Contents lists available at ScienceDirect

Chemical Engineering Research and Design

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

Post cyber-attack state reconstruction for nonlinear processes using machine learning

Zhe Wu[a], Scarlett Chen[a], David Rincon[a], Panagiotis D. Christofides[a,b,*]

[a] Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095-1592, USA
[b] Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095-1592, USA

ARTICLE INFO

Article history:
Received 21 March 2020
Received in revised form 2 April 2020
Accepted 6 April 2020

Keywords:
Cyber-security
State reconstruction
Machine learning
Neural networks
Nonlinear processes
Model predictive control

ABSTRACT

This work proposes state-reconstruction strategies to effectively regain and/or maintain controllability of the system following the detection of cyber-attacks on sensor measurements. Working with a general class of nonlinear systems, of which the sensor measurements may be subject to cyber-attacks, robust control frameworks have been previously proposed to maintain the stability of the process in the presence of cyber-attacks. Moreover, machine-learning-based detection mechanisms could be employed to effectively detect the presence of and distinguish the particular types of cyber-attacks. This work further explores recuperation measures to be taken after the detection of cyber-attacks to mitigate their impact, and proposes a machine-learning-based state reconstruction approach to provide estimated state measurements based on the falsified state measurements. This approach ensures stable operation of the process before reliable sensor measurements are installed back online.

1. Introduction

With the expansion of computer networks and interconnectivity of devices in cyber-physical systems, the digitalization of large-scale industrial systems encourages the use of big data analytics in troubleshooting plant outages, device breakdowns, fault detection, and maintaining cyber-security. The issue of cyber-defense has garnered research interests in recent years, especially when cyber-physical systems are increasingly vulnerable to malicious cyber-attacks targeting potentially every aspect of the exploitable cyber surface. This highlights the importance of implementing effective detection algorithms and having robust response measures in place that safeguards both the cyber and the physical aspects of the system – a key requirement for developing operational technology. Particularly, the process automation and control community has made many contributions in this area of operational technology.

A robust event-triggered model predictive control problem was investigated in Sun and Yang (2019) when the process is subject to bounded disturbances and denial-of-service cyber-attacks. Cumulative Sum (CUSUM) detection method was used in Chamanbaz et al. (2019) in conjunction with model predictive control to operate a nonlinear system under false data injection attacks. Moreover, a robust two-tier control architecture was proposed in Chen et al. (2020b) that provided convenient system reconfiguration strategies to maintain cyber-security.

In the case of attacked systems, countermeasures for malicious attacks have been explored with different strategies in classical techniques like state estimation and observer among other algorithms. In particular, the attacked inputs

* *Corresponding author at*: Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095-1592, USA.
E-mail address: pdc@seas.ucla.edu (P.D. Christofides).
https://doi.org/10.1016/j.cherd.2020.04.018

are reconstructed based on a sliding mode observer and the attacks are identified with a proposed sparse recovery algorithm in Nateghi et al. (2018) by assuming that only a portion of outputs are likely to be attacked. Conditions for detection and reconstruction for a cyber-physical system under cyber-attacks are studied in Ao et al. (2016) by using adaptive sliding mode observers to estimate state attacks and sensor attacks, respectively. Additionally, for state estimation under a set of corrupted measurements, a methodology has been proposed for a class of nonlinear systems with the assumption that the attack in the sensors are sparse in Hu et al. (2017), and then applied to a power system by reconstructing the attacked signal correctly.

Measurement reconstruction has been of interest for many decades in the process fault detection field (Venkatasubramanian et al., 2003; Alcala and Qin, 2010). For example, fault detection and diagnosis have been studied to build a central chilling system with the goal of reconstructing faulty sensors using principal component analysis (PCA) in Wang and Chen (2004). For systems with non-Gaussian distribution, independent component analysis is used for fault detection, identification and reconstruction, which shows better performance than PCA in a metro system in Kim et al. (2013). As a natural extension, nonlinear PCA based on a five-layer neural network has been proposed to reconstruct faulty measurement in Harkat et al. (2007). Multiscale PCA is another extension that combines PCA and wavelet analysis, which permits to reconstruct the signals by means of the inverse wavelet transform Bakshi (1998). Additionally, in order to improve reconstruction, a variance of reconstruction error is proposed for selecting the number of principal components for PCA models Qin and Dunia (2000). Moreover, by proposing a regression-based variable reconstruction method based on a radial basis function network, issues of classical reconstruction techniques like limited number of possible reconstructed variables and reconstructions based on linear projections specially when working with complex process are addressed and proved with simulated and experimental data from a reactive distillation unit in Lieftucht et al. (2009).

However, following the successful detection of cyber-attacks in Wu et al. (2018), Durand (2018), Chen et al. (2020b), the issues of handling the compromised sensor measurements and continuing process control without reliable sensor measurements were not yet addressed. As it is important to develop accurate detectors to promptly report the intrusion of a cyber-attack as well as building robust frameworks to mitigate the impact of cyber-attacks before the detector is activated, it is equally important to have recuperating measures in place to maintain controllability of the system in the absence of reliable sensors.

Machine learning methods have been used in many applications to address the problem of unreliable state measurements. While previous works have proposed effective strategies for the data-based detection of cyber-attacks and resilient operation under cyber-attacks, this work addresses the gap of post-attack handling strategies by (1) proposing a state-reconstruction algorithm using machine-learning methods, (2) considering two advanced control frameworks (Lyapunov-based model predictive control (MPC) and economic MPC) when subject to three types of deception attacks (i.e., min–max, surge, and geometric attacks) on sensor measurements, and (3) applying the proposed strategies to a benchmark nonlinear chemical process example.

2. Preliminaries

2.1. Nonlinear system formulation

In this work, $|\cdot|$ is used to denote the Euclidean norm of a vector; x^T denotes the transpose of x; $\mathrm{R}^{\mathrm{n}}_{+}$ denotes the set of vector functions of dimension n whose domain is $[0, \infty)$. Set subtraction is denoted by "\", i.e., $A\backslash B := \{x \in \mathbf{R}^n | x \in A, x \notin B\}$. Class $\mathcal{K}$ functions $\alpha(\cdot) : [0, a) \to [0, \infty)$ are defined as strictly increasing scalar functions with $\alpha(0) = 0$. The class of continuous-time nonlinear systems considered is described by the following state-space form:

$$\dot{x}(t) = f(x(t), u(t)) \tag{1a}$$

$$\bar{x}(t) = h(x(t)) \tag{1b}$$

where $x(t) \in \mathbf{R}^n$ is the state vector, and $u(t) \in \mathbf{R}^m$ is the manipulated input vector, which is constrained by $u \in U := \{u_i^{\min} \leq u_i \leq u_i^{\max}, i = 1, \ldots, m\} \subset \mathbf{R}^m$, where $u_i^{\min}$ and $u_i^{\max}$ are the lower and upper bounds for the input vector. We will denote the vector of state measurements from sensors, which may be compromised by sparse sensor cyber-attacks, with $\bar{x}(t) \in \mathbf{R}^n$. When no cyber-attacks are present in the system, $\bar{x}(t) = x(t)$. Without loss of generality, the initial time t_0 is taken to be zero ($t_0 = 0$). It is assumed that $f(\cdot)$ is a sufficiently smooth vector function of x and u, and $h(\cdot)$ is a sufficiently smooth vector function of x where $f(0, 0) = 0$, $h(0) = 0$. Therefore, the origin is an equilibrium point of the system of Eq. (1) under $u(t) = 0$.

We assume that there exists an explicit feedback controller of the form $u(t) = \phi(x(t)) \in U$ that can render the origin of the nonlinear closed-loop system of Eq. (1) asymptotically stable. The stabilizability assumption implies the existence of a positive definite control Lyapunov function $V(x)$ that satisfies the following conditions:

$$\alpha_1(|x|) \leq V(x) \leq \alpha_2(|x|), \tag{2a}$$

$$\frac{\partial V(x)}{\partial x} f(x, \phi(x)) \leq -\alpha_3(|x|), \tag{2b}$$

$$\left| \frac{\partial V(x)}{\partial x} \right| \leq \alpha_4(|x|) \tag{2c}$$

for all $x \in D \subseteq \mathbf{R}^n$, where D is an open neighborhood around the origin, and $\alpha_i(\cdot)$, $i = 1, 2, 3, 4$, are class $\mathcal{K}$ functions. The universal Sontag control law (Lin and Sontag, 1991) can be utilized as a candidate controller for $u = \phi(x)$, and will be further saturated to account for the input constraint $u \in U$. Additionally, the Lipschitz property of $f(x, u)$ and the boundedness of $u \in U$ imply that there exist positive constants M, L_x, L'_x such that the following inequalities hold for all x, x' in a neighborhood around the origin:

$$|f(x, u)| \leq M \tag{3a}$$

$$|f(x, u) - f(x', u)| \leq L_x |x - x'| \tag{3b}$$

$$\left| \frac{\partial V(x)}{\partial x} f(x, u) - \frac{\partial V(x')}{\partial x} f(x', u) \right| \leq L'_x |x - x'| \tag{3c}$$

We first characterize a set of states D, in which the time-derivative of the Lyapunov function $V(x)$ under the controller $u = \phi(x) \in U$ satisfies Eq. (2b) for $x \neq 0$. Then we construct a level set of $V(x)$ inside D as $\Omega_\rho := \{x \in D | V(x) \leq \rho, \rho > 0\}$, which represents the stability region of the closed-loop system of Eq. (1). Since Ω_ρ is an invariant set in state-space, it is guaranteed that starting from any initial state $x_0 := x(t_0)$ in Ω_ρ, the state trajectory of the closed-loop system of Eq. (1) remains within Ω_ρ and asymptotically converges to the origin under $u = \phi(x) \in U$. Therefore, the control law $u = \phi(x) \in U$ is able to stabilize the nonlinear system of Eq. (1) at the origin for any initial conditions $x_0 \in \Omega_\rho$ provided that the sensor measurements received by the controller are secure (i.e., $\bar{x}(t) = x(t)$).

3. Previous work on cyber-secure control framework and cyber-attack detection

In this section, we will recap the advanced control schemes that stabilize the nonlinear system of Eq. (1) while optimizing process performance. Then, we will introduce the cyber-secure control framework developed in Wu et al. (2018), Chen et al. (2020b, a), which includes cyber-attack detection using machine learning techniques and resilient secure control that can mitigate the impact of cyber-attacks upon detection.

3.1. *Lyapunov-based model predictive control*

The Lyapunov-based model predictive control (LMPC) was proposed to stabilize the nonlinear system of Eq. (1) at the origin based on secure state measurements. The optimization problem of LMPC is shown as follows:

$$\mathcal{J} = \min_{u \in S(\Delta)} \int_{t_k}^{t_{k+N}} l_t(\tilde{x}(t), u(t)) dt \tag{4a}$$

$$\text{s.t.} \quad \dot{\tilde{x}}(t) = f(\tilde{x}(t), u(t)) \tag{4b}$$

$$\tilde{x}(t_k) = \bar{x}(t_k) \tag{4c}$$

$$u(t) \in U, \quad \forall \ t \in [t_k, t_{k+N}) \tag{4d}$$

$$\dot{V}(\bar{x}(t_k), u(t_k)) \leq \dot{V}(\bar{x}(t_k), \phi(\bar{x}(t_k))), \quad \text{if} \ \ V(\bar{x}(t_k)) > \rho_{\min}, \tag{4e}$$

$$V(\bar{x}(t)) \leq \rho_{\min}, \quad \forall \ t \in [t_k, t_{k+N}), \quad \text{if} \ \ V(\bar{x}(t_k)) \leq \rho_{\min} \tag{4f}$$

where $\tilde{x}(t)$ is the predicted state trajectory based on state measurement $\bar{x}(t_k)$ at $t = t_k$, $S(\Delta)$ is the set of piecewise constant functions with the sampling period Δ, and N is the number of sampling periods in the prediction horizon. $\dot{V}(x(t_k), u(t_k))$ represents the time derivative of $V(x)$, i.e., $\frac{\partial V}{\partial x} f(x(t_k), u(t_k))$. The states of the closed-loop system are measured at the end of each sampling period, and are used as the initial condition for the optimization problem of LMPC in the next sampling step. The control actions calculated by the LMPC of Eq. (4) are implemented in a sample-and-hold fashion in the sense that the optimal solution $u^*(t)$ over the prediction horizon $t \in [t_k, t_{k+N})$ is obtained by solving the optimization problem of Eq. (4) based on the measured state $\bar{x}(t_k)$, and then only the first control action of $u^*(t)$, i.e., $u^*(t_k)$, will be sent to the control actuators to be applied over the next sampling period. At the next sampling time $t_{k+1} := t_k + \Delta$, the LMPC of Eq. (4) will be solved again, and the horizon will be rolled one sampling time forward.

In the optimization problem of Eq. (4), the objective function of Eq. (4a) is to minimize the integral of $l_t(\tilde{x}(t), u(t))$ over the prediction horizon. The function $l_t(x, u)$ is generally in a quadratic form (i.e., $l_t(x, u) = x^T Q_1 x + u^T Q_2 u$, where Q_1 and Q_2 are positive definite matrices) such that the minimum value of $l_t(x, u)$ is achieved at the origin. The constraint of Eq. (4b) is the nonlinear system of Eq. (1) used to predict the evolution of the closed-loop state x over the prediction horizon. The initial condition $\tilde{x}(t_k)$ for the LMPC of Eq. (4) is the state measurement $\bar{x}(t_k)$ at $t = t_k$. The constraint of Eq. (4d) defines the input constraints over the prediction horizon. The constraint of Eqs. (4e) and (4f) are utilized to maintain closed-loop stability. Specifically, the constraint of Eq. (4e) drives the process state towards the origin by decreasing the value of Lyapunov function $V(x)$ at least at the rate under $u = \phi(\bar{x})$ when $V(\bar{x}(t_k)) > \rho_{\min}$. However, if $\bar{x}(t_k)$ enters a small neighborhood around the origin $\Omega_{\rho_{\min}} := \{x \in \phi_n | V(x) \leq \rho_{\min}\}$, the constraint of Eq. (4f) is used to maintain the state inside $\Omega_{\rho_{\min}}$ afterwards.

Under the LMPC of Eq. (4), the process state x is guaranteed to remain in the stability region Ω_ρ for all times, and ultimately converge to $\Omega_{\rho_{\min}}$, for any initial condition $x_0 \in \Omega_\rho$, provided that the true state measurement x is available at each sampling step. However, in the presence of cyber-attacks that compromise state measurements, closed-loop stability is no longer guaranteed as the LMPC of Eq. (4) is solved based on falsified state measurements. Therefore, in Wu et al. (2018), the LMPC switches to secure back-up sensors to re-stabilize the nonlinear system of Eq. (1) at the origin once the cyber-attack is detected. Additionally, in Chen et al. (2020b), a two-tier cyber-secure control architecture in which the lower-tier controllers are used to stabilize the nonlinear system of Eq. (1), and the upper-tier LMPC improves closed-loop performance with networked sensor measurements, was proposed to maintain closed-loop stability by using the lower-tier controller only upon detection.

3.2. *Lyapunov-based economic model predictive control*

To optimize the overall economic gain, an economic model predictive control (EMPC) framework was proposed where the process will be operated in an off-steady-state manner, to replace the real-time optimization combination with steady-state operation driven by tracking controllers. Moreover, a cyber-secure resilient operation mode of the economic model predictive control within a conservative secure region with combined open-loop and closed-loop control actions was proposed in Chen et al. (2020a) to detect and mitigate the impact of cyber-attacks. Specifically, the Lyapunov-based economic model predictive control (LEMPC) scheme is represented by the following optimization problem:

$$\mathcal{J} = \max_{u \in S(\Delta)} \int_{t_k}^{t_{k+N}} l_e(\tilde{x}(t), u(t)) dt \tag{5a}$$

$$\text{s.t.} \quad \dot{\tilde{x}}(t) = f(\tilde{x}(t), u(t)) \tag{5b}$$

$$u(t) \in U, \quad \forall \ t \in [t_k, t_{k+N}) \tag{5c}$$

$$\tilde{x}(t_k) = \bar{x}(t_k) \tag{5d}$$

$$V(\tilde{x}(t)) \leq \rho_e, \quad \forall \ t \in [t_k, t_{k+N}), \quad \text{if} \ \ \bar{x}(t_k) \in \Omega_{\rho_e} \tag{5e}$$

$$\dot{V}(\bar{x}(t_k), u) \leq \dot{V}(\bar{x}(t_k), \phi(\bar{x}(t_k))), \quad \text{if} \ \ \bar{x}(t_k) \in \Omega_\rho \backslash \Omega_{\rho_e} \tag{5f}$$

where the notations follow those in Eq. (4). Unlike the LMPC of Eq. (4) that drives the process state to its steady-state, the LEMPC of Eq. (5) optimizes process economic benefits by dynamically operating the system in the stability region Ω_ρ to achieve economic optimality and closed-loop stability simultaneously. Specifically, the objective function of Eq. (5a) optimizes the time integral of the cost function $l_e(\bar{x}(t), u(t))$ that accounts for process economic benefits over the prediction horizon. The constraints of Eqs. (5b)–(5d) are the same as the constraints of Eqs. (4b)–(4c). The constraint of Eq. (5e) maintains the process state in Ω_{ρ_e} if $\bar{x}(t_k) \in \Omega_{\rho_e}$, where Ω_{ρ_e} is designed to ensure the invariance of the stability region Ω_ρ in the presence of sufficiently small disturbances. If $\bar{x}(t_k)$ leaves Ω_{ρ_e} due to disturbances, the constraint of Eq. (5f) will drive the process state towards the origin, and ultimately into Ω_{ρ_e} within finite sampling periods. Therefore, under the LEMPC of Eq. (5), the process state x is guaranteed to be maintained in Ω_ρ for all times for any $x_0 \in \Omega_\rho$.

However, closed-loop stability cannot be guaranteed when the state measurements are compromised by cyber-attacks (i.e., $\bar{x} \neq x$). To maintain cyber-security, an effective cyber-attack detector and a resilient control strategy were proposed in Chen et al. (2020a). Specifically, upon the successful detection of cyber-attacks in sensors, one strategy is to utilize the response plan proposed in Wu et al. (2018) that involved physical replacements of problematic sensors with their redundant back-up sensors. While sensor device replacement is an effective measure, there may be circumstances where redundant sensors cannot be deployed immediately, during which time the process must be operated in open-loop without reliable feedback measurements (Chen et al., 2020a).

4. Intelligent cyber-attacks and detection

In a cyber-physical system vulnerable to cyber-attacks, sensors, actuators, communication channels between them, as well as the control system itself can all be targeted. These intelligent cyber-attacks intentionally degrade closed-loop performance by disrupting accurate control implementation. As it has been previously considered in relevant works, sensor cyber-attacks will be the area of focus in this work. During feed-back closed-loop control, sensor measurements must accurately reflect the true process state to ensure closed-loop stability. If the measurements received by the controller are false, they will result in incorrect control actions that may result in the true states exiting from the bounded operating region and eventually going outside of the stability region.

Being process and controller behavior aware, the cyber-attacks will have access to information on the operating region of the process, and existing alarms configured on the input and output ranges. Specifically, when attacks intend to induce maximum disruption (i.e., in min–max or surge attacks), the attacked value will be set to the maximum or minimum value beyond which an alarm monitoring the current state measurement will be immediately triggered. These intelligent cyber-attacks are designed such that no alarms will be sounded (i.e., the falsified state measurement is not outside the operating stability region or the alarm window) and the controller is still able to compute feasible control actions, but have large enough variations such that economic optimality and closed-loop stability will be lost.

In this work, we consider three types of standard cyber-attacks discussed in the literature Singh and Nene (2013): min–max, surge, and geometric attacks; all of which have been previously studied in developing effective data-based detection algorithms. The algorithms developed in Wu et al. (2018), Chen et al. (2020a) demonstrated capabilities in detecting min–max, surge, and geometric attacks promptly while operating the process under LMPC and LEMPC. Furthermore, min–max, surge, geometric, and replay attacks can be detected using the neural network detector proposed in Chen et al. (2020b) when the process is operating using a two-tier control architecture.

4.1. *Min–max cyber-attack*

To achieve maximum disruptive impact within shortest amount of time, min–max attacks are designed while avoiding triggering any alarms. The compromised sensor measurements will be set to values that are furthest from the equilibrium, either the minimum or the maximum, but not outside of the normal operating region. Attacks generated will ensure that no conventional detection alarms based on statistical threshold will be triggered. The min–max attack can be formulated as follows:

$$\bar{x}(t_i) = \min / \max_{x \in R^n} \{x | V(x(t_i)) = \rho\}, \quad \forall \ \ i \in [i_0, i_0 + L_a] \tag{6}$$

where ρ is the level set of the Lyapunov function $V(x)$ that represents the operating region of the closed-loop system of Eq. (1) under either LMPC or LEMPC, $\bar{x}$ is the compromised sensor measurement, i_0 is the time instant that the attack is introduced, and L_a is the total duration of the attack in terms of sampling periods.

4.2. *Surge cyber-attack*

Surge attacks are designed such that the cumulative deviation from the true process state values will not exceed the cumulative statistical threshold examined by some conventional detection methods such as CUSUM. Surge attacks achieve this goal by setting the state measurements to the maximum or minimum value to induce maximum destabilizing effects for an initial short period of time, then the falsified measurement will be reduced and maintained at a lower value for the remainder of the attack duration. Therefore, surge attacks act like min–max attacks initially. The length of the initial surge period and the reduced value thereafter can be selected many ways, as long as the cumulative error for the entire attack duration is maintained below the alarm-triggering threshold. In this work, the reduced value after the surge is set to a sufficiently small bounded noise added on the attacked sensor.

The formulation of the surge attack is presented below:

$$\begin{aligned} &\bar{x}(t_i) = \min / \max_{x \in R^n} \{x | V(x(t_i)) = \rho\}, \quad \text{if} \ \ i_0 \le i \le i_0 + L_s \\ &\bar{x}(t_i) = x(t_i) + \eta(t_i), \quad \text{if} \ \ i_0 + L_s < i \le i_0 + L_a \end{aligned} \tag{7}$$

where i_0 is the start time of the attack, L_s is the duration of the initial surge, and $\eta_l \le \eta(t_u) \le \eta_u$ is the bounded noise added on the sensor measurement after the initial surge period, where η_l and η_u are the lower and upper bounds of the noise, respectively.

4.3. Geometric cyber-attack

When sensor measurements are under geometric cyber-attacks, the falsified measurement will deteriorate at a geometric speed until it reaches the limit boundary characterized by the secure operating region, at which an alarm will sound. A small constant $\beta \in \mathbf{R}$ is added to the true measured output $x(t_{i_0})$ at the beginning of the attack, where $x(t_{i_0}) + \beta$ is well below the alarm threshold. At each subsequent time step, β is multiplied by a factor $(1+\alpha)$, where $\alpha \in (0, 1)$, until $\bar{x}$ reaches the maximum allowable attack value bounded by Ω_ρ. Attackers will choose the two parameters α and β based on Ω_ρ and the attack duration. Geometric attacks can be written in the form as follows:

$$\bar{x}(t_i) = x(t_i) + \beta \times (1+\alpha)^{i-i_0}, \quad \forall \ i \in [i_0, i_0 + L_a] \tag{8}$$

where β and α are parameters that define the magnitude and speed of the geometric attack.

4.4. Neural network detector

Neural network detection algorithm was developed in Wu et al. (2018), Chen et al. (2020b, a) to identify the presence of certain types of targeting cyber-attacks on sensor measurements. A feed-forward neural network structure is used for classification of data signals, where input data in the form of the time-varying closed-loop trajectory of a nonlinear function of sensor data over the detection window was collected, and the output of the neural network detection indicates the presence and potential type of the cyber-attack (depending on how the neural network classifier is trained). Additionally, in Wu et al. (2018), Chen et al. (2020b, a), it was demonstrated that a well-trained neural network can successfully distinguish cyber-attacks from common process disturbances that should not be considered as attacks, and therefore, reduce false alarm rate during process operation. The neural network detector is implemented in real-time using a moving window that collects the most recent process data to identify the occurrence of known cyber-attacks. The LMPC of Eq. (4) and the LEMPC of Eq. (5) will switch to secure back-up sensors or open-loop control strategy to maintain closed-loop stability upon detection of cyber-attacks.

Although the neural network detectors in Wu et al. (2018), Chen et al. (2020b, a) were demonstrated to successfully detect cyber-attacks in a timely manner, the post-attack handling strategies (i.e., secure back-up sensors and open-loop control strategy) may not be optimal in reality due to unavailability of back-up devices or unknown process disturbances. In light of these considerations, in the next section, we propose a state reconstruction strategy following the detection of a cyber-attack, to continue closed-loop control using a reconstructed state value based on the faulty measurements.

5. State reconstruction

The state reconstructor is developed to estimate the true state values using state measurements $\bar{x}$ and control actions u applied in real-time operation. In this section, we first introduce the recurrent neural network that is used to develop the state reconstructor using open-loop simulation data of the nonlinear system of Eq. (1). Subsequently, the state reconstructor is implemented in real-time to obtain estimated true state values based on closed-loop simulation data under attacks.

5.1. Recurrent neural network

Recurrent neural network (RNN) has been widely used in developing nonlinear dynamic functions based on time-series data to predict future states. The RNN structure is shown in Fig. 1 and its mathematical formulation can be found in Wu et al. (2019). Since there exists a feedback loop in its neurons, RNN models exhibit temporal behavior, and therefore, can be utilized to represent dynamic systems. In this work, an RNN-based state reconstructor is proposed to estimate true state values in real time based on faulty measurements $\bar{x}$ and control actions u. Specifically, the inputs to the RNN model are $\bar{x}(t)$ and $u(t)$, $\forall t \in [t_k, t_{k+r})$, where r is the number of sampling periods in the reconstruction window, and the output of the RNN models is the estimate of the true state x over $t \in [t_k, t_{k+r})$.

To develop RNN-based state reconstructors for the nonlinear system of Eq. (1) under the min–max, surge, and geometric cyber-attacks on sensor measurements that were introduced in the previous section, we first perform extensive open-loop simulations for the nonlinear system of Eq. (1) with various $x \in \Omega_\rho$ and $u \in U$ under each of the different cyber-attacks, respectively. Specifically, starting from an initial condition $x_0 \in \Omega_\rho$, we apply a set of open-loop input sequences to the nonlinear system of Eq. (1) and introduce the above cyber-attacks at the second sampling period of each simulation run to obtain the trajectories of measured states and true states over a certain period of time (i.e., reconstruction window length $r\Delta$), respectively. Subsequently, the dataset that consists of extensive open-loop simulation runs is split into training, validation and testing datasets, and the training process of RNN models is conducted following the standard procedure as introduced in Wu et al. (2019) to minimize the difference between the predicted and the actual true state trajectories. Additionally, to ensure that the obtained RNN model can provide reliable state estimation for closed-loop operation of the nonlinear system of Eq. (1), the RNN model needs to be well trained such that the error between estimated states $\hat{x}$ and actual states x satisfies $|x - \hat{x}| \leq \gamma$, where $\gamma > 0$ is a sufficiently small bound.

The RNN models are demonstrated to be able to capture the attacking patterns, for example, the zigzag pattern of measured states in the presence of min–max cyber-attack as shown in Fig. 2(a), and provide the corresponding estimate of true state trajectory under a certain type of cyber-attack.

Remark 1. The RNN model is developed using a state-of-the-art machine learning library termed Keras. The original dataset is generated by simulating the continuous system of Eq. (1) under the sample-and-hold implementation of a sequence of piecewise constant inputs $u \in U$ (i.e., $u(t) = u(t_k)$, $\forall t \in [t_k, t_{k+1})$, where $t_{k+1} := t_k + \Delta$ and Δ is the sampling period). Explicit Euler method with a sufficiently small integration time step $h_c < \Delta$ is used to integrate the continuous system of Eq. (1). Additionally, since the RNN estimator is designed to reconstruct process true state values with the length of $r\Delta$, where r is a positive integer, all the data points of integration time step within $t \in [t_k, t_{k+r})$ are used as the internal states for the RNN model. The optimal structure of the RNN models (i.e., number of layers and neurons) are determined through a grid search, and finally, the optimization problem of RNN training process is solved using the Adam solver in

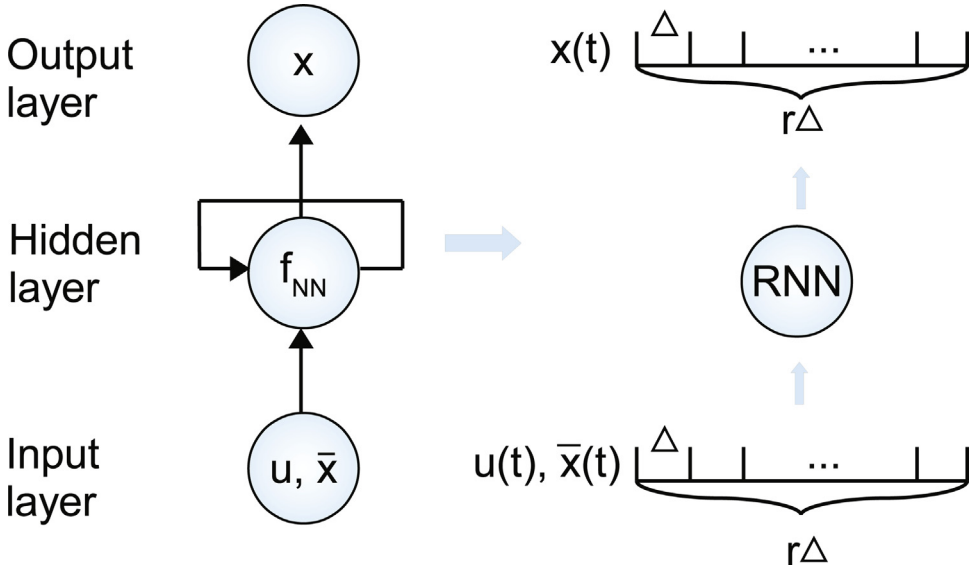

Fig. 1 – Recurrent neural network structure (left) and time-series input and output data (right), where $\bar{x}$, u are the input vectors, x is the output vector, Δ is the sampling period, rΔ is the length of reconstruction window of RNN model, and f_{NN} represents the hidden neurons that are used to capture the nonlinear relationship between input and output.

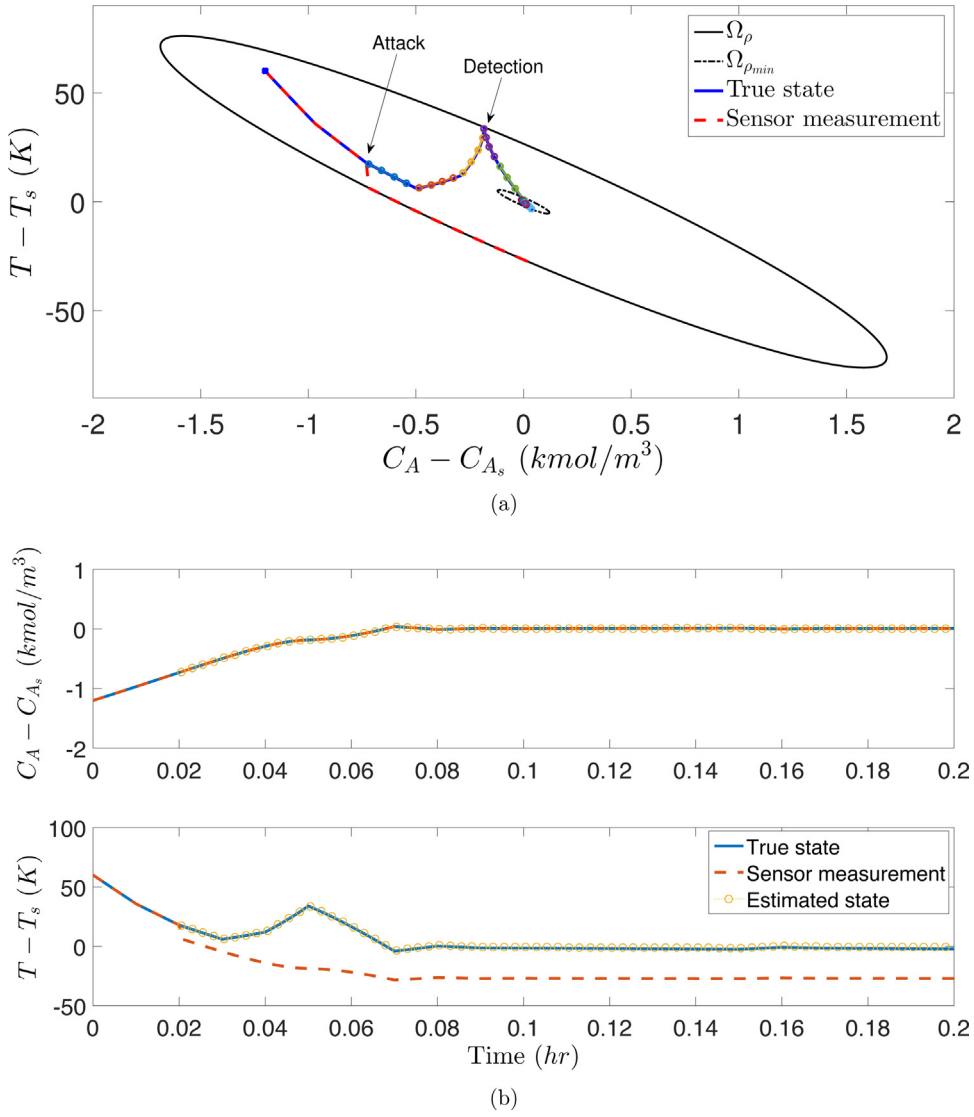

Fig. 2 – (a) State-space trajectories, and (b) closed-loop profiles of true state (blue), measured state (red), and reconstructed state (marked by colored circles) for the CSTR system of Eq. (17) with an initial condition $x_0 = (-1.2, 60)$ under LMPC when a min–max cyber-attack is introduced at t = 0.02 h on the temperature sensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Keras with the modeling error constraint and an additional early-stopping criterion to avoid over-fitting.

Remark 2. It is noted that since the dataset is generated using extensive open-loop simulations, the application of the above RNN-based state reconstructor is not restricted to the closed-loop system of Eq. (1) using the LMPC of Eq. (4) or the LEMPC of Eq. (5). It can be applied to the closed-loop system of Eq. (1) using any other controller, for example, proportional-integral-derivative controller, provided that the state measurement is available at each sampling step. Therefore, the RNN-based state reconstruction provides a general approach of state estimation for the nonlinear system of Eq. (1) under sparse sensor attacks.

Remark 3. Another approach to reconstructing state information under cyber-attacks is to predict state evolution from a secure checkpoint (i.e., from a past state that is not attacked) based on a process model for the nonlinear system of Eq. (1), and this process model can be derived either from first-principles knowledge or using RNN modeling approach as introduced in Wu et al. (2019). However, unlike the RNN model that was developed in Wu et al. (2019) to predict future states based on the current state measurement and control actions, the RNN model developed in this work takes the real-time faulty state measurements and applied control actions as inputs to estimate the corresponding true state trajectory in the same timeframe, and therefore, can provide a more accurate state estimation.

Remark 4. It should be noted that the data-based state reconstruction approach can be applied in the closed-loop simulation of the nonlinear system of Eq. (1) provided that a part of process state measurements remains secure since the RNN model essentially generates the estimate of true states for those compromised sensors based on other secure sensor measurements and applied control actions. Under the worst-case scenario that all the state measurements are under attacks, for example, the measured states remain unchanged for all times under attacks, it becomes barely possible for data-based state reconstructor to estimate the true states without any reliable information of secure sensors. In this case, an open-loop model-based control strategy could be applied to mitigate the impact of cyber-attacks to the greatest extent.

5.2. *Online reconstruction*

Once cyber-attacks are detected by neural-network-based detectors developed in Wu et al. (2018), Chen et al. (2020b, a), online state reconstruction will be implemented from the last secure checkpoint. Specifically, the RNN-based state reconstruction will be performed with the following steps. (1) Since the neural-network-based detectors in Wu et al. (2018), Chen et al. (2020b, a) are implemented in real time with a moving detection window to confirm the occurrence of cyber-attacks only if the cyber-attacks have been detected multiple times, the secure checkpoint will be set at the sampling step before the first detection to make sure the initial state measurement for the RNN reconstructor is not attacked. (2) Subsequently, the state reconstructor is applied to predict the state evolution from the last secure checkpoint to the current time step $t = t_k$ based on the sensor measurements and control actions in this period. Since the RNN model is developed with a reconstruction window length $r\Delta$, the estimated state in the second sampling period in the window will be used as the initial conditions for the next reconstruction as it moves one sampling step forward every time. (3) The estimated state $x(t_k)$ at the current time step will be sent to the controller (e.g., the LMPC of Eq. (4) or the LEMPC of Eq. (5)) to calculate the control action $u(t_k)$ for the next sampling period $t \in [t_k, t_{k+1})$. (4) After the control action $u(t_k)$ is applied and the new state measurements $\overline{x}(t), t \in [t_k, t_{k+1})$ are received, the state reconstruction window will be rolled one sampling time forward to estimate the true state value at $t = t_{k+1}$ using compromised state measurements and control actions in real-time.

Table 1 – Parameter values of the CSTR.

$T_0 = 300\,K$	$F = 5\,m^3/h$
$V = 1\,m^3$	$E = 5 \times 10^4\,kJ/kmol$
$k_0 = 8.46 \times 10^6\,m^3/kmol\,h$	$\Delta H = -1.15 \times 10^4\,kJ/kmol$
$C_p = 0.231\,kJ/kg\,K$	$R = 8.314\,kJ/kmol\,K$
$\rho_L = 1000\,kg/m^3$	$C_{A0_s} = 4\,kmol/m^3$
$Q_s = 0.0\,kJ/h$	$C_{A_s} = 1.95\,kmol/m^3$
$T_s = 401.87\,K$	

Remark 5. A key requirement for accurate state estimation is that the initial state for the RNN-based reconstructor should remain secure in order to provide a correct initial condition for the prediction of state evolution. Since the neural-network-based detectors developed in Wu et al. (2018), Chen et al. (2020b, a) are demonstrated to be able to detect cyber-attacks in a timely-manner (i.e., within a few sampling steps) with a sufficiently high accuracy, the secure checkpoint determined based on the detection outcome of neural-network-based detectors are also reliable for RNN-based state reconstructor. However, in the case that the detector is not well developed, we can choose a state measurement at an earlier time instance, or even at the beginning of operation to make sure that the initial state for reconstruction is secure.

Remark 6. It is noted that the proposed RNN-based state reconstruction method is not restricted to the cyber-attacks discussed in the manuscript since it is a data-driven approach that does not require any first-principles knowledge of process model or of cyber-attacks. For example, it can be applied to deception attacks such as optimization-based deception attack, randomly injected attacks and scheduled attacks on sensor measurements. However, there is one restriction, that is the cyber-attacks should target sensor measurements instead of blocking the communication networks between sensors and controllers, such that the RNN reconstructor can continuously receive (falsified) state measurements to make estimation. Therefore, the proposed approach may not be applied to cyber-attacks such as denial-of-service attack that makes sensor measurement unavailable to its intended users by temporarily disrupting network services.

5.3. *Closed-loop control with reconstructed states*

After the estimated state $\hat{x}(t_k)$ at the current time step $t = t_k$ is obtained through state reconstruction, the LMPC of Eq. (4) or the LEMPC of Eq. (5) will use the estimated state $\hat{x}$ instead of sensor measurement $\overline{x}$ to solve for the optimal control actions afterwards. However, considering that there may exist a state

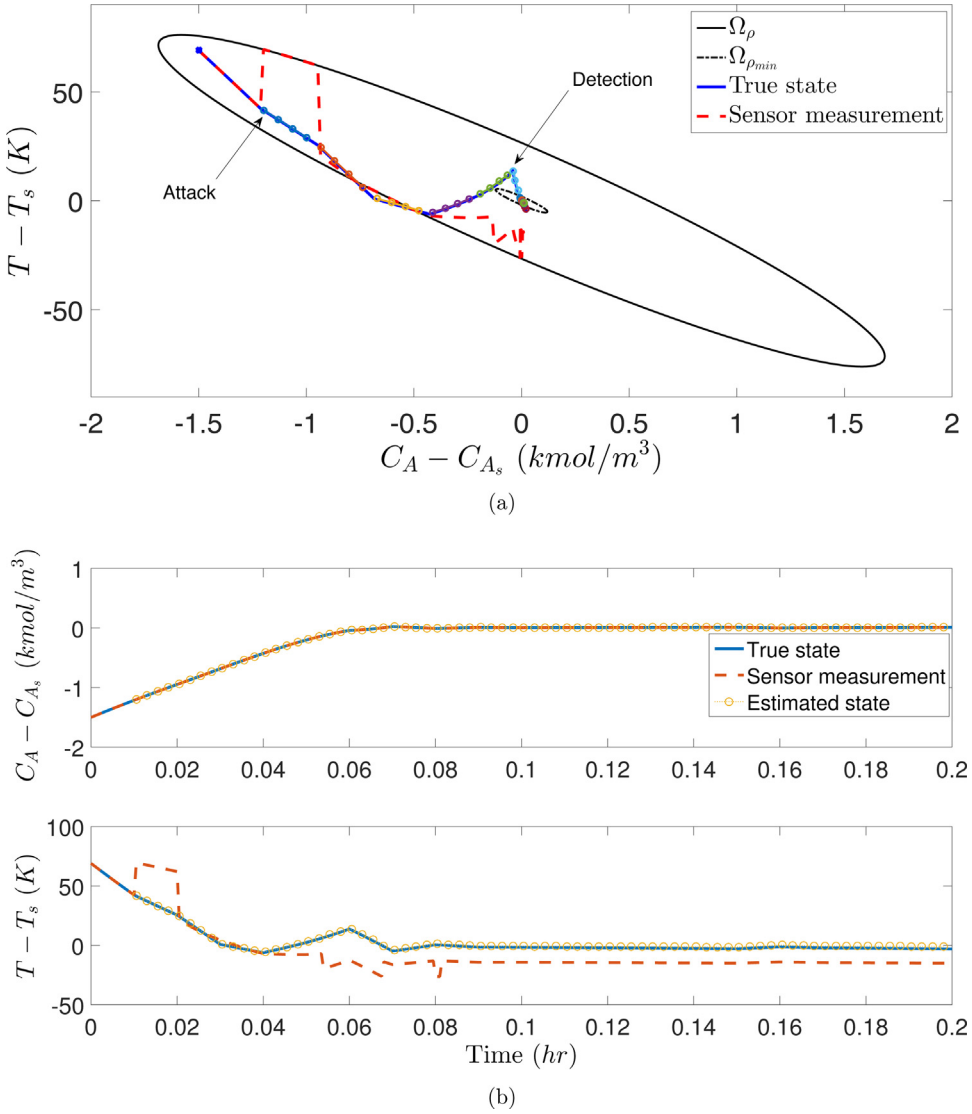

Fig. 3 – (a) State-space trajectories, and (b) closed-loop profiles of true state (blue), measured state (red), and reconstructed state (marked by colored circles) for the CSTR system of Eq. (17) with an initial condition $x_0 = (-1.5, 69)$ under LMPC when a surge cyber-attack is introduced at t = 0.01 h on the temperature sensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

estimation error, in this section, we demonstrate that the RNN model needs to be well trained to achieve a desired estimation accuracy such that closed-loop stability is still guaranteed under the LMPC of Eq. (4) and the LEMPC of Eq. (5) with state reconstruction. The following proposition is developed to demonstrate that the error between true state trajectories x and the trajectories based on estimated states $\hat{x}$ of the nonlinear system of Eq. (1) is bounded under the same control actions for finite time.

Proposition 1. *Consider the solution $x(t)$ of the nonlinear system $\dot{x} = f(x, u)$ of Eq. (1) based on the actual state x, and the solution $\hat{x}(t)$ of the nonlinear system $\dot{\hat{x}} = f(\hat{x}, u)$ based on the estimated state $\hat{x}$ with the initial condition $|x_0 - \hat{x}_0| \leq \gamma$, where $\gamma > 0$. If $x(t), \hat{x}(t) \in \Omega_\rho$ for all times, then there exists a positive constant κ such that the following inequalities hold $\forall x, \hat{x} \in \Omega_\rho$:*

$$|x(t) - \hat{x}(t)| \leq \gamma e^{L_x t} \tag{9a}$$

$$V(x) \leq V(\hat{x}) + \alpha_4(\alpha_1^{-1}(\rho))|x - \hat{x}| + \kappa|x - \hat{x}|^2 \tag{9b}$$

Proof. We define the state error vector as $e(t) = x(t) - \hat{x}(t)$ and derive the time-derivative of $e(t)$, $\forall x, \hat{x} \in \Omega_\rho$ and $u \in U$ using Eq. (3c) as follows:

$$|\dot{e}| = |f(x, u) - f(\hat{x}, u)| \leq L_x|e(t)| \tag{10}$$

$$|e(t)| = |x(t) - \hat{x}(t)| \leq \gamma e^{L_x t} \tag{11}$$

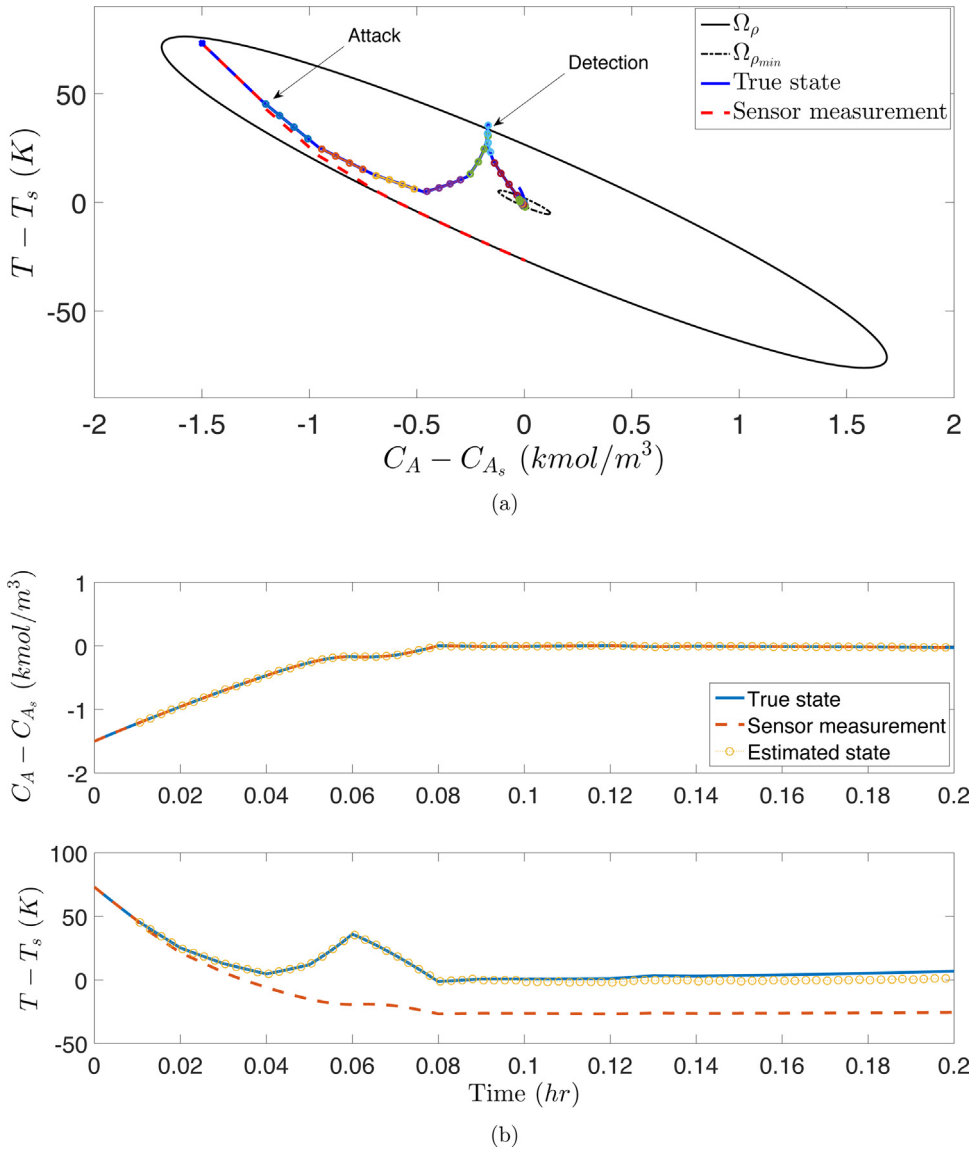

Fig. 4 – (a) State-space trajectories, and (b) closed-loop profiles of true state (blue), measured state (red), and reconstructed state (marked by colored circles) for the CSTR system of Eq. (17) with an initial condition $x_0 = (-1.5, 73)$ under LMPC when a geometric cyber-attack is introduced at $t = 0.01$ h on the temperature sensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Additionally, we derive Eq. (9b) based on Eq. (2a), Eq. (2c) and the Taylor series expansion of $V(x)$ around $\hat{x}$ for all $x, \hat{x} \in \Omega_\rho$ as follows:

$$\begin{aligned} V(x) &\leq V(\hat{x}) + \frac{\partial V(\hat{x})}{\partial x}|x - \hat{x}| + \kappa|x - \hat{x}|^2 \\ &\leq V(\hat{x}) + \alpha_4(\alpha_1^{-1}(\rho))|x - \hat{x}| + \kappa|x - \hat{x}|^2 \end{aligned} \tag{12}$$

where κ is a positive real number. □

The following proposition is developed to demonstrate that by implementing the stabilizing controller $u = \phi(\hat{x}) \in U$ based on estimated states $\hat{x}$ in a sample-and-hold fashion after detection of cyber-attacks, $\dot{V}(x)$ for the nonlinear system of Eq. (1) can be rendered negative for all times such that the true state x can be driven towards the origin.

Proposition 2. *Consider the nonlinear system of Eq. (1) under the sample-and-hold implementation of the controller $u = \phi(\hat{x}) \in U$ based on the estimated state $\hat{x}$ that satisfies $|\hat{x} - x| \leq \gamma$. Let $\epsilon_s > 0$, $\Delta > 0$ and $\rho > \rho_s > 0$ satisfy*

$$-\alpha_3(\alpha_2^{-1}(\rho_s)) + L'_x(\gamma + M\Delta) \leq -\epsilon_s \tag{13}$$

Then, $\dot{V}(x) \leq -\epsilon_s$ holds for any $x(t_k) \in \Omega_\rho \backslash \Omega_{\rho_s}$.

Proof. The proof follows closely to that for Proposition 4 in Wu et al. (2019) except that we account for the estimation error γ in $\dot{V}(x)$ as follows:

$$\begin{aligned} \dot{V}(x(t_k)) = &\frac{\partial V(x(t_k))}{\partial x} f(x(t_k), \phi(\hat{x}(t_k))) \\ &= \frac{\partial V(\hat{x}(t_k))}{\partial x} f(\hat{x}(t_k), \phi(\hat{x}(t_k))) + \frac{\partial V(x(t_k))}{\partial x} f(x(t_k), \phi(\hat{x}(t_k))) \\ &- \frac{\partial V(\hat{x}(t_k))}{\partial x} f(\hat{x}(t_k), \phi(\hat{x}(t_k))) \end{aligned} \tag{14}$$

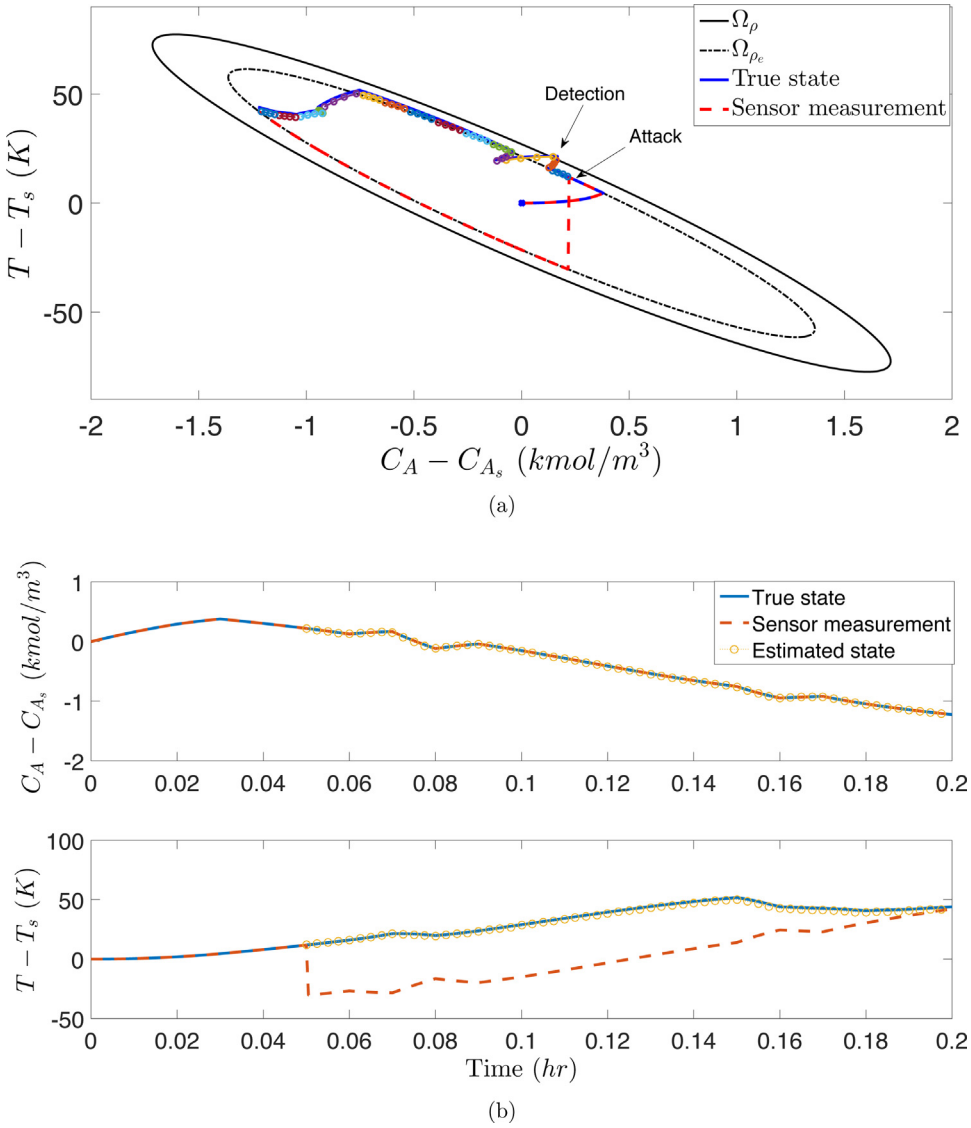

Fig. 5 – (a) State-space trajectories, and (b) closed-loop profiles of true state (blue), measured state (red), and reconstructed state (marked by colored circles) for the CSTR system of Eq. (17) under LEMPC when a min–max cyber-attack is introduced at t = 0.05 h on the temperature sensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We can further derive the following inequalities using Eq. (2a), (2b) and the Lipschitz condition of Eq. (3):

$$\begin{aligned}\dot{V}(x(t_k)) \leq\ & -\alpha_3(\alpha_2^{-1}(\rho_s)) + L'_x|x(t_k) - \hat{x}(t_k)| \\ \leq\ & -\alpha_3(\alpha_2^{-1}(\rho_s)) + L'_x\gamma\end{aligned} \tag{15}$$

Therefore, $\dot{V}(x) \leq -\epsilon_s$ can be proved by further accounting for the impact of sample-and-hold implementation of control actions following the proof in Wu et al. (2019) provided that Eq. (13) is satisfied. □

Based on the above proposition, closed-loop stability can be readily proved for the nonlinear system of Eq. (1) under the LMPC of Eq. (4), and therefore, is omitted here. The next proposition demonstrates that Ω_{ρ_e} needs to be carefully chosen for the closed-loop system under LEMPC to ensure the invariance of the stability region Ω_ρ accounting for the estimation error.

Proposition 3. *Consider the nonlinear system of Eq. (1) under the sample-and-hold implementation of the LEMPC of Eq. (5). Let $\Delta > 0$ and $\rho > \rho_e > \rho_s > 0$ satisfy the following inequality:*

$$\rho_e \leq \rho - \alpha_4(\alpha_1^{-1}(\rho))\gamma e^{L_x\Delta} - \kappa(\gamma e^{L_x\Delta})^2 \tag{16}$$

If the state estimation error $|\hat{x} - x|$ is bounded by γ for all times, then, the true state of the nonlinear system of Eq. (1) under LEMPC is guaranteed to remain inside the stability region Ω_ρ, $\forall t \geq 0$, for any $x_0 \in \Omega_\rho$.

Proof. Following the results of Proposition 1, ρ_e is determined accounting for the error between true state trajectories x of the nonlinear system of Eq. (1) and the predicted trajectories based on estimated state $\hat{x}$ under the sample-and-hold implementation of control actions. The proof follows closely to that for Theorem 2 in Heidarinejad et al. (2012), and is omitted here. □

Therefore, given that the RNN model is well trained to achieve a sufficiently small estimation error, i.e., $|x - \hat{x}| \leq \gamma$, closed-loop stability is guaranteed for the nonlinear system of Eq. (1) under resilient LMPC and LEMPC using estimated state $\hat{x}$ upon detection of cyber-attacks.

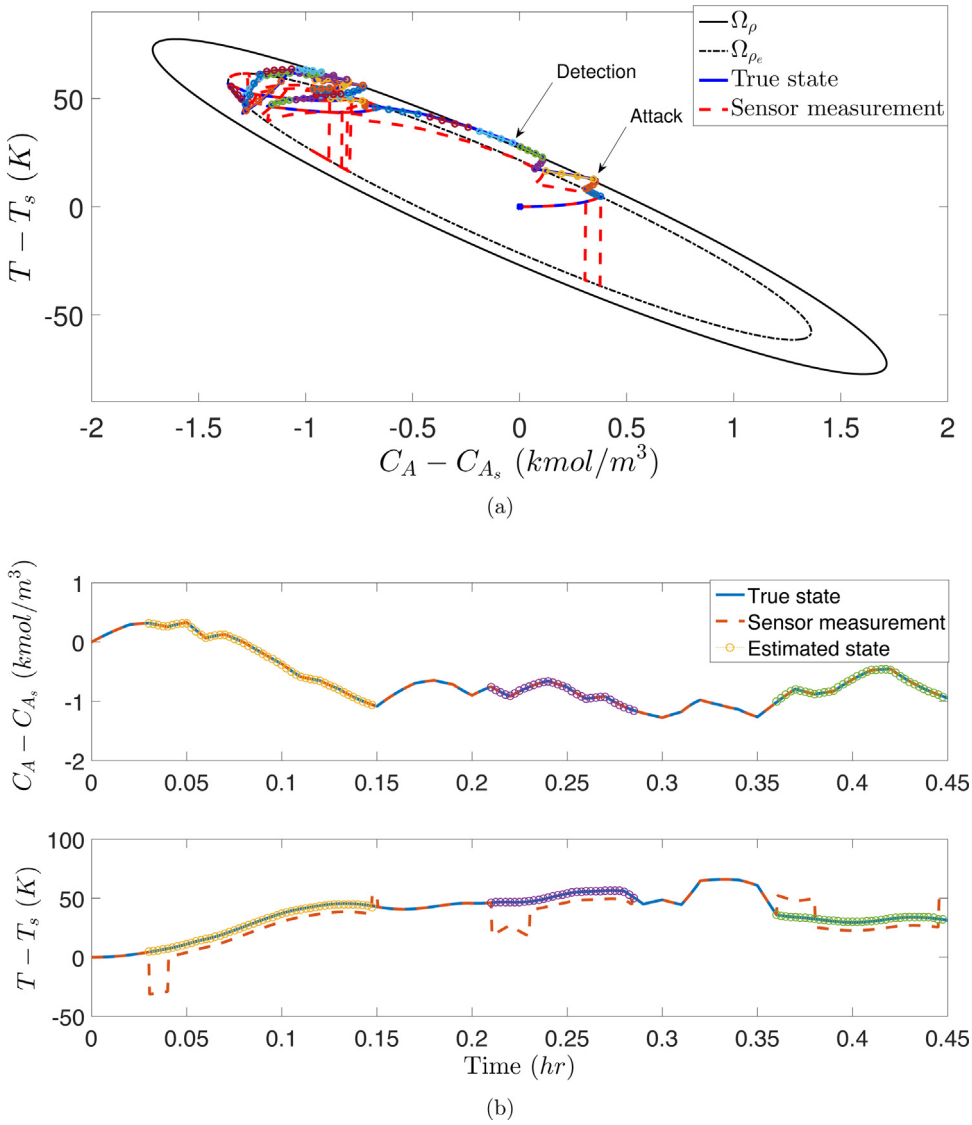

Fig. 6 – (a) State-space trajectories, and (b) closed-loop profiles of true state (blue), measured state (red), and reconstructed state (marked by colored circles) for the CSTR system of Eq. (17) under LEMPC when surge cyber-attacks are introduced at $t = 0.03$ h, $t = 0.21$ h, and $t = 0.36$ h on the temperature sensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Remark 7. In this work, we assume no measurement noise, and thus, the RNN state reconstructor takes the compromised state measurement under cyber-attacks as the inputs to estimate the true state values. However, in the presence of measurement noise, which is very common in practical systems, the RNN reconstructor can still work well as long as the training dataset is developed from simulations/industrial process data that also account for the measurement noise with the same distribution. Additionally, closed-loop stability of MPC is still guaranteed provided that the modeling error of the RNN reconstructor is sufficiently small, which will be implemented as a constraint in the training process.

6. Application to a chemical process example

In this section, a chemical process example is utilized to illustrate the application of the tracking MPC and EMPC schemes that incorporate the proposed real-time state reconstruction upon detection of cyber-attacks. Specifically, we consider a well-mixed, non-isothermal continuous stirred tank reactor (CSTR), within which an irreversible second-order exothermic reaction takes place. The second-order reaction, $A \rightarrow B$, transforms reactant A to product B at a reaction rate $r_B = k_0 e^{-E/(RT)} C_A^2$. The CSTR is equipped with a heating jacket that supplies or removes heat at a rate Q. The dynamic model of this CSTR process is described by the following material and energy balance equations:

$$\frac{dC_A}{dt} = \frac{F}{V}(C_{A0} - C_A) - k_0 e^{\frac{-E}{RT}} C_A^2 \tag{17a}$$

$$\frac{dT}{dt} = \frac{F}{V}(T_0 - T) + \frac{-\Delta H}{\rho_L C_p} k_0 e^{\frac{-E}{RT}} C_A^2 + \frac{Q}{\rho_L C_p V} \tag{17b}$$

where C_A is the concentration of reactant A in the reactor, V is the volume of the reacting liquid in the reactor (assuming the vessel has constant holdup), T is the temperature of the reactor and Q denotes the heat input rate. The concentration of reactant A in the feed is C_{A0}. The feed temperature and volumetric flow rate are T_0 and F, respectively. The reacting liquid has a constant density of ρ_L and a heat capacity of C_p. ΔH, k_0, R, and E represent the enthalpy of reaction, pre-exponential con-

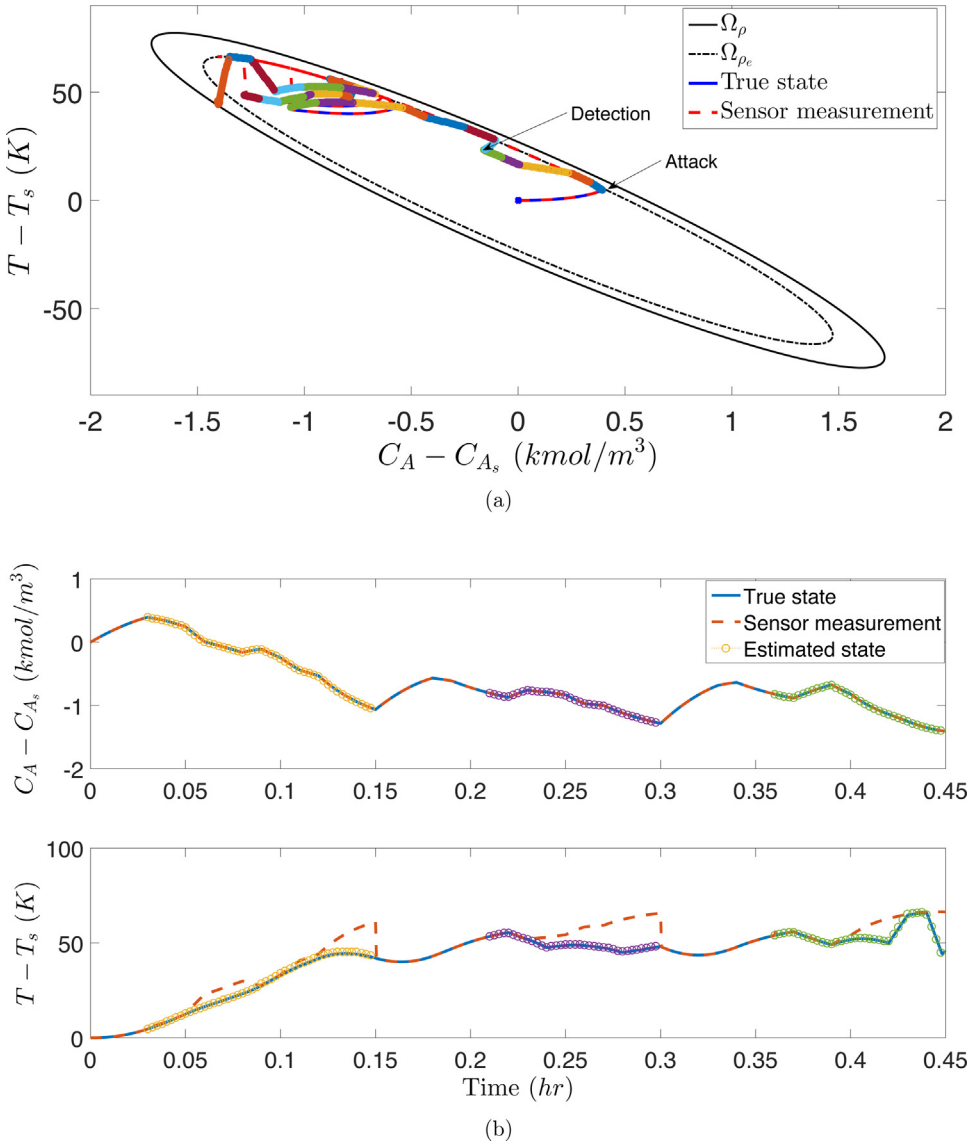

Fig. 7 – (a) State-space trajectories, and (b) closed-loop profiles of true state (blue), measured state (red), and reconstructed state (marked by colored circles) for the CSTR system of Eq. (17) under LEMPC when geometric cyber-attacks are introduced at t = 0.03 h, t = 0.21 h, and t = 0.36 h on the temperature sensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

stant, ideal gas constant, and activation energy, respectively. The process parameter values are shown in Table 1.

The manipulated inputs are the inlet concentration of reactant A and the heat input rate, which are represented by the deviation variables $\Delta C_{A0} = C_{A0} - C_{A0_s}$, $\Delta Q = Q - Q_s$, respectively. The manipulated inputs are bounded as follows: $|\Delta C_{A0}| \leq 3.5\,\text{kmol/m}^3$ and $|\Delta Q| \leq 5 \times 10^5\,\text{kJ/h}$. Therefore, the states and the inputs of the closed-loop system are $x^T = [C_A - C_{As} \quad T - T_s]$ and $u^T = [\Delta C_{A0} \quad \Delta Q]$, respectively, such that the equilibrium point of the system is at the origin of the state-space, (i.e., $x_s^T = [0, 0]$, $u_s^T = [0, 0]$).

The dynamic model of Eq. (17) is numerically simulated using the explicit Euler method with an integration time step of $h_c = 1 \times 10^{-4}$ h. The optimization problem of the LMPC of Eq. (4) and the LEMPC of Eq. (5) are solved using the python module of the python module of the IPOPT software package Wachter and Biegler, 2006, named PyIpopt with the sampling period $\Delta = 0.01$ h.

6.1. *Neural-network-based state reconstructor*

To develop state reconstructors for various types of cyber-attacks, extensive open-loop simulations are performed in the stability region Ω_ρ for the nonlinear system of Eq. (17) under cyber-attacks to generate the dataset for RNNs. The closed-loop stability region Ω_ρ for the CSTR is characterized as a level set of Lyapunov function $V(x) = x^T P x$ with $P = \begin{bmatrix} 1060 & 22 \\ 22 & 0.52 \end{bmatrix}$ and $\rho = 372$, from which the origin can be rendered asymptotically stable under the controller $u = \phi(x) \in U$. We choose various initial states $x_0 \in \Omega_\rho$ and inputs $u \in U$ and run open-loop simulations for finite sampling periods for the nonlinear system of Eq. (17) under min–max, surge, and geometric cyber-attacks, respectively. Sensor measurements $\bar{x}$ and control actions u saved every integration time step h_c are utilized as the input to the RNN model, and the corresponding true states x are the RNN output. Two-hidden-layer

RNN models with 60 neurons in each layer are designed using the state-of-the-art machine learning library, Keras to train the state reconstructors for min–max, surge, and geometric cyber-attacks, respectively with datasets consisting of around 150,000 data sequences. The averaged mean square errors of the three state reconstructors on training and validation datasets are maintained below 10^{-5}. The averaged training time for each neural network is around 2.5 h. The training is done off-line, and the obtained RNN model is used on-line for state estimation within MPC. It is noted that the state estimation within MPC is completed almost instantaneously because the RNN model after training is essentially a nonlinear function that calculates estimated values (output) given the past state measurements (input). Therefore, the computational time for running estimation using RNN models is negligible compared to the process sampling time.

6.2. *Tracking MPC*

We first carry out the closed-loop simulation results under the LMPC of Eq. (4) with state reconstruction for the nonlinear system of Eq. (17) under cyber-attacks. The control objective of LMPC is to track process state to its unstable steady-state $[C_{As}, T_s] = [1.9537\ \text{kmol/m}^3, 401.87\ \text{K}]$. The LMPC cost function of Eq. (4a) is designed to be $l_t(x, u) = |x|^2_{Q_1} + |u|^2_{Q_2}$, where $Q_1 = [500\ \ 0; 0\ \ 0.5]$ and $Q_2 = [1\ \ 0; 0\ \ 8 \times 10^{-11}]$ to balance the magnitudes of states and inputs. The detector is activated at every sampling step to identify the occurrence of cyber-attacks using the most recent process data. The neural-network-based detectors developed in Wu et al. (2018), Chen et al. (2020b, a) are used in this work to detect cyber-attacks in closed-loop simulations. The details of the neural-network-based detectors are omitted here due to limited space.

The closed-loop state trajectories and profiles for min–max, surge, and geometric cyber-attacks are shown in Figs. 2(a) and (b), 3(a) and (b) and 4(a) and (b), respectively. Specifically, in Fig. 2(a), it is shown that starting from the initial condition $x_0 = (-1.2, 60)$ the system of Eq. (17) is initially operated without any attacks. Then, the min–max cyber-attack is introduced on the temperature sensor at $t = 0.02$ h, and it is shown that the sensor measurement (dashed red trajectory) stays on the lower boundary of Ω_ρ, while the true state trajectory (blue) starts deviating from the direction towards the origin. Once the cyber-attack is detected at $t = 0.05$ h, we reconstruct the true states (colored dotted trajectories) based on past sensor measurements and control actions, and subsequently, the LMPC of Eq. (4) restabilizes the CSTR system at the steady-state by using the estimated state. In Fig. 2(b), it is demonstrated that the reconstructed concentration and temperature are very close to the true states in closed-loop simulation, and therefore, provide reliable state estimation for the feedback control with LMPC. During online implementation, state reconstruction will be ideally activated after the first positive detection given by the cyber-attack detector to save computational power, given that detection happens in real-time and promptly reports the occurrence of a cyber-attack. However, starting state reconstruction is not limited to only when the detector gives a positive detection. Here, we have plotted the reconstructed states right after the attack occurs to demonstrate the effectiveness of this RNN-based state reconstruction method throughout the attack duration. Moreover, even in the case that the sensor measurements are not faulty, the NN-based state reconstructor is also capable of predicting the true process states successfully with a sufficiently small bounded error. Therefore, state reconstruction could start at the beginning of the operation period, as long as the sensor measurements prior to which time are reliable.

Similarly, Fig. 3(a) and (b) demonstrates the closed-loop results under surge cyber-attack that is introduced on temperature sensor at $t = 0.01$ h, and is detected at $t = 0.05$ h. It is noted that in Fig. 3(a) the state trajectory leaves Ω_ρ during the third sampling since the attack remains undetected at that time; however, once the attack is detected, the closed-loop state can be driven towards the origin and maintained in $\Omega_{\rho_{\min}}$ under LMPC with state reconstruction. Additionally, Fig. 4(a) and (b) demonstrates the impact of geometric attack and of state reconstruction upon detection. It is shown in Fig. 4(b) that the pattern of geometric attack is very similar to that of min–max cyber-attack except that it gradually decreases to its lower bound to avoid sudden increase/decrease in sensor measurement. As shown in Fig. 4(a), the geometric cyber-attack is detected at $t = 0.06$ h, and the LMPC with the integration of RNN-based state reconstructor stabilizes the system at the steady-state within 0.1 h.

6.3. *Economic MPC*

The control objective of EMPC is to maximize the economic profit of the CSTR process of Eq. (17) by manipulating the inlet concentration ΔC_{A0} and the heat input rate ΔQ, while maintaining the closed-loop state trajectories in the stability region Ω_ρ for all times under LEMPC. The CSTR is initially operated at the unstable steady-state $[C_{As}, T_s] = [1.9537\ \text{kmol/m}^3, 401.87\ \text{K}]$, and $[C_{A0_s}\ Q_s] = [4\ \text{kmol/m}^3, 0\ \text{kJ/h}]$. The objective function of the LEMPC of Eq. (5) optimizes the production rate of B as follows:

$$l_e(\tilde{x}, u) = r_B(C_A, T) = k_0 e^{-E/(RT)} C_A^2 \tag{18}$$

Additionally, to make the averaged reactant material available within one operating period t_{N_p} to be its steady-state value, C_{A0s} (i.e., the averaged reactant material in deviation form, u_1, is equal to 0), we introduce the following material constraint into the formulation of the LEMPC of Eq. (5).

$$\frac{1}{t_{N_p}} \int_0^{t_{N_p}} u_1(\tau) d\tau = 0\ \text{kmol/m}^3 \tag{19}$$

The closed-loop stability region Ω_ρ is the same as the one for LMPC simulations. The CSTR system of Eq. (17) is normally operated in the region Ω_{ρ_e} with $\rho_e = 280$ under no attacks. When cyber-attacks occur, the true state trajectory may leave Ω_{ρ_e} under LEMPC, and therefore, the size of Ω_{ρ_e} is carefully chosen to make sure the state does not leave the stability region Ω_ρ before the attacks can be detected.

The closed-loop state trajectories and profiles for min–max, surge, and geometric cyber-attacks under LEMPC are shown in Figs. 5(a) and (b), 6(a) and (b) and 7(a) and (b), respectively. Specifically, as shown in Fig. 5(a) and (b), we introduce min–max cyber-attack on temperature sensor at $t = 0.05$ h, and the true state leaves Ω_{ρ_e} due to faulty sensor measurement. After detection, the closed-loop state can be maintained in Ω_{ρ_e} again under LEMPC using real-time state reconstruction. In Fig. 6(a) and (b), we perform closed-loop simulation under surge cyber-attack for multiple EMPC operating periods. It is demonstrated in Fig. 6(b) that the surge cyber-attacks are introduced in each operating period (i.e., from $t = 0$ h to $t = 0.15$ h,

from $t = 0.15$ h to $t = 0.3$ h, and from $t = 0.3$ h to $t = 0.45$ h with $t_{N_p} = 0.15$ h), from which the compromised sensor measurement first reaches its maximum allowable value and remains a small deviation from true states afterwards. Similarly, RNN-based state reconstructor successfully estimates the true state trajectory and provides a reliable correction for sensor measurement for LEMPC. Additionally, Fig. 7(a) and (b) show the simulation results of closed-loop CSTR system under geometric cyber-attack, for which the analysis is similar to the above, and is omitted here.

7. Conclusion

In this work, an RNN-based state reconstruction approach was proposed for state estimation of nonlinear processes following the detection of cyber-attacks on sensor measurements. Specifically, based on machine-learning-based detection mechanisms that were developed to detect the occurrence of cyber-attacks in closed-loop operation, we developed an RNN model to reconstruct process states using falsified state measurements and used them to calculate control actions. The RNN-based state reconstructor was applied in real-time within LMPC and LEMPC to provide reliable state estimation such that closed-loop stability of the nonlinear processes can be guaranteed upon cyber-attack detection. The application of state reconstuctors to a chemical process example under min–max, surge and geometric cyber-attacks demonstrated its effectiveness of reconstructing process states for both LMPC and LEMPC.

Conflict of interest

None declared.

Acknowledgments

Financial support from the National Science Foundation and the Department of Energy is gratefully acknowledged.

References

Alcala, C.F., Qin, S.J., 2010. Reconstruction-based contribution for process monitoring with kernel principal component analysis. Ind. Eng. Chem. Res. 49, 7849–7857.

Ao, W., Song, Y., Wen, C., 2016. Adaptive cyber-physical system attack detection and reconstruction with application to power systems. IET Control Theory Appl. 10, 1458–1468.

Bakshi, B., 1998. Multiscale PCA with application to multivariate statistical process monitoring. AIChE J. 44, 1596–1610.

Chamanbaz, M., Dabbene, F., Bouffanais, R., 2019. A physics-based attack detection technique in cyber-physical systems: a model predictive control co-design approach. Proceedings of the Australian & New Zealand Control Conference, Auckland, New Zealand, 18–23.

Chen, S., Wu, Z., Christofides, P.D., 2020a. Cyber-attack detection and resilient operation of nonlinear processes under Lyapunov-based economic model predictive control. Comput. Chem. Eng. 136, 106806.

Chen, S., Wu, Z., Christofides, P.D., 2020b. A cyber-secure control-detector architecture for nonlinear processes. AIChE J. 66, e16907.

Durand, H., 2018. A nonlinear systems framework for cyberattack prevention for chemical process control systems. Mathematics 6, 169.

Harkat, M., Djelel, S., Doghmane, N., Benouaret, M., 2007. Sensor fault detection, isolation and reconstruction using nonlinear principal component analysis. Int. J. Autom. Comput. 4, 149–155.

Heidarinejad, M., Liu, J., Christofides, P.D., 2012. Economic model predictive control of nonlinear process systems using Lyapunov techniques. AIChE J. 58, 855–870.

Hu, Q., Fooladivanda, D., Chang, Y.H., Tomlin, C.J., 2017. Secure state estimation for nonlinear power systems under cyber attacks. Proceedings of the American Control Conference, Seattle, Washington, 2779–2784.

Kim, M., Liu, H., Kim, J.T., Yoo, C., 2013. Sensor fault identification and reconstruction of indoor air quality (IAQ) data using a multivariate non-Gaussian model in underground building space. Energy Build. 66, 384–394.

Lieftucht, D., Völker, M., Sonntag, C., Kruger, U., Irwin, G., Engell, S., 2009. Improved fault diagnosis in multivariate systems using regression-based reconstruction. Control Eng. Pract. 17, 478–493.

Lin, Y., Sontag, E.D., 1991. A universal formula for stabilization with bounded controls. Syst. Control Lett. 16, 393–397.

Nateghi, S., Shtessel, Y., Barbot, J., Edwards, C., 2018. Cyber attack reconstruction of nonlinear systems via higher-order sliding-mode observer and sparse recovery algorithm. Proceedings of the IEEE Conference on Decision and Control, Miami Beach, Florida, 5963–5968.

Qin, S., Dunia, R., 2000. Determining the number of principal components for best reconstruction. J. Process Control 10, 245–250.

Singh, J., Nene, M., 2013. A survey on machine learning techniques for intrusion detection systems. Int. J. Adv. Res. Comput. Commun. Eng. 2, 4349–4355.

Sun, Y., Yang, G., 2019. Robust event-triggered model predictive control for cyber-physical systems under denial-of-service attacks. Int. J. Robust Nonlinear Control 29, 4797–4811.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., Yin, K., 2003. A review of process fault detection and diagnosis: Part III: process history based methods. Comput. Chem. Eng. 27, 327–346.

Wachter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Programm. 106, 25–57.

Wang, S., Chen, Y., 2004. Sensor validation and reconstruction for building central chilling systems based on principal component analysis. Energy Convers. Manag. 45, 673–695.

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, 22 pages.

Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019. Machine learning-based predictive control of nonlinear processes. Part I: theory. AIChE J. 65, e16729.