{z1}\})$ , so that if Eq. 101 is satisfied,  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_q}$  for all  $t \in [t_0, \max\{\Delta, t_{z1}\})$ .

*Part 3.* In this part, we demonstrate that before any attack or change in the underlying dynamics occurs, the closed-loop state is maintained within  $\Omega_{\hat{\rho}_{samp,q}} \subset \Omega_{\hat{\rho}_q}$  and the state estimate is maintained within  $\Omega_{\hat{\rho}_q}$ . In this case, either  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{e,1,q}}$  so that the constraint of Eq. 26f is activated, or  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,1,q}}$  so that the constraint of Eq. 26g is activated.

Consider first the case that  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{e,1,q}}$ . Eq. 26f ensures that  $\tilde{x}_{b,q}(t)$  is maintained within  $\Omega_{\hat{\rho}_{e,1,q}}$  throughout the prediction horizon. Following steps similar to those in Eq. 72 but using Proposition 8 to note that  $|z_{q,1}(t_k) - \bar{x}_{a,i}(t_k)| \leq \epsilon_{M,i,q}^*$  in this case gives:

$$\begin{aligned} \hat{V}_q(\bar{x}_{a,i}(t)) &\leq \hat{V}_q(z_{q,1}(t_k)) + f_{V,q}(|\bar{x}_{a,i}(t) - \bar{x}_{a,i}(t_k) + \bar{x}_{a,i}(t_k) - z_{q,1}(t_k)|) \\ &\leq \hat{\rho}_{e,1,q} + f_{V,q}(\epsilon_{M,i,q}^* + M_{i,0}\Delta) \end{aligned} \quad (115)$$

for  $t \in [t_k, t_{k+1})$  if  $\bar{x}_{a,i}(t)$  and  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{e,1,q}}$ . If Eq. 103 holds, then  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{samp,q}}$  for  $t \in [t_k, t_{k+1})$  when  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{e,1,q}}$ .

To ensure that  $z_{q,1}(t_{k+1}) \in \Omega_{\hat{\rho}_q}$  for  $t \in [t_k, t_{k+1})$ , Eq. 115 and Proposition 2 give:

$$\begin{aligned} \hat{V}_q(z_{q,1}(t_{k+1})) &\leq \hat{V}_q(\bar{x}_{a,i}(t_{k+1})) + f_{V,q}(|z_{q,1}(t_{k+1}) - \bar{x}_{a,i}(t_{k+1})|) \\ &\leq \hat{\rho}_{e,1,q} + f_{V,q}(\epsilon_{M,i,q}^* + M_{i,0}\Delta) + f_{V,q}(\epsilon_{M,i,q}^*) \end{aligned} \quad (116)$$

When Eq. 104 holds, Eq. 116 gives that  $z_{q,1}(t_{k+1}) \in \Omega_{\hat{\rho}_q}$  when  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{e,1,q}}$ .

Next, we evaluate the case that  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}}/\Omega_{\hat{\rho}_{e,1,q}}$  (i.e., Eq. 26g is activated). Using similar steps as in Eqs. 74 and 75 gives:

$$\begin{aligned} \dot{\hat{V}}_q(\bar{x}_{a,i}(t)) &\leq -\hat{\alpha}_{3,q}(|z_{q,1}(t_k)|) + \frac{\partial \hat{V}_q(\bar{x}_{a,i}(t))}{\partial x} \bar{f}_i(\bar{x}_{a,i}(t), \bar{u}_i(t_k), w_i(t)) - \frac{\partial \hat{V}_q(z_{q,1}(t_k))}{\partial x} \bar{f}_i(z_{q,1}(t_k), \bar{u}_i(t_k), 0) \\ &\quad + \frac{\partial \hat{V}_q(z_{q,1}(t_k))}{\partial x} \bar{f}_i(z_{q,1}(t_k), \bar{u}_i(t_k), 0) - \frac{\partial \hat{V}_q(z_{q,1}(t_k))}{\partial x} \bar{f}_{NL,q}(z_{q,1}(t_k), \bar{u}_q(t_k)) \\ &\leq -\hat{\alpha}_{3,q}(|z_{q,1}(t_k)|) + L'_{x,i}|\bar{x}_{a,i}(t) - \bar{x}_{a,i}(t_k) + \bar{x}_{a,i}(t_k) - z_{q,1}(t_k)| + L'_{w,i}\theta + \left| \frac{\partial \hat{V}_q(z_{q,1}(t_k))}{\partial x} \right| M_{deriv,i,q} \\ &\leq -\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i}(\epsilon_{M,i,q}^* + M_{i,0}\Delta) + L'_{w,i}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q)) M_{deriv,i,q} \end{aligned} \quad (117)$$

If the condition of Eq. 105 holds, then:

$$\hat{V}_q(\bar{x}_{a,i}(t)) \leq \hat{V}_q(\bar{x}_{a,i}(t_k)) - \frac{\epsilon'_{w,q,i}(t - t_k)}{\Delta}, \quad t \in [t_k, t_{k+1}) \quad (118)$$

so that when  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,1,q}}$ ,  $\hat{V}_q(\bar{x}_{a,i}(t))$  decreases over the subsequent sampling period. Since  $\hat{\rho}_{e,1,q} > \hat{\rho}_{s,1,q}$ ,  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,1,q}}$  only if  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{s,1,q}}$ .  $\bar{x}_{a,i}(t_k)$  is guaranteed to be within

$\Omega_{\hat{\rho}_{s\text{amp},q}}$  when the conditions of the theorem are satisfied, as demonstrated using a similar proof as in a similar part of Part 2 of the proof of Theorem 1 and with Eqs. 102, 103, 115, and 118 (i.e., regardless of where  $\bar{x}_{a,i}(t_k)$  is located in  $\Omega_{\hat{\rho}_{s\text{amp},q}}$  at  $t_k$ , since  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,1,q}}$  or  $\Omega_{\hat{\rho}_{e,1,q}} \supset \Omega_{\hat{\rho}_{\min,1,i,q}}$ ,  $\bar{x}_{a,i}(t_k)$  either stays within  $\Omega_{\hat{\rho}_{s\text{amp},q}}$  in the former case by Eqs. 105 and 117, or it remains within  $\Omega_{\hat{\rho}_{s\text{amp},q}}$  by Eqs. 102, 103, and 116). Demonstrating that  $z_{q,1}(t_{k+1}) \in \Omega_{\hat{\rho}_q}$  when  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{s\text{amp},q}}$  uses similar steps as in Eq. 77 with  $z_{q,1}(t)$  replacing  $x(t_{k+1})$  when Eq. 106 holds. Furthermore, Eq. 94 demonstrates that when there is no attack or model change, the condition  $|z_{q,r}(t) - z_{q,l}(t)| \leq \epsilon_{\max,q,s}$ ,  $r = 1, \dots, M$ ,  $l = 1, \dots, M$ , always holds when  $\epsilon_{\max,q} \leq \epsilon_{\max,q,s}$  (i.e., there will be no false alarms with this detection threshold).

*Part 4.* In this part, we demonstrate that after a model change occurs, if there is no attack, the closed-loop state stays in  $\Omega_{\hat{\rho}_{s\text{afe},q}}$  for at least  $t_{h,q}$  sampling periods after  $t_{d,q}$  and no attack is detected after  $t_{d,q}$  (i.e., there are no false alarms under the proposed implementation strategy).

Until  $t_{s,i+1}$ ,  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{s\text{amp},q}}$  and  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}$  under the implementation strategy of Section 3.2.3 as proven in Part 3. After  $t_{s,i+1}$  and at  $t_{d,q}$ , either: 1)  $z_{q,1}(t_{d,q}) \in \Omega_{\hat{\rho}_q}$  but  $|z_{q,r}(t_{d,q}) - z_{q,l}(t_{d,q})| > \epsilon_{\max,q,s}$  at  $t_{d,q}$  for some  $r, l \in \{1, \dots, M\}$ ; 2)  $z_{q,1} \notin \Omega_{\hat{\rho}_q}$  at  $t_{d,q}$  but still  $|z_{q,r}(t_{d,q}) - z_{q,l}(t_{d,q})| \leq \epsilon_{\max,q,s}$  for all  $r, l \in \{1, \dots, M\}$ ; or 3) both  $z_{q,1}(t_{d,q}) \notin \Omega_{\hat{\rho}_q}$  and  $|z_{q,r}(t_{d,q}) - z_{q,l}(t_{d,q})| > \epsilon_{\max,q,s}$  for some  $r, l \in \{1, \dots, M\}$ . In any of these cases, the upper bound on  $|z_{q,r}(t) - z_{q,l}(t)|$  is changed to  $\epsilon_{\max,q,l}$  and re-checked, and the worst-case value of  $\hat{V}_q(\bar{x}_{a,i+1}(t_{d,q}))$  is determined from Eq. 107 and Eq. 117 formulated using the  $i+1$  model. From Proposition 2:

$$\begin{aligned} \hat{V}_q(\bar{x}_{a,i+1}(t_{d,q} - \Delta)) &\leq \hat{V}_q(z_{q,1}(t_{d,q} - \Delta)) + f_{V,q}(|\bar{x}_{a,i+1}(t_{d,q} - \Delta) - z_{q,1}(t_{d,q} - \Delta)|) \\ &\leq \hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^*) \end{aligned} \quad (119)$$

From the integration of Eq. 79, the worst-case value of  $\hat{V}_q(\bar{x}_{a,i+1}(t_k))$  is given by  $\hat{\rho}'_{s\text{amp}2,i+1,q} = \hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^*) + \bar{e}'_{w,q,i+1}$ , from which it can be derived from Eq. 79 that there are  $\text{floor}\left(\frac{\hat{\rho}_{s\text{afe},q} - \hat{\rho}'_{s\text{amp}2,i+1,q}}{\bar{e}'_{w,q,i+1}}\right)$  sampling periods before the closed-loop state leaves  $\Omega_{\hat{\rho}_{s\text{afe},q}}$ , as required, following  $t_{d,q}$ . Finally, Eqs. 94 and 96 demonstrate that if there is no attack and the attack detection strategy is updated at  $t_{d,q}$  to become  $|z_{q,r}(t) - z_{q,l}(t)| \leq \epsilon_{\max,q,l}$ , then no cyberattack will be flagged after the underlying process dynamics change so that there are no false detections.

*Part 5.* If there is no model change but an undetected attack occurs at  $t_A$ , if the attack causes

neither  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,s}$  nor  $z_{q,1}(t_k) \notin \Omega_{\hat{\rho}_q}$ , then from Proposition 2:

$$\begin{aligned} \hat{V}_q(\bar{x}_{a,i}(t)) &\leq \hat{V}_q(z_{q,1}(t_k)) + f_{V,q}(|\bar{x}_{a,i}(t) - \bar{x}_{a,i}(t_k) + \bar{x}_{a,i}(t_k) - z_{q,2}(t_k) + z_{q,2}(t_k) - z_{q,1}(t_k)|) \\ &\leq \hat{\rho}_{e,1,q} + f_{V,q}(M_{i,0}\Delta + \epsilon_{M,i,q}^* + \epsilon_{\max,q,s}) \end{aligned} \quad (120)$$

for  $t \in [t_k, t_{k+1})$ , where  $z_{q,2}$  represents a state estimator that is not impacted by the attack, if  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{e,1,q}}$ , leading to:

$$\begin{aligned} \hat{V}_q(z_{q,1}(t_{k+1})) &\leq \hat{V}_q(\bar{x}_{a,i}(t_{k+1})) + f_{V,q}(|z_{q,1}(t_{k+1}) - z_{q,2}(t_{k+1}) + z_{q,2}(t_{k+1}) - \bar{x}_{a,i}(t_{k+1})|) \\ &\leq \hat{\rho}_{e,1,q} + f_{V,q}(M_{i,0}\Delta + \epsilon_{M,i,q}^* + \epsilon_{\max,q,s}) + f_{V,q}(\epsilon_{\max,q,s} + \epsilon_{M,i,q}^*) \end{aligned} \quad (121)$$

In contrast, if  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}/\Omega_{\hat{\rho}_{e,1,q}}$ , then through similar steps as in Eq. 117, we obtain:

$$\begin{aligned} \dot{\hat{V}}_q(\bar{x}_{a,i}(t)) &\leq -\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i}|\bar{x}_{a,i}(t) - \bar{x}_{a,i}(t_k) + \bar{x}_{a,i}(t_k) - z_{q,2}(t_k) + z_{q,2}(t_k) - z_{q,1}(t_k)| \\ &\quad + L'_{w,i}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q))M_{deriv,i,q} \\ &\leq -\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i}(M_{i,0}\Delta + \epsilon_{M,i,q}^* + \epsilon_{\max,q,s}) + L'_{w,i}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_q))M_{deriv,i,q} \end{aligned} \quad (122)$$

$$\begin{aligned} \hat{V}_q(z_{q,1}(t_{k+1})) &\leq \hat{V}_q(\bar{x}_{a,i}(t_{k+1})) + f_{V,q}(|z_{q,1}(t_{k+1}) - z_{q,2}(t_{k+1}) + z_{q,2}(t_{k+1}) - \bar{x}_{a,i}(t_{k+1})|) \\ &\leq \hat{V}_q(\bar{x}_{a,i}(t_k)) + f_{V,q}(\epsilon_{\max,q,s} + \epsilon_{M,i,q}^*) \\ &\leq \hat{\rho}_{samp,q} + f_{V,q}(\epsilon_{\max,q,s} + \epsilon_{M,i,q}^*) \end{aligned} \quad (123)$$

If Eqs. 102, and 110-113 hold, this ensures that the closed-loop state stays in  $\Omega_{\hat{\rho}_{samp,q}}$  and the state estimate stays within  $\Omega_{\hat{\rho}_q}$ . In contrast, if the attack occurs at  $t_{d,q}$  but remains undetected (i.e., inconclusively detected) and occurs at  $t_k = t_A$ , one of several cases occurred: 1)  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{safe,q}}/\Omega_{\hat{\rho}_q}$  but  $|z_{q,r}(t_k) - z_{q,l}(t_k)| \leq \epsilon_{\max,q,s}$ ; 2)  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_q}$  and  $|z_{q,r}(t_k) - z_{q,l}(t_k)| \leq \epsilon_{\max,q,l}$  but  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,s}$ ; or 3)  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{safe,q}}/\Omega_{\hat{\rho}_q}$  and  $|z_{q,r}(t_k) - z_{q,l}(t_k)| \leq \epsilon_{\max,q,l}$  but  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,s}$ . In each case, however, because there was no model change,  $\bar{x}_{a,i}(t_k) \in \Omega_{\hat{\rho}_{samp,q}}$  according to the proof in Part 3. However, in some of these cases, the implementation strategy of Section 3.2.3 dictates that  $h_{NL,q}$  be used starting at  $t_k$  given the above conditions, and in some of these cases, the LEMPC of Eq. 26 continues to be used. Using similar steps as in Eqs. 115-116 indicates that when Eq. 108 holds,  $\bar{x}_{a,i}(t_{k+1}) \in \Omega_{\hat{\rho}_{safe,q}}$  though there is an undetected attack at  $t_k$ , demonstrated as follows with Eq. 108:

$$\begin{aligned} \hat{V}_q(\bar{x}_{a,i}(t)) &\leq \hat{V}_q(\bar{x}_{a,i}(t_k)) + f_{V,q}(|\bar{x}_{a,i}(t) - \bar{x}_{a,i}(t_k)|) \\ &\leq \hat{\rho}_{samp,q} + f_{V,q}(M_{i,0}\Delta) \end{aligned} \quad (124)$$

*Part 6.* In this case, we consider that there are both a model change and an undetected attack, where one of those could occur before the other, or both may occur at the same time. From Eqs. 107 and 79, and given  $\hat{\rho}_{far} := \hat{\rho}'_{samp2,i+1,q} + \frac{\bar{\epsilon}'_{w,q,i+1}(t_{h,q}-\Delta)}{\Delta}$  as the largest possible value of  $\hat{V}_q$  at  $\bar{x}_{a,i+1}(t_{ID,q} - \Delta)$ , the worst-case value of  $\hat{V}_q$  evaluated along the state trajectory for  $t \in [t_k, t_{k+1})$  when any attack at  $t_k$  is undetected (whether or not a model change occurs before, after, or at the same time as the attack) occurs when  $\hat{V}_q$  is increasing at its maximal possible rate in  $\Omega_{\hat{\rho}_{safe,q}}$  for an undetected attack given by Eq. 107 with the initial state as far from  $\Omega_{\hat{\rho}_q}$  at that time as it can be, which is given by  $\hat{\rho}_{far} + \bar{\epsilon}'_{w,q,i+1}$ . If Eq. 109 holds, then the closed-loop state is still within  $\Omega_{\hat{\rho}_{safe,q}}$  after a sampling period after an attack occurs, regardless of whether the model change occurs first or not.  $\square$

**Remark 10.** If  $\epsilon_{M,i,q}^* > \epsilon_{M,i,q+1}^*$  when defined as in Proposition 8,  $\epsilon_{M,i,q+1}^*$  can be set to  $\epsilon_{M,i,q}^*$  instead of to the value in Proposition 8.

*3.2.3.3. Detection Strategy 3, Method 2* The second method to be proposed for Detection Strategy 3 takes advantage of the closed-loop stability properties in Part 3 of the proof of Theorem 2 above. Assumption 2 guarantees that the original state estimator, designed using the  $q$ -th empirical model, is able to be used to derive a characterizable upper bound on the worst-case difference between the closed-loop state and state estimate in the presence of an undetected attack before the model update. An updated implementation strategy for the second method for Detection Strategy 3 utilizes the two bounds  $\epsilon_{\max,q,s}$  and  $\epsilon_{\max,q,l}$ , but ensures that the closed-loop state is maintained within  $\Omega_{\hat{\rho}_{safe,q}}$  for  $t_{h,q}$  time periods after  $t_{d,q}$  regardless of whether an undetected attack or a model change occurred at  $t_{d,q}$ . If the conditions of Theorem 2 hold, then until  $t_{d,q}$ , the closed-loop state should be maintained within  $\Omega_{\hat{\rho}_{samp,q}}$  with the state estimate in  $\Omega_{\hat{\rho}_q}$  following the proof of Parts 1-3 of Theorem 2. If a state estimate is received where  $z_{q,1}(t_k) \notin \Omega_{\hat{\rho}_q}$  and/or  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,s}$ , for some  $r = 1, \dots, M$  and  $l = 1, \dots, M$ , then the detection bound for the second condition is changed to  $\epsilon_{\max,q,l}$  and it is checked whether  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,l}$ . If it is, a cyberattack is flagged and backup strategies are employed. In addition, a cyberattack should be flagged if  $z_{q,1}(t_k) \notin \Omega_{\hat{\rho}_{safe,q}}$  before  $t_{ID,q}$ , since this implementation strategy should maintain the process state within  $\Omega_{\hat{\rho}_{safe,samp,q}}$

(a subset of  $\Omega_{\hat{\rho}_{safe,q}}$ ) while maintaining the estimate at  $t_k$  in  $\Omega_{\hat{\rho}_{safe,q}}$  in the presence of an undetected attack after  $t_{d,q}$  and before  $t_{ID,q}$ , as discussed below. When no attack is flagged with this method, then at  $t_{ID,q}$ , instead of using  $h_{NL,q}$ , the LEMPC of Eq. 26 is updated to utilize a subset  $\Omega_{\hat{\rho}_{safe,q,e}}$  of  $\Omega_{\hat{\rho}_{safe,q}}$  ( $\Omega_{\hat{\rho}_q} \subset \Omega_{\hat{\rho}_{safe,q,e}}$ ) in place of  $\Omega_{\hat{\rho}_{e,1,q}}$ .

We note that the only differences between what must hold in this case and what has already been proven in Theorem 2, assuming that the requirements of that theorem hold, are: 1) feasibility of the LEMPC of Eq. 26 after it is updated at  $t_{d,q}$  as described in the prior paragraph must be demonstrated; 2) after  $t_{d,q}$ , if only a cyberattack occurred to cause the first detection bound to be breached, the closed-loop state would not leave  $\Omega_{\hat{\rho}_{safe,q}}$  before  $t_{h,q}$  time units pass; 3) the state estimate is maintained within  $\Omega_{\hat{\rho}_{safe,q}}$  at every sampling time before  $t_{ID,q}$  when there are a model change, an undetected attack, or both; and 4) the proof of Part 6 of Theorem 2 can be extended to reflect that  $t_{h,q}$  sampling periods can be made available after  $t_{d,q}$  before the closed-loop state exits  $\Omega_{\hat{\rho}_{safe,q}}$  under a combined model change and cyberattack. To demonstrate the first point, it is noted that the proof of Part 1 of Theorem 1 holds for the updated LEMPC formulation after  $t_{d,q}$  if the conditions of Theorem 2 are met,  $\Omega_{\hat{\rho}_{safe,q}} \subset X_q$ ,  $\hat{\rho}_q$  is replaced by  $\hat{\rho}_{safe,q}$  in Eq. 49, the Lipschitz requirements (e.g., on  $h_{NL,q}$ , derivatives of  $\hat{V}_q$  along model trajectories, and derivatives of  $\hat{V}_q$  with respect to the states) hold in  $\Omega_{\hat{\rho}_{safe,q}}$ , and if the state estimate always remains within  $\Omega_{\hat{\rho}_{safe,q}}$ , which can be guaranteed under some additional conditions clarified below. To demonstrate the second point, the proof of Part 5 of Theorem 2 can be performed with  $\Omega_{\hat{\rho}_{samp,q}}$  replaced by  $\Omega_{\hat{\rho}_{safe,samp,q}}$  (a subset of  $\Omega_{\hat{\rho}_{safe,q}}$ ),  $\Omega_{\hat{\rho}_q}$  replaced by  $\Omega_{\hat{\rho}_{safe,q}}$ , and  $\Omega_{\hat{\rho}_{safe,q,e}}$  used in place of  $\Omega_{\hat{\rho}_{e,1,q}}$  (where  $\Omega_{\hat{\rho}_q} \subset \Omega_{\hat{\rho}_{safe,q,e}} \subset \Omega_{\hat{\rho}_{safe,q}}$ ) if  $\bar{x}_{a,i}(t_{d,q}) \in \Omega_{\hat{\rho}_{safe,samp,q}}$  and  $z_{q,1}(t_{d,q}) \in \Omega_{\hat{\rho}_{safe,q}}$ . In that case, if the following conditions are added to the requirements of Theorem 2:

$$\hat{\rho}_{safe,q,e} + f_{V,q}(\epsilon_{M,i,q}^* + M_{i,0}\Delta + \epsilon_{\max,q,l}) \leq \hat{\rho}_{safe,samp,q} \quad (125)$$

$$\hat{\rho}_{safe,q,e} + f_{V,q}(\epsilon_{M,i,q}^* + M_{i,0}\Delta + \epsilon_{\max,q,l}) + f_{V,q}(\epsilon_{M,i,q}^* + \epsilon_{\max,q,l}) \leq \hat{\rho}_{safe,q} \quad (126)$$

$$-\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i}(\epsilon_{M,i,q}^* + M_{i,0}\Delta + \epsilon_{\max,q,l}) + L'_{w,i}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_{safe,q}))M_{deriv,i,q} \leq -\epsilon_{w,i,q}'''/\Delta \quad (127)$$

$$\hat{\rho}_{safe,samp,q} + f_{V,q}(\epsilon_{\max,q,l} + \epsilon_{M,i,q}^*) \leq \hat{\rho}_{safe,q} \quad (128)$$

where  $\epsilon'''_{w,i,q} > 0$ , then  $\bar{x}_{a,i}(t) \in \Omega_{\hat{\rho}_{safe,samp,q}}$  and  $z_{q,1}(t_k) \in \Omega_{\hat{\rho}_{safe,q}}$  for all times after  $t_{d,q}$  before  $t_{ID,q}$ . To verify that this holds, we therefore must define the conditions under which  $\bar{x}_{a,i}(t_{d,q}) \in \Omega_{\hat{\rho}_{safe,samp,q}}$  and  $z_{q,1}(t_{d,q}) \in \Omega_{\hat{\rho}_{safe,q}}$ .

Under the assumption that at  $t_k = t_{d,q}$  one of the detection bounds is first breached, then one of several cases have occurred: 1) an undetected attack occurred at  $t_{d,q}$  or in a sampling period before it. In this case, Part 3 of the proof of Theorem 2 indicates that  $z_{q,1}(t_{k-1}) \in \Omega_{\hat{\rho}_q}$ ; 2) a model change at a time prior to  $t_{d,q}$ , but no subsequent cyberattack. In this case, as the detection is first triggered at  $t_{d,q}$ ,  $z_{q,1}(t_{k-1}) \in \Omega_{\hat{\rho}_q}$ ; or 3) both an attack and model change have affected the system prior to or at  $t_{d,q}$ , either one occurring before the other, or both simultaneously. Again the lack of detection at a prior sampling time implies  $z_{q,1}(t_{k-1}) \in \Omega_{\hat{\rho}_q}$ .

In the first case, the cyberattack remains undetected by maintaining  $z_{q,1}(t_{d,q} - \Delta) \in \Omega_{\hat{\rho}_q}$ , and this implies that  $\bar{x}_{a,i}(t_{d,q}) \in \Omega_{\hat{\rho}_{samp,q}}$  according to the proof in Parts 3 and 5 of Theorem 2. In the second case, and assuming  $\epsilon^*_{M,i,q} \leq \epsilon^*_{M,i+1,q}$  so that the case of a model change at or before  $t_{k-1}$  is to be considered, Proposition 2 gives:

$$\hat{V}_q(\bar{x}_{a,i+1}(t_{k-1})) \leq \hat{V}_q(z_{q,1}(t_{k-1})) + f_{V,q}(|\bar{x}_{a,i+1}(t_{k-1}) - z_{q,1}(t_{k-1})|) \leq \hat{\rho}_q + f_{V,q}(\epsilon^*_{M,i+1,q}) \quad (129)$$

If Eq. 107 holds, with similar steps as used for equations Eqs. 107 and 79, the worst-case value of  $\hat{V}_q(\bar{x}_{a,i+1}(t_k))$  in this case is given by  $\hat{\rho}_q + f_{V,q}(\epsilon^*_{M,i+1,q}) + \bar{e}'_{w,q,i+1}$ . If the following condition is satisfied:

$$\hat{\rho}_q + f_{V,q}(\epsilon^*_{M,i+1,q}) + \bar{e}'_{w,q,i+1} \leq \hat{\rho}_{safe,samp,q} \quad (130)$$

then  $\bar{x}_{a,i+1}(t_{d,q}) \in \Omega_{\hat{\rho}_{safe,samp,q}}$ . When there is no attack, it can also be demonstrated that  $z_{q,1}(t_{d,q}) \in \Omega_{\hat{\rho}_{safe,q}}$  by using Eqs. 129 and 130 and Proposition 2 as follows with  $t_k = t_{d,q}$ :

$$\begin{aligned} \hat{V}_q(z_{q,1}(t_k)) &\leq \hat{V}_q(\bar{x}_{a,i+1}(t_k)) + f_{V,q}(|z_{q,1}(t_k) - \bar{x}_{a,i+1}(t_k)|) \\ &\leq \hat{\rho}_q + f_{V,q}(\epsilon^*_{M,i+1,q}) + \bar{e}'_{w,q,i+1} + f_{V,q}(\epsilon^*_{M,i+1,q}) \end{aligned} \quad (131)$$

If

$$\hat{\rho}_q + f_{V,q}(\epsilon^*_{M,i+1,q}) + \bar{e}'_{w,q,i+1} + f_{V,q}(\epsilon^*_{M,i+1,q}) \leq \hat{\rho}_{safe,q} \quad (132)$$

then  $z_{q,1}(t_{d,q}) \in \Omega_{\hat{\rho}_{safe,q}}$ .



In the third case (i.e., a cyberattack and a change in the dynamics of the process both occur), if it is assumed for simplicity of presentation that  $L'_{x,i+1} > L'_{x,i}$ ,  $L'_{w,i+1} > L'_{w,i}$ ,  $M_{i+1,0} > M_{i,0}$ , and  $M_{deriv,i+1,q} > M_{deriv,i,q}$  (to avoid the need to present separately whether the model change occurs before  $t_{k-1}$  or not), Proposition 2 gives:

$$\hat{V}_q(\bar{x}_{a,i+1}(t_{k-1})) \leq \hat{V}_q(z_{q,1}(t_{k-1})) + f_{V,q}(|\bar{x}_{a,i+1}(t_{k-1}) - z_{q,2}(t_{k-1}) + z_{q,2}(t_{k-1}) - z_{q,1}(t_{k-1})|) \leq \hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,s}) \quad (133)$$

If

$$-\hat{\alpha}_{3,q}(\hat{\alpha}_{2,q}^{-1}(\hat{\rho}_{s,1,q})) + L'_{x,i+1}(\epsilon_{M,i+1,q}^* + M_{i+1,0}\Delta + \epsilon_{\max,q,l}) + L'_{w,i+1}\theta + \hat{\alpha}_{4,q}(\hat{\alpha}_{1,q}^{-1}(\hat{\rho}_{safe,q}))M_{deriv,i+1,q} \leq \tilde{\epsilon}_{w,i+1,q}/\Delta \quad (134)$$

which is derived in a similar manner to Eq. 117 but for the  $i + 1$  model and accounting for the possibility of an attack, then the worst-case value of  $\hat{V}(\bar{x}_{a,i+1}(t_{d,q}))$  is given by  $\hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,s}) + \tilde{\epsilon}_{w,i+1,q}$  and the worst-case value  $t_{h,q}$  time units after  $t_{d,q}$  from Eqs. 129, 130, and 79 is  $\hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,s}) + \tilde{\epsilon}_{w,i+1,q}(t_{h,q} + 1)$ . The worst-case value of  $z_{q,1}(t_{d,q} + t_{h,q})$  is given by:

$$\begin{aligned} \hat{V}_q(z_{q,1}(t_{ID,q})) &\leq \hat{V}_q(\bar{x}_{a,i+1}(t_{ID,q})) + f_{V,q}(|z_{q,1}(t_{ID,q}) - z_{q,2}(t_{ID,q}) + z_{q,2}(t_{ID,q}) - \bar{x}_{a,i+1}(t_{ID,q})|) \\ &\leq \hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,s}) + \tilde{\epsilon}_{w,i+1,q}(t_{h,q} + 1) + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,l}) \end{aligned} \quad (135)$$

Therefore, if the following hold:

$$\hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,s}) + \tilde{\epsilon}_{w,i+1,q}(t_{h,q} + 1) \leq \hat{\rho}_{safe,samp,q} \quad (136)$$

$$\hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,s}) + \tilde{\epsilon}_{w,i+1,q}(t_{h,q} + 1) + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,l}) \leq \hat{\rho}_{safe,q} \quad (137)$$

then the state estimate at each sampling time and the closed-loop state trajectory are maintained from  $t_{d,q}$  to  $t_{ID,q}$  in  $\Omega_{\hat{\rho}_{safe,q}}$  and  $\Omega_{\hat{\rho}_{safe,samp,q}}$ , respectively, even in the presence of a model change, undetected cyberattack, or both, where  $t_{h,q} = \text{floor}\left(\frac{\hat{\rho}_{safe,q} - (\hat{\rho}_q + f_{V,q}(\epsilon_{M,i+1,q}^* + \epsilon_{\max,q,s}) + \tilde{\epsilon}_{w,i+1,q})}{\tilde{\epsilon}_{w,i+1,q}}\right)$ .

**Remark 11.** *In Method 2 for Detection Strategy 3, because the estimation error remains bounded after the model change, the attack detection policy makes it certain that if  $|z_{q,r}(t_k) - z_{q,l}(t_k)| > \epsilon_{\max,q,l}$ , for  $r = 1, \dots, M$  or  $z_{1,q}(t_k) \notin \Omega_{\hat{\rho}_{safe,q}}$  (assuming that no further model changes occur until after  $t_{ID,q}$  and the closed-loop state is driven into  $\Omega_{\hat{\rho}_{q+1}}$  using  $h_{NL,q+1}$  after  $t_{ID,q}$ ), an attack that would*

compromise closed-loop stability is flagged with certainty. Thus, despite the consideration in this work that sensor measurement attacks and model changes cannot be distinguished from the sensor data, this method indicates that under certain assumptions, there is no need to distinguish them in order to ensure that the closed-loop state remains in a bounded operating region for some time after the attack or model change. However, a consideration is that if there is a new model change before the closed-loop state enters  $\Omega_{\hat{\rho}_{q+1}}$ , then there may not be much time before the closed-loop state leaves  $\Omega_{\hat{\rho}_{safe,q}}$  (i.e., the system is not robust to a second model update before  $\bar{x}_{a,i+1} \in \Omega_{\hat{\rho}_{q+1}}$ ). This indicates that the time required for the re-identification and for driving the closed-loop state from  $\Omega_{\hat{\rho}_{safe,samp,q}}$  may also be a consideration to investigate during the design of the system to ensure that sufficient time is expected between model updates to prevent loss of closed-loop stability.

**Remark 12.** The proofs presented provide a way to guarantee that the closed-loop state is maintained within  $\Omega_{\hat{\rho}_{safe,q}}$  until  $t_{ID,q}$ , regardless of whether an undetected cyberattack on the sensors caused the closed-loop state measurements to leave  $\Omega_{\hat{\rho}_q}$ , or whether a change in the underlying dynamics caused this. However, for verifying safety at run-time for an autonomous system in the presence of changes in the underlying process dynamics and potentially also undetected cyberattacks on the sensors further requires that the closed-loop state be driven into  $\Omega_{\hat{\rho}_{q+1}}$  and subsequently safely operated within that region after  $t_{ID,q}$  in either the case that a model was re-identified at  $t_{ID,q}$  from accurate process data, or that it was re-identified from corrupted data (i.e., data that has been impacted by an undetected attack on the sensors). To achieve this,  $M_{deriv,i+1,q+1}$  cannot be too large, as reflected from the fact that when the closed-loop state is initialized in  $\Omega_{\hat{\rho}_{e,q+1}}$ , under both Detection Strategy 2 and Detection Strategy 3, that variable would need to be sufficiently small to allow the guarantees of Theorems 1 and 2 to hold. If data used in the model re-identification is potentially falsified, it must be asked what the conditions are under which that data would prevent  $M_{deriv,i+1,q+1}$  from being too large. One way to begin to consider this is to consider that the model to be re-identified is chosen from a set of models, each of which has a bounded prediction error. Sufficient data must be available to distinguish which of these models is most accurate. Furthermore, the data is not arbitrarily bad due to the detection mechanisms utilized (e.g., for Detection Strategies 2 and 3, the data is only being used to re-identify the model if the attack was undetected,

implying that there is a bound on how far off the sensor data is from the actual state). Theoretically, one could evaluate the potential worst-case impact of the falsified data on the model selection by evaluating what models would be selected from the set of possible models for each possible data set with data that could be generated within the bounds allowed by the detection strategies, and then for each of the identified models, evaluate  $M_{deriv,i+1,q+1}$ . Though perhaps computationally this would be difficult, it provides insights into the fact that falsified data does not necessarily correspond to a model that would cause closed-loop stability to be compromised being identified, which is consistent with simulation results presented in Durand (2020a). It also suggests how stealthy attacks could be carried out by an attacker who is aware of the model identification algorithm and could seek to determine whether there are state measurement trajectories that could be provided to the sensors that could keep the closed-loop state in  $\Omega_{\hat{\rho}_{saf,e,q}}$  until  $t_{ID,q}$  but then cause the re-identified model to be insufficient for maintaining closed-loop stability under the LEMPC. However, it may be possible to evaluate whether these stealthy attacks can occur a priori and then to attempt to evaluate how or whether they might be mitigated via the control design.

**Remark 13.** We note that the discussion in Remark 12 provides one of the major motivations for explicitly considering numerical error in the LEMPC formulation in this work, and differentiating it from model error. Specifically, the guarantees regarding model re-identification above describe how far off the new dynamic model right-hand side is from that of the empirical model. However, they do not describe how far off it would be in the presence of numerical error. Though the traditional assumption is that they would not be far off in the presence of numerical error, explicitly accounting for the numerical error as a function of the numerical method used and trading off accuracy with computation time in the design of a safe and cybersecure operating policy for run-time verification provides a comprehensive picture of how changing the various parameters involved in the implementation of the control design should be expected to change the guarantees that can be made. We have assumed throughout, however, that the results returned by the numerical methods can take any value (i.e., we have not accounted for finite precision).

**Remark 14.** From a verification standpoint, tests which would be required to be run for this method

include evaluating many of the parameters which correspond to the next model to be identified (e.g., parameters that depend only on the next model such as those in Eq. 9b, or those which depend on differences between models including the next model such as  $M_{deriv,i+1,q}$ ).

**Remark 15.** *The proofs have considered sudden changes in the underlying dynamics. One could attempt to use this method when a gradual change progressively causes the plant/model mismatch to grow over time. The theoretical results lead to an admittedly conservative Ames et al. (2016) concept for run-time verification. However, the conservatism of this initial approach allows the results in Oyama and Durand (2020) developed specifically for nonlinear systems under Lyapunov-based economic model predictive control to be readily extended to this case.*

#### 4. Process Example Demonstration: LEMPC with a Truncated Taylor Series Model

In Oyama et al., simulation results are presented for a continuous stirred tank reactor (CSTR) under strategies similar to some of those proposed in this work. For example, the process was simulated under different thresholds on the smaller state estimate-based detection bound used for Detection Strategy 3 in the presence of plant/model mismatch or sensor measurement attacks. The conclusion of Oyama et al. based on the simulation studies was that without a theoretical basis for developing the various parameters in the simulation (e.g., appropriate values of the detection bounds and different subsets of  $\Omega_{\hat{\rho}_{safe,q}}$ ), it may be less obvious how to tune all of these parameters appropriately. Thus, the work in Oyama et al. motivated the theoretical results developed in this paper. However, it is outside the scope of the present work to develop algorithms for attempting to obtain control law parameters which meet conditions of the theorems systematically. We therefore leave the determination of algorithms for obtaining control law parameters which meet the requirements in this work to future work, and focus instead in this section on addressing the impact that data-driven modeling and different numerical methods could have on sensor measurement cyberattacks via a process example.

The process example under consideration is a continuous stirred tank reactor (CSTR) in which a second-order reaction  $A \rightarrow B$  occurs. The states are the reactant concentration of  $A$  ( $C_A$ ) and

the temperature ( $T$ ), where the dynamics are given by:

$$\dot{C}_A = \frac{F}{V}(C_{A0} - C_A) - k_0 e^{-\frac{E}{R_g T}} C_A^2 \quad (138)$$

$$\dot{T} = \frac{F}{V}(T_0 - T) - \frac{\Delta H k_0}{\rho_L C_p} e^{-\frac{E}{R_g T}} C_A^2 + \frac{Q}{\rho_L C_p V} \quad (139)$$

Here,  $R_g$  is the ideal gas constant,  $E$  is the activation energy,  $\Delta H$  is the enthalpy of reaction, and  $k_0$  is the pre-exponential constant. The inlet/outlet volumetric flow rate,  $F$ , is considered fixed, as are the liquid density,  $\rho_L$ , heat capacity,  $C_p$ , and liquid volume in the tank,  $V$ . The parameter values are as shown in Table 1.

Parameter	Value	Unit	Parameter	Value	Unit
$V$	1	m <sup>3</sup>	$T_0$	300	K
$C_p$	0.231	kJ/kg·K	$k_0$	$8.46 \times 10^6$	m <sup>3</sup> /h·kmol
$F$	5	m <sup>3</sup> /h	$\rho_L$	1000	kg/m <sup>3</sup>
$E$	$5 \times 10^4$	kJ/kmol	$R_g$	8.314	kJ/kmol·K
$\Delta H$	$-1.15 \times 10^4$	kJ/kmol			

Table 1: Parameters for the CSTR model of Eqs. 138-139

The manipulated inputs are the inlet reactant  $A$  concentration ( $C_{A0}$ , which is bounded as follows:  $0.5 \leq C_{A0} \leq 7.5$  kmol/m<sup>3</sup>) and the rate of heat transferred to the system ( $Q$ , which is bounded as follows:  $-5 \times 10^5 \leq Q \leq 5 \times 10^5$  kJ/h). Vectors of deviation variables for the states and inputs from their steady-state values,  $C_{As} = 1.22$  kmol/m<sup>3</sup>,  $T_s = 438.2$  K,  $C_{A0s} = 4.0$  kmol/m<sup>3</sup>, and  $Q_s = 0$  kJ/h, respectively, are  $x^T = [x_1 \ x_2] = [\bar{C}_A \ \bar{T}]$ , where  $\bar{C}_A = C_A - C_{As}$  and  $\bar{T} = T - T_s$ , and  $u^T = [u_1 \ u_2] = [\bar{C}_{A0} \ \bar{Q}]$ , where  $\bar{C}_{A0} = C_{A0} - C_{A0s}$  and  $\bar{Q} = Q - Q_s$ . An LEMPC is used to control the process with the objective function:

$$\int_{t_k}^{t_{k+N}} [-k_0 e^{-\frac{E}{RT(\tau)}} C_A(\tau)^2] d\tau \quad (140)$$

Lyapunov-based stability constraints are designed using the quadratic function,  $\hat{V}_q = x^T P x$ , where  $P = [1200, 5; 5, 0.1]$ . The Lyapunov-based controller, denoted by  $h_{NL,1}(x) = [h_{NL,1,1}(x) \ h_{NL,1,2}(x)]^T$  has  $h_{NL,1,1}(x) \equiv 0$  kmol/m<sup>3</sup> and  $h_{NL,1,2}$  is governed by Sontag's control law Lin and Sontag (1991):

$$h_{NL,1,2}(x) = \begin{cases} -\frac{L_{\hat{f}} \hat{V}_q + \sqrt{L_{\hat{f}}^2 \hat{V}_q^2 + L_{\hat{g}_2} \hat{V}_q^4}}{L_{\hat{g}_2} \hat{V}_q}, & \text{if } L_{\hat{g}_2} \hat{V}_q \neq 0 \\ 0, & \text{if } L_{\hat{g}_2} \hat{V}_q = 0 \end{cases} \quad (141)$$

$h_{NL,1,2}$  is saturated at the input bounds.  $\tilde{f}$  and  $\tilde{g}$  represent the vector-valued and matrix-valued functions that do not and do multiply the inputs, respectively, in Eqs. 138-139 ( $\tilde{g}_2$  is the second column of  $\tilde{g}$ ).  $L_{\tilde{f}}\hat{V}_q$  and  $L_{\tilde{g}_2}\hat{V}_q$  are Lie derivatives of  $\hat{V}_q$  with respect to  $\tilde{f}$  and  $\tilde{g}_2$ .  $\hat{\rho}_1$  was selected to be 300, with  $\hat{\rho}_{e,1} = 225$ . The process state was initialized from  $x_{init} = [-0.4 \text{ kmol/m}^3 \ 8 \text{ K}]^T$ ,  $N$  is 10, and  $\Delta$  is 0.01 h.

In the first part of this example, we analyze how plant/model mismatch introduced via numerical integration techniques in a controller could impact the results of a state measurement cyberattack on the controller. To do this, we develop two LEMPC formulations: one which numerically integrates the process model of Eqs. 138-139 using the explicit Euler numerical integration method with an integration step of  $10^{-4}$  h, and one which uses a truncated Taylor series model as discussed in this work. We choose to employ three terms in the truncated Taylor series model. Specifically, the truncated Taylor series model employed was:

$$\begin{aligned} \tilde{C}_A(t) = & \tilde{C}_A(t_j) + \left[ \left( \frac{F}{V}(C_{A0}(t_j) - \tilde{C}_A(t_j)) \right) - k_0 e^{-\frac{E}{R_g \tilde{T}(t_j)}} (\tilde{C}_A(t_j))^2 \right] \frac{(t - t_j)}{1!} + \\ & \left[ \left( -\frac{F}{V} - (2k_0 e^{-\frac{E}{R_g \tilde{T}(t_j)}} \tilde{C}_A(t_j)) \right) \frac{dC_A(t_j)}{dt} + \left( -k_0 \frac{E}{R_g (\tilde{T}(t_j))^2} e^{-\frac{E}{R_g \tilde{T}(t_j)}} (\tilde{C}_A(t_j))^2 \right) \frac{dT(t_j)}{dt} \right] \frac{(t - t_j)^2}{2!} \end{aligned} \quad (142)$$

$$\begin{aligned} \tilde{T}(t) = & \tilde{T}(t_j) + \left[ \left( \frac{F}{V}(T_0 - \tilde{T}(t_j)) \right) - \frac{\Delta H k_0}{\rho_L C_p} e^{-\frac{E}{R_g \tilde{T}(t_j)}} (\tilde{C}_A(t_j))^2 + \frac{Q(t_k)}{\rho_L C_p V} \right] \frac{(t - t_j)}{1!} + \\ & \left[ \left( -\frac{2k_0 \Delta H}{\rho_L C_p} e^{-\frac{E}{R_g \tilde{T}(t_j)}} \tilde{C}_A(t_j) \right) \frac{dC_A(t_j)}{dt} + \left( -\frac{F}{V} - \left( \frac{k_0 \Delta H E}{\rho_L C_p R_g (\tilde{T}(t_j))^2} e^{-\frac{E}{R_g \tilde{T}(t_j)}} (\tilde{C}_A(t_j))^2 \right) \right) \frac{dT(t_j)}{dt} \right] \frac{(t - t_j)^2}{2!} \end{aligned} \quad (143)$$

for  $t \in [t_j, t_{j+1})$ ,  $j = k, \dots, k + N - 1$ . To obtain some intuition regarding how close the process model solutions from the Taylor series and explicit Euler methods are to one another for the same inputs, both the explicit Euler numerical method and the truncated Taylor series method were used to numerically integrate the process model from a number of different initial conditions in the stability region and under a number of different inputs for at least the length of a sampling period. The results of these studies indicated that over some time horizons under a constant input profile, the results from the explicit Euler and Taylor series methods were relatively close. For example, Fig. 1 shows the trajectories of the process of Eqs. 138-139 under the steady-state inputs, initialized from  $x_{init} = [-0.4 \text{ kmol/m}^3, 8 \text{ K}]^T$ , and indicates good agreement of the trajectories.

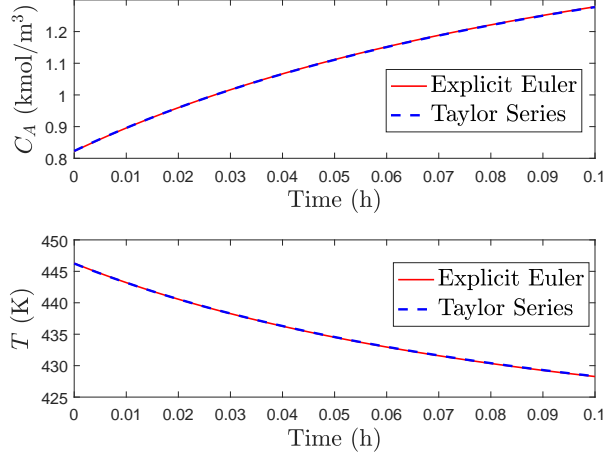


Figure 1: Plot of the state trajectories for  $C_A$  and  $T$  from Eqs. 138-139 determined using explicit Euler (“Explicit Euler”) and Eqs. 142-143 (“Taylor Series”) using the steady-state inputs and initialized from  $x_{init} = [-0.4 \text{ kmol/m}^3, 8 \text{ K}]^T$ .

Furthermore, a number of different initial conditions in state-space were used with various inputs (the initial value of  $C_A$  was varied between 0 and 4  $\text{kmol/m}^3$  in increments of 1  $\text{kmol/m}^3$  and the initial value of  $T$  was varied between 250 and 500 K in increments of 50 K, discarding any points in this discretization that were not in the stability region, and employing an input for each of the  $C_A-T$  combinations from a discretization of the input space in which  $C_{A0}$  was between 0.5  $\text{kmol/m}^3$  and 7.5  $\text{kmol/m}^3$  in increments of 3.5  $\text{kmol/m}^3$  and  $Q$  was between  $-5 \times 10^5 \text{ kJ/h}$  and  $5 \times 10^5 \text{ kJ/h}$  in increments of  $5 \times 10^5 \text{ kJ/h}$ ). These simulations resulted in 18 different scenarios being analyzed for each numerical integration technique. The average integral square error between the trajectories from the two integration methods was evaluated for each initial condition and input combination over 0.01 h as the sum of the squares of the errors between the trajectories at each 0.0001 h, divided by 100 (the number of integration steps in a sampling period). The maximum value of the mean integral square error (scaled by  $10^4$ ) was 0.1269 and the minimum value was  $1.3836 \times 10^{-6}$  among the points evaluated. Even for the case corresponding to the maximum value of the mean integral square error in 0.01 h (which occurred with  $C_{A0} = 0.5 \text{ kmol/m}^3$ ,  $Q = -5 \times 10^5 \text{ kJ/h}$ , and with the initial condition  $T = 450 \text{ K}$ , and  $C_A = 1 \text{ kmol/m}^3$ ), relatively good agreement is shown between the trajectories over 0.1 h, as demonstrated in Fig. 2.

Simulations using the LEMPC with the truncated Taylor series model and with explicit Euler

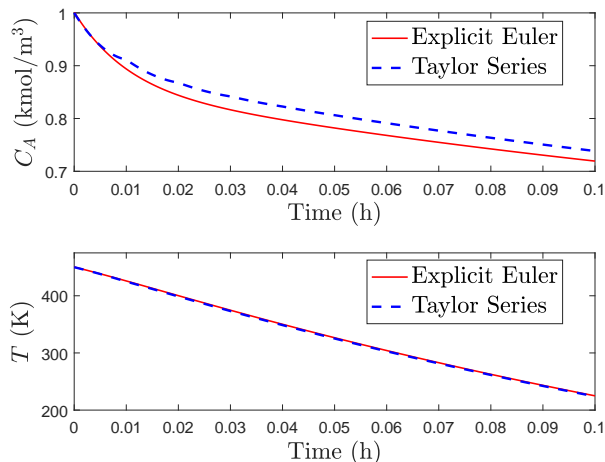


Figure 2: Plot of the state trajectories for  $C_A$  and  $T$  from Eqs. 138-139 determined using explicit Euler (“Explicit Euler”) and Eqs. 142-143 (“Taylor Series”) using the inputs  $C_{A0} = 0.5 \text{ kmol/m}^3$ ,  $Q = -5 \times 10^5 \text{ kJ/h}$ , and with the initial condition  $T = 450 \text{ K}$ , and  $C_A = 1 \text{ kmol/m}^3$ .

were performed for one hour of operation in IPOPT Wächter and Biegler (2006) with ADOL-C Walther and Griewank (2009) using  $C++$  and the code for integrating IPOPT and ADOL-C from Walther (2010), on an Intel(R) Core i7-7500U CPU at 2.70 GHz, 2.90 GHz with 16.0 GB of installed RAM (15.9 GB usable) and a 64-bit operating system with an x64-based processor running Windows 10 Enterprise. The solver indicated that a local minimum was found at each sampling time. The Lyapunov-based stability constraint with the form in Eq. 26f was enforced at the end of each sampling period when  $x(t_k) \in \Omega_{\hat{\rho}_{e,1}}$ , and at the end of each sampling period after the first otherwise. The plant was also simulated using explicit Euler, but with an integration step of  $10^{-5}$  h to introduce minor plant/model mismatch even for the case that the explicit Euler method is used both in the LEMPC and the plant simulation. The input and state trajectories are shown in Figs. 3-4. The somewhat periodic behavior of the states and inputs results from the plant-model mismatch that causes the closed-loop state to leave  $\Omega_{\hat{\rho}_{e,1}}$  when it was not predicted that it would under the computed control action in the LEMPC, resulting in switching between which of the two Lyapunov-based stability constraints is activated at various sampling times. The truncated Taylor series approach and the method with Explicit Euler had similar economic performance (i.e., an integral of the form of the negative of Eq. 140 evaluated over the full hour of operation was 32.516 for the explicit Euler-based LEMPC and 32.606 for the Taylor series method-based LEMPC), but the



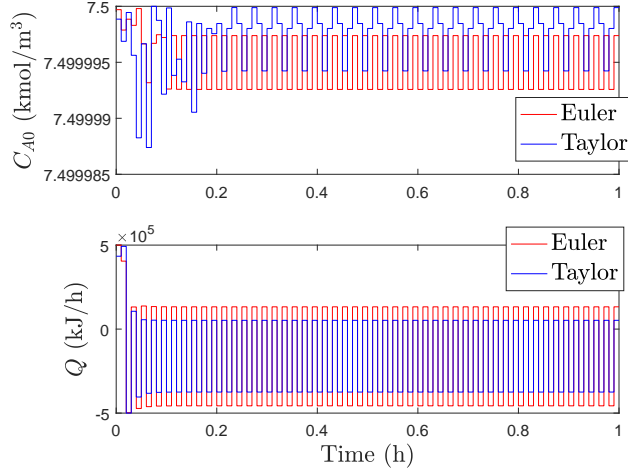


Figure 3: Inputs over one hour of operation under LEMPC’s with the truncated Taylor Series solution (“Taylor”) compared to the formulation using explicit Euler (“Euler”).

LEMPC using the truncated Taylor series did not require the value of the state at every integration step within a sampling period to be computed since the Lyapunov-based stability constraint of Eq. 26f is only enforced at the end of sampling periods.

The results in Fig. 3 indicate that if the trajectories of the state computed using the two different numerical integration techniques are only slightly different (e.g., Figs. 1-2), the two different LEMPC’s might compute similar inputs for the process when presented the same state measurement. If the process is initialized from the same initial condition under two different LEMPC’s, these initially similar input trajectories may result in similar state trajectories for the process, but if the inputs are slightly different, they may over time drive the process state to different conditions from which the state measurements are no longer the same, resulting in the inputs computed potentially being different or driving the state to different conditions over time than would have otherwise been the case (Fig. 4). This implies that if the problem formulation within a model predictive controller is not overly sensitive to slight changes in the process model (e.g., slight changes in the degree of approximation of the model solution introduced by the numerical integration technique used for the dynamic model do not introduce large changes in the optimization problem solution), then a false sensor measurement cyberattack would be likely to cause EMPC’s with different numerical methods used to develop similar inputs to the process. This is demonstrated, for example, by performing a cyberattack involving the false state measurement of  $[\bar{C}_A \ T] = [0.1 \text{ kmol/m}^3 \ -5 \text{ K}]$  on

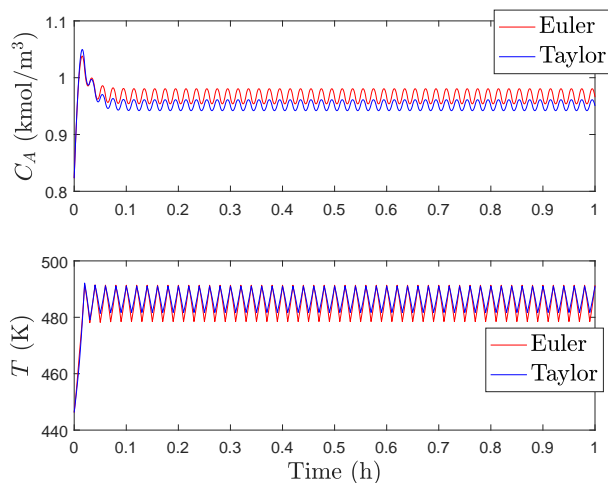


Figure 4: States over one hour of operation under LEMPC’s with the truncated Taylor Series solution (“Taylor”) compared to the formulation using explicit Euler (“Euler”).

both controllers. In this case, both computed approximately the same input to implement on the process (both LEMPC’s selected  $C_{A0} = 7.5 \text{ kmol/m}^3$  and  $Q = 5 \times 10^5 \text{ kJ/h}$  for the first sampling period of the prediction horizon). Thus, both controllers would operate the same process under this attack in approximately the same manner for the subsequent sampling period. This suggests that if an optimization algorithm and problem formulation are not highly sensitive to slight changes in the constraints and objective function, these algorithms and formulations could cause the same state measurement cyberattacks to have similar effects on the process when different levels of plant/model mismatch are present due to numerical integration techniques selected. In addition, the inputs computed for this false state measurement are at the input bounds, demonstrating that even in cases where the process models/their numerical integration method accuracies differ significantly, if the same false state measurement would saturate the inputs in either case, the two different controllers could still operate the process in a similar manner under the same cyberattack.

The above results suggest that if a data-driven model was instead used in an LEMPC, introducing plant/model mismatch due to the imperfections of the model, then if the model is sufficiently close to the model which describes the underlying dynamics, it may be that the use of a data-driven model compared to the use of a first-principles model would not cause the inputs computed by an LEMPC with a data-driven model to be sufficiently different from those which would be computed with a first-principles model even in the face of a sensor measurement cyberattack. To demon-

strate this, we develop an LEMPC that utilizes a data-driven model. To obtain a rough (and not optimized) data-driven model for the purpose of this simulation, we focused on a state-space data-driven modeling strategy to be able to perform a Taylor series approximation of the data-driven model. Inspired by Brunton et al. (2016b), we suggested 78 potential terms for the model (which are listed in Tables 2-3) and performed a regression to determine the coefficients of all of these terms in these tables. The resulting data-driven model was a sum of all of the terms multiplied by their respective coefficients. Ipopt with ADOL-C was used to perform the regression with a data set developed in MATLAB by simulating the process from an initial condition equivalent to  $(C_{As} \text{ kmol/m}^3, T_s + 8 \text{ K})$  under randomly generated inputs from the MATLAB randn function, seeded with the rng function with an argument of 15, and with means of 0, standard deviations of 1 kmol/m<sup>3</sup> and 10<sup>5</sup> kJ/h, and bounds of magnitudes 3.5 kmol/m<sup>3</sup> and  $5 \times 10^5$  kJ/h (in deviation variable form for the inputs) for  $C_{A0}$  and  $Q$  respectively, and simulated for 0.1 h over which the inputs were changed at every 10 integration steps of length  $10^{-4}$  h, and with the initial inputs of 3.5 kmol/m<sup>3</sup> and 1 kJ/h in deviation variable form. The regression was unconstrained besides lower and upper bounds on the decision variables of  $-10^9$  and  $10^9$  respectively (initial guesses of all decision variables were 0), and the objective function was the sum of the squares of the errors in the predictions of all states (with those for concentration weighted by a constant of 100 and those for temperature weighted by a constant of 1). The steady-state inputs for the empirical model for maintaining the closed-loop state at the operating steady-state noted above are 4.0 kmol/m<sup>3</sup> and 102.1 kJ/h. The stability region selected for the original system continued to be selected for the empirical model under Sontag’s control law developed for the empirical system. Though Ipopt returned a solution in this case, no attempt was made to make this model-building strategy robust to noise or to enhance predictive accuracy. Though noise can be present in the process and still be handled with the proposed method as demonstrated theoretically in Section 3.2.3, noise also can impact the data-driven model fidelity (e.g., the value of  $M_{err,i,q}$ ). Because the resulting model provided a sufficient approximation of the actual process dynamics for a sampling period in the absence of noise and therefore serves to adequately demonstrate the concept that an LEMPC using a sufficiently accurate model developed from data may compute similar inputs under a certain sensor

measurement cyberattack as an LEMPC using a first-principles model would, it was decided not to explore enhancing the data-driven modeling strategy to make it robust to noise in the data from which it is identified or to consider noise in the data that could lead to the need for a more robust identification strategy for this example. The resulting data-driven model can be poor at predicting the value of the state under an input for more than a sampling period; therefore, the LEMPC using this data-driven model will employ a prediction horizon on 1. For a single sampling period, the predictions from this model over a sampling period were evaluated in open-loop for a number of different points in state-space (specifically, for 8536 combinations of  $C_A$ ,  $T$ ,  $C_{A0}$ , and  $Q$  in the stability region in state-space, obtained by taking the combinations that result from discretizing  $C_{A0}$  between 0.5 and 7.5 kmol/m<sup>3</sup> in increments of 1 kmol/m<sup>3</sup>,  $Q$  between  $-5 \times 10^5$  and  $5 \times 10^5$  kJ/h in increments of  $10^5$  kJ/h,  $C_A$  from 0 to 4 kmol/m<sup>3</sup> in increments of 0.1 kmol/m<sup>3</sup>, and  $T$  between 250 and 500 K in increments of 10 K, where the initial conditions were in the stability region). The maximum average integral square error (scaled by  $10^4$ ) between the predictions using the actual process dynamics and the empirical process dynamics (both integrated using the explicit Euler numerical method with an integration step of  $10^{-4}$  h from each of the  $C_A - T$  combinations in the discretization) among the points tested was 4.3014. For the case that generated this maximum average integral square error (using the initial condition with  $C_A = 1.4$  kmol/m<sup>3</sup> and  $T = 480$  K, under the inputs  $C_{A0} = 7.5$  kmol/m<sup>3</sup> and  $Q = 5 \times 10^5$  kJ/h), the trajectories of the actual and empirical models over a sampling period are shown in Fig. 5; even with the error in the model for this case, the general trend of the state trajectories with both models appears to be in a similar direction.

To explore the concept in this work of using a Taylor series approximation of a data-driven model, we used three terms in the Taylor series expansion. Whereas the terms in the Taylor series were derived analytically in Eqs. 142-143, with the large number of terms in the data-driven model (78 terms), it was deemed preferable to estimate the derivatives using centered finite differences with the offset in each variable as it is increased or decreased set to  $10^{-5}$ . When three terms are included in the Taylor series expansion as in Eqs. 142-143, the largest value of the average integral mean-square error (scaled by  $10^4$ ) between the empirical model's dynamics integrated using Explicit

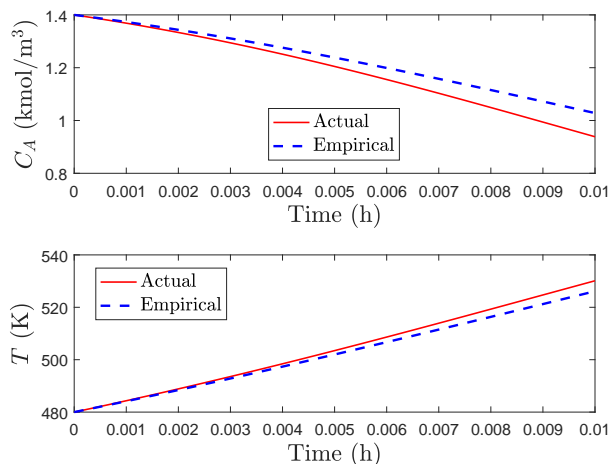


Figure 5: Comparison of state predictions under the data-driven and actual process models integrated using Explicit Euler with an integration step of  $10^{-4}$  h from the initial condition with  $C_A = 1.4$  kmol/m<sup>3</sup> and  $T = 480$  K, under the inputs  $C_{A0} = 7.5$  kmol/m<sup>3</sup> and  $Q = 5 \times 10^5$  kJ/h.

Euler with an integration step of  $10^{-4}$  h and that of the Taylor series for the empirical model was 1.99 (using the same discretization with 8536  $C_A - T$  combinations as mentioned above). This corresponds to the offset over a sampling period depicted in Fig. 6 and was initialized at  $C_A = 1.2$  kmol/m<sup>3</sup> and  $T = 490$  K, under the inputs  $C_{A0} = 0.5$  kmol/m<sup>3</sup> and  $Q = -5 \times 10^5$  kJ/h.

The EMPC formulation was modified for comparing the results of an EMPC using the explicit Euler numerical integration method and using the Taylor series version of the data-driven model. Specifically, in Eq. 140, the economics-based objective function is based on an understanding of the process dynamics and in particular on the knowledge of the reaction rate of the desired product. A challenge is, however, that the model from Tables 2-3 may not allow the reaction rate law to be clearly understood. Therefore, to demonstrate the use of the data-driven model in the EMPC, we will change the stage cost to the following tracking stage cost to avoid the need to assess the reaction rate law:

$$100x_1^2 + x_2^2 + u_1^2 + 10^{-10}u_2^2 \quad (144)$$

With slight abuse of notation,  $u_1$  and  $u_2$  in Eq. 144 represent the inputs in deviation form from the steady-state of the model being used in the EMPC (i.e., either the dynamic model of Eqs. 138-139 or the data-driven model). The constraints were also changed to eliminate those with the form in Eq. 26f and to only enforce that with the form in Eq. 26g at the beginning of a sampling period.

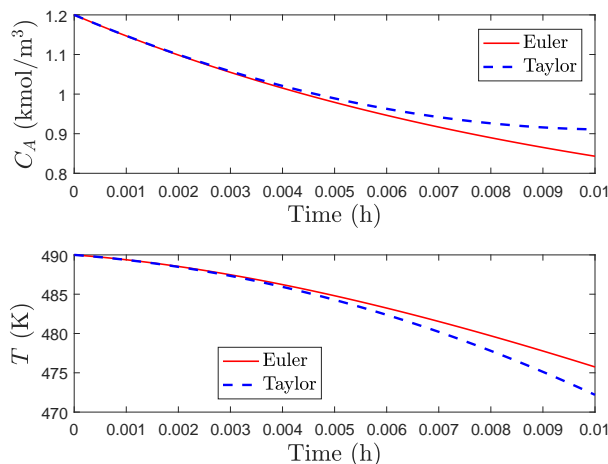


Figure 6: Comparison of state predictions under the data-driven model integrated using Explicit Euler (“Euler”) with an integration step of  $10^{-4}$  h from the initial condition with  $C_A = 1.2$  kmol/m<sup>3</sup> and  $T = 490$  K, under the inputs  $C_{A0} = 0.5$  kmol/m<sup>3</sup> and  $Q = -5 \times 10^5$  kJ/h, against the state predictions from the Taylor series approximation of the data-driven model including three terms (“Taylor”) and under the same inputs, initialized from the same state.

Simulations using the EMPC’s based on the objective function in Eq. 144 were subjected to a cyberattack where the sensor measurement provided to the controller was  $[\bar{C}_A \ \bar{T}] = [0.1 \text{ kmol/m}^3 \ -5 \text{ K}]$  despite that the actual state was  $x_{init}$ , and the inputs from the two different EMPC’s were computed using Ipopt and ADOL-C for a sampling period. The state trajectories under the inputs computed in both cases over the following sampling period are shown in Fig. 7. Even with the different process models and numerical integration techniques used in the two controllers, the inputs computed in both cases are very similar (i.e.,  $C_{A0} = 3.807$  kmol/m<sup>3</sup> and  $Q = 1.3759 \times 10^5$  kJ/h for the explicit Euler numerical integration of Eqs. 138-139, whereas  $C_{A0} = 3.809$  kmol/m<sup>3</sup> and  $Q = 1.3814 \times 10^5$  kJ/h using the Taylor series form of the data-driven model). This indicates that the cyberattacks caused both LEMPC’s to compute similar inputs and therefore to cause the state trajectories in the following sampling period to be relatively similar. If the control laws (e.g., models or numerical integration accuracy) were different enough to find different local optima for the same state measurement, then potentially the same cyberattack could impact both controllers differently. However, these results indicate that an attacker wishing to provide false state measurements to a process may not need to know all of the details of how the code is written (e.g., whether a first-principles or a data-driven model is used, and the method by which the numerical model solution is determined) to plan attacks on the control system.

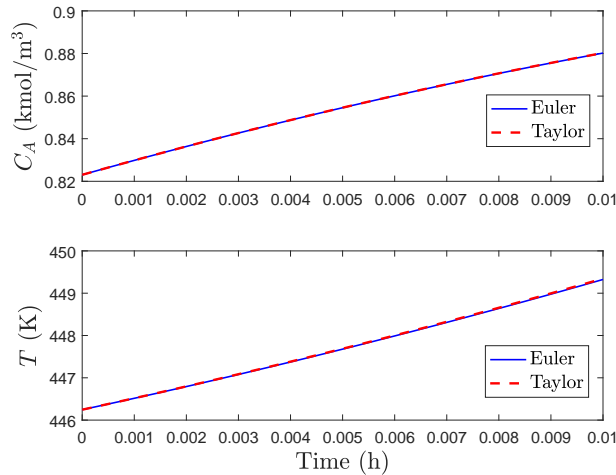


Figure 7: State trajectories over a sampling period of the system of Eqs. 138-139 under the inputs computed by EMPC’s with the stage cost of Eq. 144 and using the explicit Euler numerical integration method for the model of Eqs. 138-139 with an integration step of  $10^{-4}$  h (“Euler”) and the Taylor series form of the data-driven model solution (“Taylor”) initialized from  $x_{init}$  when the controllers were given a false state measurement  $[\bar{C}_A \ \bar{T}] = [0.1 \text{ kmol/m}^3 \ -5 \text{ K}]$ .

## 5. Conclusions

This work considers the fact that despite the desire to be able to verify, at run-time, that an LEMPC maintains safe operation even in the presence of changing dynamics, changes in the process model may violate a notion of cyberattack “discoverability” and thus could be difficult to distinguish from attacks. As a result, detection mechanisms from Oyama and Durand (2020) developed for guaranteeing that the closed-loop state under an LEMPC is maintained within a characterizable region of operation for defined time periods after attacks in the absence of changes in the process dynamics may no longer be guaranteed to do so in the presence of dynamics changes as well. Modified detection strategies with two steps of detection were developed and conditions under which the closed-loop state remains within a characterizable region of operation for a defined time period after either undetected attacks or model changes occur were characterized. However, a challenge with the presented approaches, which we consider a step toward run-time verification of safety with cybersecurity considered as a part of this, is that they may be difficult to utilize practically. Developing techniques for automatically obtaining control law parameters that meet the requirements remains an open direction.

## Acknowledgment

Financial support from the Air Force Office of Scientific Research (award number FA9550-19-1-0059), National Science Foundation CNS-1932026 and CBET-1839675, and Wayne State University is gratefully acknowledged.

## Literature Cited

- Alanqar, A., Durand, H., Christofides, P.D., 2015a. On identification of well-conditioned nonlinear systems: Application to economic model predictive control of nonlinear processes. *AIChE Journal* 61, 3353–3373.
- Alanqar, A., Ellis, M., Christofides, P.D., 2015b. Economic model predictive control of nonlinear process systems using empirical models. *AIChE Journal* 61, 816–830.
- Ames, A.D., Xu, X., Grizzle, J.W., Tabuada, P., 2016. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control* 62, 3861–3876.
- Amin, S., Litrico, X., Sastry, S.S., Bayen, A.M., 2012. Cyber security of water scada systems—part ii: Attack detection using enhanced hydrodynamic models. *IEEE Transactions on Control Systems Technology* 21, 1679–1693.
- Amin, S., Schwartz, G.A., Hussain, A., 2013. In quest of benchmarking security risks to cyber-physical systems. *IEEE Network* 27, 19–24.
- Ani, U.P.D., He, H.M., Tiwari, A., 2017. Review of cybersecurity issues in industrial critical infrastructure: manufacturing in perspective. *Journal of Cyber Security Technology* 1, 32–74.
- Barboni, A., Boem, F., Parisini, T., 2018. Model-based detection of cyber-attacks in networked mpc-based control systems. *IFAC-PapersOnLine* 51, 963–968.
- Brunton, S.L., Proctor, J.L., Kutz, J.N., 2016a. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* 113, 3932–3937.



- Brunton, S.L., Proctor, J.L., Kutz, J.N., 2016b. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* 113, 3932–3937.
- Davis, J., Edgar, T., Graybill, R., Korambath, P., Schott, B., Swink, D., Wang, J., Wetzel, J., 2015. Smart manufacturing. *Annual review of chemical and biomolecular engineering* 6, 141–160.
- Durand, H., 2018. A nonlinear systems framework for cyberattack prevention for chemical process control systems. *Mathematics* 6, 44 pages.
- Durand, H., 2020a. Anomaly-handling in lyapunov-based economic model predictive control via empirical models, in: *Proceedings of the 2020 IFAC World Congress, Berlin, Germany*.
- Durand, H., 2020b. Responsive economic model predictive control for next-generation manufacturing. *Mathematics* 2020 8, 259.
- Durand, H., Wegener, M., 2020. Mitigating safety concerns and profit/production losses for chemical process control systems under cyberattacks via design/control methods. *Mathematics* 8, 499.
- Ellis, M., Durand, H., Christofides, P.D., 2014a. A tutorial review of economic model predictive control methods. *Journal of Process Control* 24, 1156–1178.
- Ellis, M., Zhang, J., Liu, J., Christofides, P.D., 2014b. Robust moving horizon estimation based output feedback economic model predictive control. *Systems & Control Letters* 68, 101–109.
- Giuliani, L., Durand, H., 2018. Data-based nonlinear model identification in economic model predictive control. *Smart and Sustainable Manufacturing Systems* 2, 61–109.
- Heidarinejad, M., Liu, J., Christofides, P.D., 2012. Economic model predictive control of nonlinear process systems using Lyapunov techniques. *AIChE Journal* 58, 855–870.
- Hoehn, A., Zhang, P., 2016. Detection of replay attacks in cyber-physical systems, in: *Proceedings of the American Control Conference, IEEE*. pp. 290–295.
- Khalil, H.K., 2002. *Nonlinear Systems*. Third ed., Prentice Hall, Upper Saddle River, NJ.

- Krantz, S.G., Parks, H.R., 2002. A primer of real analytic functions. Birkhäuser Basel .
- Lao, L., Ellis, M., Durand, H., Christofides, P.D., 2015. Real-time preventive sensor maintenance using robust moving horizon estimation and economic model predictive control. *AICHE Journal* 61, 3374–3389.
- Lezzi, M., Lazoi, M., Corallo, A., 2018. Cybersecurity for industry 4.0 in the current literature: A reference framework. *Computers in Industry* 103, 97–110.
- Li, Y., Voos, H., Rosich, A., Darouach, M., 2015. A stochastic cyber-attack detection scheme for stochastic control systems based on frequency-domain transformation technique, in: *International Conference on Network and System Security*, Springer. pp. 209–222.
- Lin, Y., Sontag, E.D., 1991. A universal formula for stabilization with bounded controls. *Systems & Control Letters* 16, 393–397.
- Liu, S., Wei, G., Song, Y., Liu, Y., 2016. Extended kalman filtering for stochastic nonlinear systems with randomly occurring cyber attacks. *Neurocomputing* 207, 708–716.
- Oyama, H., Durand, H., 2020. Integrated cyberattack detection and resilient control strategies using lyapunov-based economic model predictive control. *AICHE Journal* .
- Oyama, H., Rangan, K.K., Durand, H., . Handling of stealthy sensor and actuator attacks on evolving nonlinear process systems. *Journal of Advanced Manufacturing and Processing* , submitted.
- P. Mhaskar, J.L., Christofides, P.D., 2013. *Fault-Tolerant Process Control: Methods and Applications*. Springer-Verlag, London, England.
- Pasqualetti, F., Dörfler, F., Bullo, F., 2013. Attack detection and identification in cyber-physical systems. *IEEE transactions on automatic control* 58, 2715–2729.
- Qin, S.J., Badgwell, T.A., 2003. A survey of industrial model predictive control technology. *Control Engineering Practice* 11, 733–764.

- Rangan, K.K., Durand, H., 2020. Lyapunov-based economic model predictive with Taylor series state approximations, in: Proceedings of the American Control Conference, Denver, Colorado. pp. 1980–1985.
- Rawlings, J.B., Angeli, D., Bates, C.N., 2012. Fundamentals of economic model predictive control, in: Proceedings of the Conference on Decision and Control, Maui, Hawaii. pp. 3851–3861.
- Ren, A., Wu, D., Zhang, W., Terpenney, J., Liu, P., . Cyber security in smart manufacturing: Survey and challenges.
- Stewart, J., 2003. Calculus. fifth ed., Brooks/Cole-Thomson Learning, Belmont, CA.
- Teixeira, A., Shames, I., Sandberg, H., Johansson, K.H., 2012. Revealing stealthy attacks in control systems, in: 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE. pp. 1806–1813.
- Tuptuk, N., Hailes, S., 2018. Security of smart manufacturing systems. *Journal of Manufacturing Systems* 47, 93–106.
- Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106, 25–57.
- Walther, A., 2010. source:trunk/adol-c/examples/additional\_examples/ipoprt/mittelmannndistcntrlneuma@78 [https://projects.coin-or.org/ADOL-C/browser/trunk/ADOL-C/examples/additional\\_examples/ipoprt/MittelmannDistCntrlNeuma?rev=78](https://projects.coin-or.org/ADOL-C/browser/trunk/ADOL-C/examples/additional_examples/ipoprt/MittelmannDistCntrlNeuma?rev=78).
- Walther, A., Griewank, A., 2009. Getting started with ADOL-C. *Combinatorial Scientific Computing* , 181–202.
- Wu, Z., Albalawi, F., Zhang, J., Zhang, Z., Durand, H., Christofides, P.D., 2018. Detecting and handling cyber-attacks in model predictive control of chemical processes. *Mathematics* 6, 173.
- Xiang, W., Johnson, T.T., 2018. Reachability analysis and safety verification for neural network control systems. arXiv preprint arXiv:1805.09944 .

Xu, X., Tabuada, P., Grizzle, J.W., Ames, A.D., 2015. Robustness of control barrier functions for safety critical control. IFAC-PapersOnLine 48, 54–61.

Coefficient Number	Coefficient Value	Term
1	4.9996	$C_{A0}$
2	-71.4819	$C_A$
3	1873.4916	$e^{-\frac{13000}{R_g T}} (C_A)^2$
4	-1164.1370	$e^{-\frac{16000}{R_g T}} (C_A)^2$
5	-6571.2291	$e^{-\frac{19000}{R_g T}} (C_A)^2$
6	-6171.4487	$e^{-\frac{23000}{R_g T}} (C_A)^2$
7	-1053.5566	$e^{-\frac{26000}{R_g T}} (C_A)^2$
8	-146.256	$e^{-\frac{29000}{R_g T}} (C_A)^2$
9	713.2755	$e^{-\frac{33000}{R_g T}} (C_A)^2$
10	614.1902	$e^{-\frac{36000}{R_g T}} (C_A)^2$
11	428.4638	$e^{-\frac{39000}{R_g T}} (C_A)^2$
12	265.2503	$e^{-\frac{43000}{R_g T}} (C_A)^2$
13	155.5268	$e^{-\frac{46000}{R_g T}} (C_A)^2$
14	94.0699	$e^{-\frac{49000}{R_g T}} (C_A)^2$
15	44.8221	$e^{-\frac{53000}{R_g T}} (C_A)^2$
16	24.6876	$e^{-\frac{56000}{R_g T}} (C_A)^2$
17	13.1956	$e^{-\frac{59000}{R_g T}} (C_A)^2$
18	5.7048	$e^{-\frac{63000}{R_g T}} (C_A)^2$
19	2.9842	$e^{-\frac{66000}{R_g T}} (C_A)^2$
20	1.5408	$e^{-\frac{69000}{R_g T}} (C_A)^2$
21	11237.9225	$e^{-\frac{13000}{R_g T}} C_A$
22	-17294.489	$e^{-\frac{16000}{R_g T}} C_A$
23	-7849.0201	$e^{-\frac{19000}{R_g T}} C_A$
24	1597.6678	$e^{-\frac{23000}{R_g T}} C_A$
25	2547.4864	$e^{-\frac{26000}{R_g T}} C_A$
26	2854.5928	$e^{-\frac{29000}{R_g T}} C_A$
27	1845.8469	$e^{-\frac{33000}{R_g T}} C_A$
28	1134.8829	$e^{-\frac{36000}{R_g T}} C_A$
29	696.6413	$e^{-\frac{39000}{R_g T}} C_A$
30	337.1544	$e^{-\frac{43000}{R_g T}} C_A$
31	188.4947	$e^{-\frac{46000}{R_g T}} C_A$
32	103.4909	$e^{-\frac{49000}{R_g T}} C_A$
33	46.5209	$e^{-\frac{53000}{R_g T}} C_A$
34	24.4744	$e^{-\frac{56000}{R_g T}} C_A$
35	12.8937	$e^{-\frac{59000}{R_g T}} C_A$
36	5.3838	$e^{-\frac{63000}{R_g T}} C_A$
37	2.7827	$e^{-\frac{66000}{R_g T}} C_A$
38	1.4225	$e^{-\frac{69000}{R_g T}} C_A$

Table 2: Coefficients and terms determined for the data-driven model for the terms on the right-hand side of  $\dot{C}_A$  (i.e.,  $\dot{C}_A$  equals the sum of all of the terms in this table multiplied by their respective coefficients).

Coefficient Number	Coefficient Value	Term
39	-223.992	1
40	3275.1834	$C_A$
41	0.0043	$Q$
42	-152969.9125	$e^{-\frac{13000}{R_g T}} (C_A)^2$
43	205377.9155	$e^{-\frac{16000}{R_g T}} (C_A)^2$
44	295222.1403	$e^{-\frac{19000}{R_g T}} (C_A)^2$
45	234061.6455	$e^{-\frac{23000}{R_g T}} (C_A)^2$
46	162756.8662	$e^{-\frac{26000}{R_g T}} (C_A)^2$
47	103695.0183	$e^{-\frac{29000}{R_g T}} (C_A)^2$
48	52141.5254	$e^{-\frac{33000}{R_g T}} (C_A)^2$
49	29747.4961	$e^{-\frac{36000}{R_g T}} (C_A)^2$
50	16492.8407	$e^{-\frac{39000}{R_g T}} (C_A)^2$
51	7255.0603	$e^{-\frac{43000}{R_g T}} (C_A)^2$
52	3838.9721	$e^{-\frac{46000}{R_g T}} (C_A)^2$
53	2002.212	$e^{-\frac{49000}{R_g T}} (C_A)^2$
54	824.9108	$e^{-\frac{53000}{R_g T}} (C_A)^2$
55	419.2063	$e^{-\frac{56000}{R_g T}} (C_A)^2$
56	211.1986	$e^{-\frac{59000}{R_g T}} (C_A)^2$
57	83.6753	$e^{-\frac{63000}{R_g T}} (C_A)^2$
58	41.4646	$e^{-\frac{66000}{R_g T}} (C_A)^2$
59	20.4283	$e^{-\frac{69000}{R_g T}} (C_A)^2$
60	-319541.5717	$e^{-\frac{13000}{R_g T}} C_A$
61	255551.8217	$e^{-\frac{16000}{R_g T}} C_A$
62	360356.6269	$e^{-\frac{19000}{R_g T}} C_A$
63	270756.4013	$e^{-\frac{23000}{R_g T}} C_A$
64	182069.3959	$e^{-\frac{26000}{R_g T}} C_A$
65	112987.0472	$e^{-\frac{29000}{R_g T}} C_A$
66	55245.1489	$e^{-\frac{33000}{R_g T}} C_A$
67	31012.5856	$e^{-\frac{36000}{R_g T}} C_A$
68	16951.9104	$e^{-\frac{39000}{R_g T}} C_A$
69	7344.6616	$e^{-\frac{43000}{R_g T}} C_A$
70	3847.5611	$e^{-\frac{46000}{R_g T}} C_A$
71	1989.7173	$e^{-\frac{49000}{R_g T}} C_A$
72	811.7477	$e^{-\frac{53000}{R_g T}} C_A$
73	409.9647	$e^{-\frac{56000}{R_g T}} C_A$
74	205.36	$e^{-\frac{59000}{R_g T}} C_A$
75	80.8220	$e^{-\frac{63000}{R_g T}} C_A$
76	39.8696	$e^{-\frac{66000}{R_g T}} C_A$
77	19.5618	$e^{-\frac{69000}{R_g T}} C_A$
78	-1.1485	$T$

Table 3: Coefficients and terms determined for the data-driven model for the terms on the right-hand side of  $\dot{T}$  (i.e.,  $\dot{T}$  equals the sum of all of the terms in this table multiplied by their respective coefficients).