Hindawi
Journal of Electrical and Computer Engineering
Volume 2018, Article ID 7204939, 14 pages
https://doi.org/10.1155/2018/7204939

Research Article

A Novel Covert Agent for Stealthy Attacks on Industrial Control Systems Using Least Squares Support Vector Regression

Weize Li, Lun Xie, and Zhiliang Wang

School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Lun Xie; xielun@ustb.edu.cn

Received 12 July 2017; Accepted 24 December 2017; Published 1 February 2018

Academic Editor: Vinod Sharma

Research on stealthiness has become an important topic in the field of data integrity (DI) attacks. To construct stealthy DI attacks, a common assumption in most related studies is that attackers have prior model knowledge of physical systems. In this paper, such assumption is relaxed and a covert agent is proposed based on the least squares support vector regression (LSSVR). By estimating a plant model from control and sensory data, the LSSVR-based covert agent can closely imitate the behavior of the physical plant. Then, the covert agent is used to construct a covert loop, which can keep the controller's input and output both stealthy over a finite time window. Experiments have been carried out to show the effectiveness of the proposed method.

1. Introduction

Industrial control systems (ICSs) are widely deployed in modern critical infrastructures (CIs), and their incapacitation can cause serious damage to equipment, environment, or even people's lives [1]. During the past ten years, many efforts have been made to improve the security of ICSs [2, 3]. Among the existing research on ICSs security, a great deal of attention has been given to the study of stealthy data integrity (DI) attacks [4, 5], which can violate the integrity of control and sensory data. The purpose of such attacks is to disrupt the physical process while remaining stealthy with respect to anomaly detectors [6].

To construct stealthy DI attacks, a common assumption in most related studies is that attackers have prior model knowledge of physical systems. Kwon et al. [7] investigated three kinds of stealthy deception attacks on a linear time-invariant system with Gaussian noise. Their results showed that if an attacker had perfect model knowledge of the target system, he could carefully design a stealthy attack to avoid being detected by the monitoring system. Pang et al. [8] proposed stealthy false data injection (FDI) attacks for both feedback and forward channels of the networked control systems. It was assumed that the attacker knew the detailed system parameters. Such assumption can also be found in the recent work of Teixeira et al. [9], Sedghi and Jonckheere [10], Manandhar et al. [11], and Dutta and Langbort [12]. In particular, in [9], the authors also considered a more moderate scenario where the attacker's model knowledge contains some uncertainties. In [13], the authors presented a covert agent structure and showed that the better the covert agent's model of the plant, the easier it was for the covert agent to hide its actions.

Besides the perfect model knowledge of physical systems, there is a more rigorous assumption that attackers also have other model knowledge of target systems. Cárdenas et al. [14] studied three types of stealthy attacks that aimed at raising the pressure in a tank without being detected. The powerful attacker was assumed to have prior knowledge of the exact plant model and the anomaly detection scheme. In the work of Teixeira et al. [15], the model knowledge was divided into three categories: the model of the physical system, the model of the feedback controller, and the model of the anomaly detector. Attacks constrained by different levels of prior model knowledge were illustrated by experiments on a quadruple-tank process control testbed. In [16], the authors considered a stronger adversary who not only knew the physical model and the detection scheme, but also could adapt to different detection thresholds.

As discussed before, most prior works on stealthy DI attacks are based on various assumptions that attackers have model knowledge of target systems at different levels. However, there is no description of how such model knowledge can be obtained by an attacker. Although the assumptions of model knowledge are very useful for identifying subtle and stealthy malicious attacks, it may be difficult to acquire such prior knowledge in many practical scenarios, where explicit models of physical systems are usually not available directly [17].

Recently, increasing attention has been paid to stealthy DI attacks without the prior model knowledge of physical systems. Unlike the studies discussed before, Yu and Chin [18] proposed a principal component analysis (PCA) based method to design blind FDI attacks, which did not need any prior knowledge of Jacobian matrix in smart grid. Furthermore, Anwar and Mahmood [19] clarified that the PCA based blind attack strategy was only valid for the measurements with Gaussian noises. In the case of gross errors, they proposed the accelerated proximal gradient (APG) method to circumvent the gross error issue and construct stealthy attacks. Most recently in [20], the authors proposed a sparse optimization based stealthy attacks construction strategy and demonstrated how FDI attacks could be constructed blindly, that is, without the system model knowledge. However, unfortunately, these three studies were closely related to the smart grid, and the proposed methods were designed for the approximation of Jacobian matrix.

In the framework of a general dynamic cyberphysical system (CPS), Yuan and Mo [21] applied the classical system identification technique to the construction of stealthy attacks. The spectral factorization based method was used to identify the transfer function of the physical system by observing the input-output data from the system. Furthermore, they proved a necessary condition and a sufficient condition, under which the perfect model of the system could be successfully identified. However, such conditions are overly restrictive for widespread applications. In fact, it is more realistic to consider that the identified model of the system is not perfect. That is, there is a model error between the identified model and the real system model. Motivated by this consideration, we explored the possibility that an attacker can carry out stealthy DI attacks on the ICS by identifying a not so perfect model of the system.

The most similar work to ours is the recent study of Kim et al. [22], where a subspace estimation method was used to estimate a system operating subspace from sensor measurements. Based on the subspace information, stealthy attacks could be constructed without the need of prior system model knowledge. As shown in Figure 1(a), the unobservable attack is launched by adding a corresponding perturbation to the sensor data, and the modified sensor data can avoid being detected by the anomaly detector. However, because the ultimate objective of the attack is to disrupt the system's behavior, the controller's output will be abnormal. Another similar case is the replay attack, which also does not require any prior knowledge. It gathers sequences of data for a certain amount of time and afterwards just repeats the recorded data. Teixeira et al. [15] introduced an interesting instance of this attack scenario which consists of applying a physical attack to the plant while using the replay attack to render the physical attack stealthy. However, the replay attack on the sensor data could also cause anomalies in the controller's output, and this point will be revealed later in our experiments.

Our goal is to design a covert agent to keep the controller's input and output both stealthy over a finite time window. To this end, we propose a function estimation based covert agent as shown in Figure 1(b). The proposed covert agent can be used to construct a two-loop covert structure in Figure 1(c), which consists of two loops: the covert loop and the attack loop. In comparison, Figure 1(d) shows a typical structure of the prior model knowledge based covert attack [13]. The core idea of such structure is to calculate the attack effect on the plant output measurements and subtract the effect from the measured plant output. By contrast, in the two-loop covert structure, the covert loop covers up the effect of the real attack on the physical plant by closely imitating the expected behavior of the physical plant over a finite time window. For the sake of concentrating on the stealthiness, this paper will be restricted to the construction of the covert loop and will not deal with the attack loop.

The main contribution of this paper is the exploratory attempt to establish the feasibility of machine learning based stealthy DI attacks. In this paper, we use the least squares support vector regression (LSSVR) to demonstrate that point. The LSSVR has emerged as a popular data-driven modeling method, and it has uniform approximation ability for any complex nonlinear system [23]. As far as we know, there is no LSSVR-based DI attack reported in the literature. Overall, the contributions of this work are threefold. First, we give a formal description of the LSSVR-based covert agent. Second, we present the procedure of how to train a covert agent model. Third, we provide a case study of a continuous stirred tank heater (CSTH) pilot plant to illustrate and demonstrate the effectiveness of the covert agent.

It is necessary to mention that the purpose of this work is not to facilitate stealthy attacks but to disclose the potential attacks, where the attackers do not need any prior model knowledge of physical systems, and to encourage the corresponding research of the defending methods. The rest of this paper is organized as follows. Section 2 introduces the LSSVR for function estimation. Section 3 gives the covert agent model and the procedure of training the model. Section 4 is an overview of the experiments, and the experimental results are presented in Section 5. Finally, conclusions and future work are summarized in Section 6.

2. Least Squares Support Vector Regression (LSSVR)

The least squares support vector machine (LSSVM) is an alteration of the standard support vector machine (SVM) [24]. By changing the inequality constraints in SVR into equality ones, the LSSVM method can avoid the long and computationally difficult convex quadratic programming

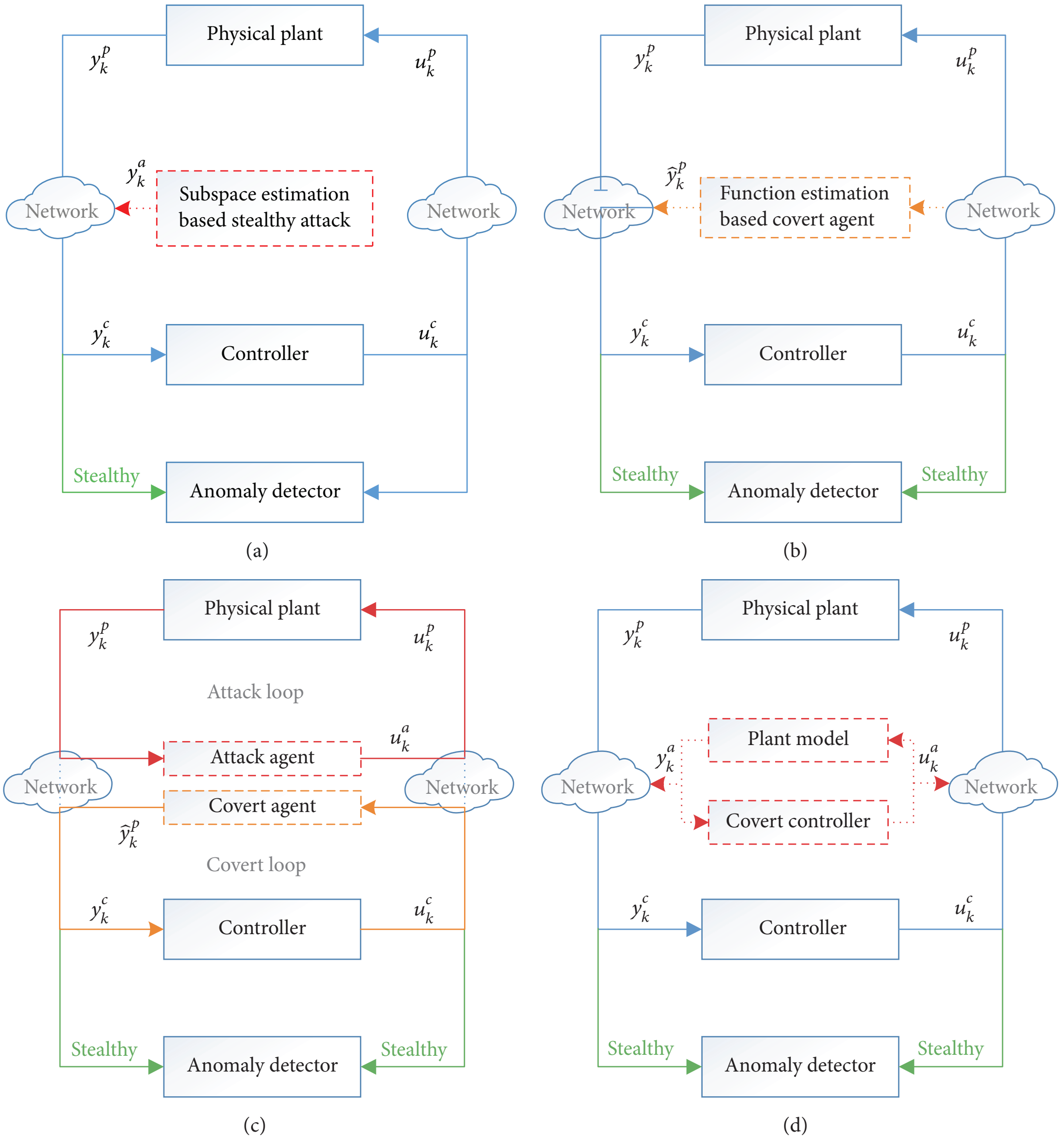


Figure 1: Schematic diagrams for the proposed covert agent and the closely related stealthy attacks. (a) The subspace estimation based unobservable attack, (b) the proposed covert agent, (c) application of the proposed covert agent, and (d) the prior model knowledge based covert structure.

and, thus, largely speeds up training. The LSSVM for regression is called LSSVR, which has been extended and applied to forecasting by many studies [25–27]. In this section, we briefly introduce the LSSVR for function estimation.

Given training set $\{x_k, y_k\}_{k=1}^{N}$, the regression function of LSSVR can be defined as follows:

$$y(x) = w^T \varphi(x) + b, \tag{1}$$

where $x \in \Re^n$, $y \in \Re$, and $\varphi(\cdot)$ is the mapping from the original feature space to the high dimensional feature space. w is the coefficient vector and b is a bias term. The optimization problem of LSSVR is given as follows:

$$\begin{aligned} \min_{w,b,e} \quad & J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \\ \text{subjective to} \quad & y_i = w^T \varphi(x_i) + b + e_i, \\ & \qquad i = 1, 2, \ldots, N, \end{aligned} \tag{2}$$

where γ is the regularization parameter and e_i is the slack variable for x_i. The Lagrangian is constructed as follows:

$$\begin{aligned} L(w,b,e,\alpha) = {} & J(w,e) \\ & - \sum_{i=1}^{N} \alpha_i \left\{ w^T \varphi(x_i) + b + e_i - y_i \right\}, \end{aligned} \tag{3}$$

where α_i $(i = 1, 2, \ldots, N)$ are the Lagrange multipliers. The conditions for optimality are

$$\begin{aligned}
&\frac{\partial L}{\partial w} = 0 \longrightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i), \\
&\frac{\partial L}{\partial b} = 0 \longrightarrow \sum_{i=1}^{N} \alpha_i = 0, \\
&\frac{\partial L}{\partial e_i} = 0 \longrightarrow \alpha_i = \gamma e_i, \quad i = 1, \ldots, N, \\
&\frac{\partial L}{\partial \alpha_i} = 0 \longrightarrow y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, \ldots, N.
\end{aligned} \tag{4}$$

With the solution of

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{5}$$

where $y = [y_1; \ldots; y_N]$, $1_v = [1; \ldots; 1]$, and $\Omega_{ij} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j)$, for $i, j = 1, \ldots, N$, the LSSVR model for function estimation is

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + b. \tag{6}$$

The kernel function $K(x_i, x)$ is any symmetric function that satisfies Mercer's condition. In this study, the radial basis function (RBF) is used as the kernel function due to its strong nonlinear modeling ability. The RBF is formulated as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \tag{7}$$

Using RBF kernels, the LSSVR has only two tuning parameters: the regularization parameter (γ) and the kernel function parameter (σ), which is lesser than the tuning parameters of standard SVR.

3. Covert Agent Based on LSSVR

3.1. Covert Agent Model. Suppose that the physical plant is a linear time-invariant (LTI) process, which is modeled in a discrete-time state-space form [28, 29]:

$$\begin{aligned}
\mathbf{x}_{k+1} &= A\mathbf{x}_k + B\mathbf{u}_k, \\
\mathbf{y}_k &= C\mathbf{x}_k + \mathbf{w}_k,
\end{aligned} \tag{8}$$

where $\mathbf{x}_k \in R^m$ is the state variable, $\mathbf{u}_k \in R^q$ is the control input, and $\mathbf{y}_k \in R^p$ is the measurement vector. The measurement noise $\mathbf{w}_k \in R^p$ is independent Gaussian noise vector with zero mean and covariance $Q > 0$. The system operates in closed loop and the control input $\mathbf{u}_k$ is given by the feedback controller:

$$\mathbf{u}_k = K(\mathbf{y}_k), \tag{9}$$

where K is the controller function that makes the closed-loop system stable.

We now consider the case where the attacker can both capture and inject the data transmitted via the network (i.e., $\mathbf{y}$ and $\mathbf{u}$). The control and sensory data are recorded by the attacker to generate the training dataset, which is described by the following notation:

(i) $T = \{1, 2, \ldots, k, \ldots, n\}$ is a set of sampling instants over a finite time window;

(ii) $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_k, \ldots, \mathbf{y}_n\}$ is a dataset of output variables captured over the sampling time window T;

(iii) $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_k, \ldots, \mathbf{u}_n\}$ is a dataset of input variables captured over the sampling time window T;

(iv) $\mathbf{y}_k = \{y_k^1, \ldots, y_k^j, \ldots, y_k^p\}$ is a data record of output variables at the kth time instant;

(v) $\mathbf{u}_k = \{u_k^1, \ldots, u_k^i, \ldots, u_k^q\}$ is a data record of input variables at the kth time instant;

(vi) y_k^j denotes the value of the jth output variable at the kth time instant;

(vii) u_k^i denotes the value of the ith input variable at the kth time instant;

(viii) $J = \{1, 2, \ldots, j, \ldots, p\}$ is a set of output variables of the physical plant;

(ix) $I = \{1, 2, \ldots, i, \ldots, q\}$ is a set of input variables of the physical plant.

From the system model in (8), we have

$$\begin{aligned}
\mathbf{y}_k &= C\mathbf{x}_k + \mathbf{w}_k, \\
\mathbf{y}_{k+1} &= CA\mathbf{x}_k + CB\mathbf{u}_k + \mathbf{w}_{k+1}.
\end{aligned} \tag{10}$$

If $C^T C$ is nonsingular, then we can obtain

$$\mathbf{y}_{k+1} - \mathbf{w}_{k+1} = CA\left(C^T C\right)^{-1} C^T \left(\mathbf{y}_k - \mathbf{w}_k\right) + CB\mathbf{u}_k. \tag{11}$$

In order to reduce the effect of Gaussian noise, a wavelet filter $W(\cdot)$ is applied to the data. The filtered data are given by

$$\begin{aligned}
\tilde{\mathbf{y}}_{k+1} &= W(\mathbf{y}_{k+1}), \\
\tilde{\mathbf{y}}_k &= W(\mathbf{y}_k),
\end{aligned} \tag{12}$$

and the estimated noises are

$$\begin{aligned}
\widehat{\mathbf{w}}_{k+1} &= \mathbf{y}_{k+1} - W(\mathbf{y}_{k+1}), \\
\widehat{\mathbf{w}}_k &= \mathbf{y}_k - W(\mathbf{y}_k).
\end{aligned} \tag{13}$$

Based on (12) and (13), (11) can be rewritten as

$$\tilde{\mathbf{y}}_{k+1} = CA\left(C^T C\right)^{-1} C^T \tilde{\mathbf{y}}_k + CB\mathbf{u}_k + \mathbf{e}_{k+1}, \tag{14}$$

where

$$\mathbf{e}_{k+1} = CA\left(C^T C\right)^{-1} C^T \left(\widehat{\mathbf{w}}_k - \mathbf{w}_k\right) + \mathbf{w}_{k+1} - \widehat{\mathbf{w}}_{k+1}. \tag{15}$$

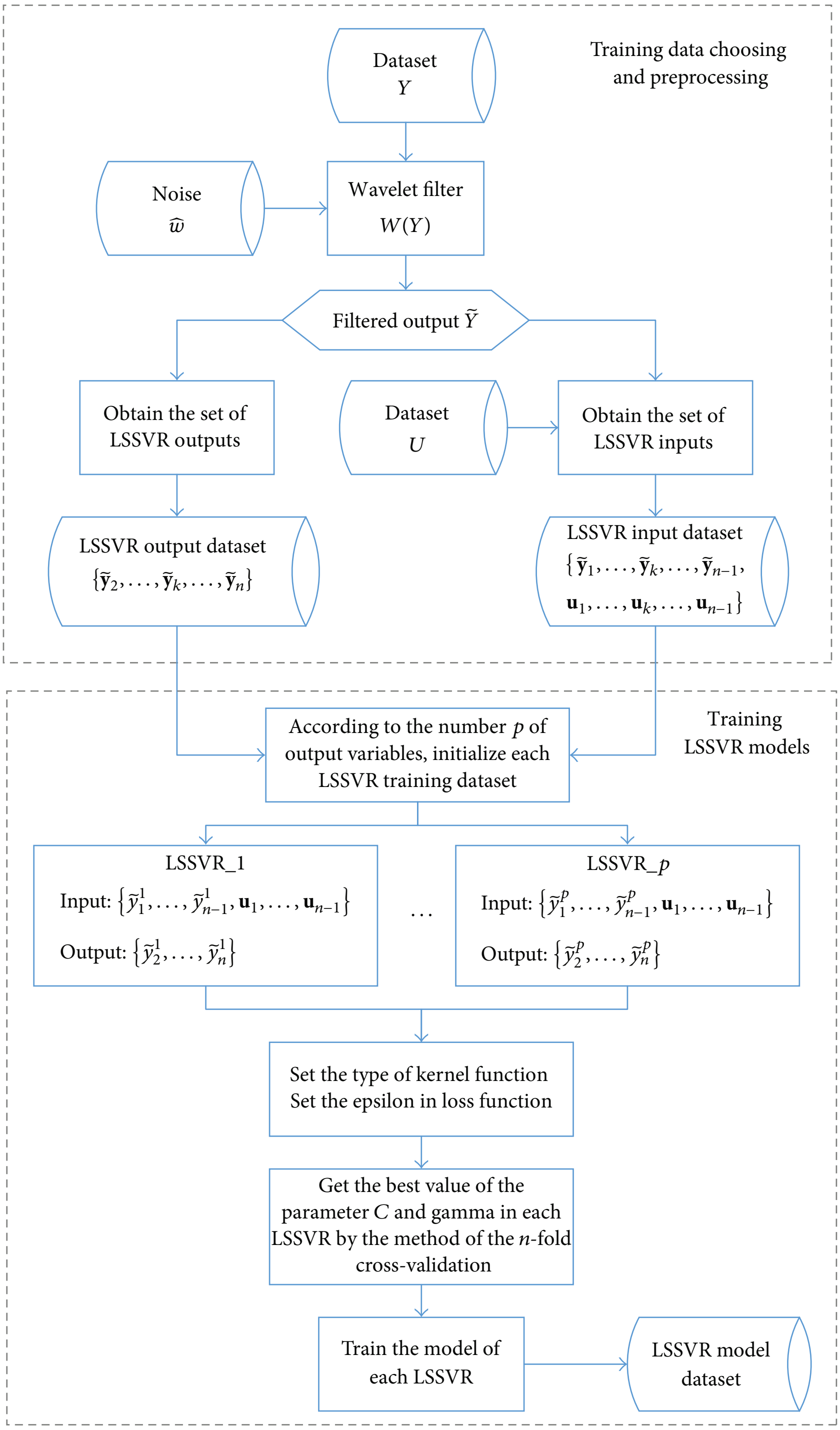

Dataset
Y
Training data choosing and preprocessing
Noise
$\widehat{w}$
Wavelet filter
$W(Y)$
Filtered output $\widetilde{Y}$
Obtain the set of LSSVR outputs
Dataset
U
Obtain the set of LSSVR inputs
LSSVR output dataset
$\{\tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_k, \dots, \tilde{\mathbf{y}}_n\}$
LSSVR input dataset
$\{\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_k, \dots, \tilde{\mathbf{y}}_{n-1}, \mathbf{u}_1, \dots, \mathbf{u}_k, \dots, \mathbf{u}_{n-1}\}$
According to the number p of output variables, initialize each LSSVR training dataset
Training LSSVR models
LSSVR_1
Input: $\{\tilde{y}^1_1, \dots, \tilde{y}^1_{n-1}, \mathbf{u}_1, \dots, \mathbf{u}_{n-1}\}$
Output: $\{\tilde{y}^1_2, \dots, \tilde{y}^1_n\}$
...
LSSVR_p
Input: $\{\tilde{y}^p_1, \dots, \tilde{y}^p_{n-1}, \mathbf{u}_1, \dots, \mathbf{u}_{n-1}\}$
Output: $\{\tilde{y}^p_2, \dots, \tilde{y}^p_n\}$
Set the type of kernel function
Set the epsilon in loss function
Get the best value of the parameter C and gamma in each LSSVR by the method of the n-fold cross-validation
Train the model of each LSSVR
LSSVR model dataset

Figure 2: Procedure of training the covert agent model.

Assume that the signal noise can be well filtered by $W(\cdot)$ and the error $\mathbf{e}_{k+1}$ can be ignored. Then, (14) changes to

$$\tilde{\mathbf{y}}_{k+1} \approx CA\left(C^T C\right)^{-1} C^T \tilde{\mathbf{y}}_k + CB\mathbf{u}_k = F\left(\tilde{\mathbf{y}}_k, \mathbf{u}_k\right). \qquad (16)$$

Without the prior knowledge of A, B, and C, we use the LSSVR to estimate $\widehat{F}$ of F from the training data with the input $[\tilde{\mathbf{y}}_k, \mathbf{u}_k]$ and the output $\tilde{\mathbf{y}}_{k+1}$. However, for each $\tilde{y}^j_{k+1}$ in $\tilde{\mathbf{y}}_{k+1}$, we do not have the knowledge of the relatedness between $\tilde{y}^j_{k+1}$ and the other variables. For the relatedness between $\tilde{y}^j_{k+1}$ and $\mathbf{u}_k$, we keep it loose and select all $\mathbf{u}_k$ as the input data. For the relatedness between $\tilde{y}^j_{k+1}$ and $\tilde{\mathbf{y}}_k$, we select $\tilde{y}^j_k$ as the input, for the reason that the sample $\tilde{y}^j_k$ is heavily correlated with the

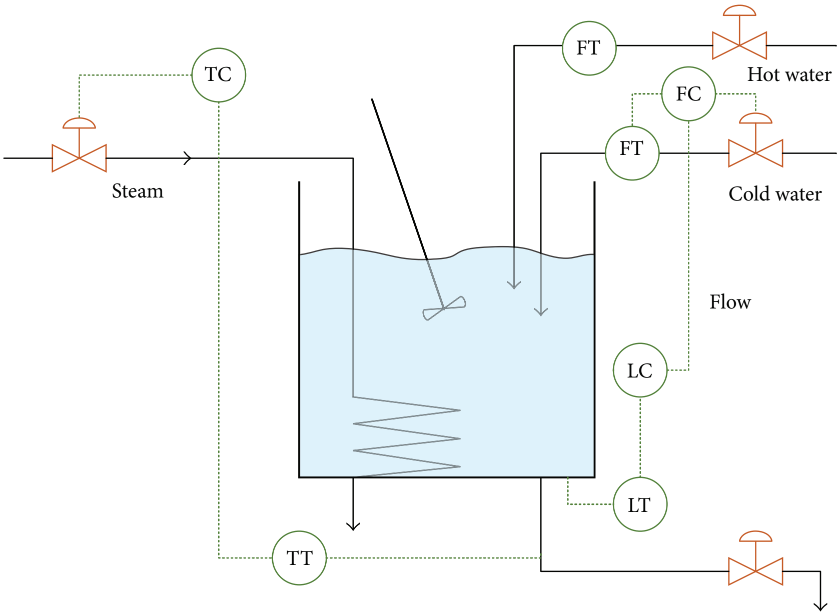


Figure 3: The continuous stirred tank heater.

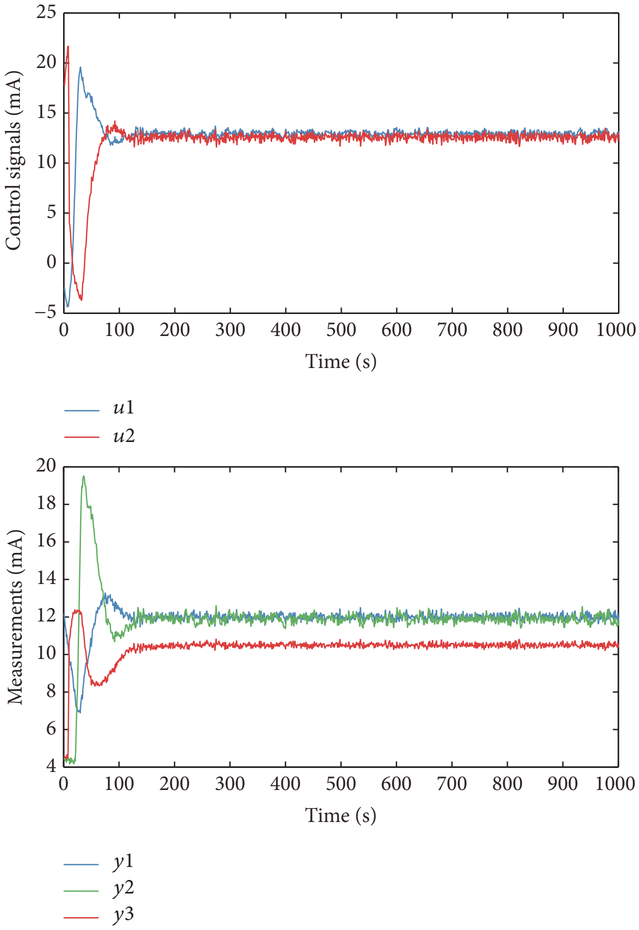


Figure 4: Normal data acquired from the closed-loop CSTH system with the standard operating condition.

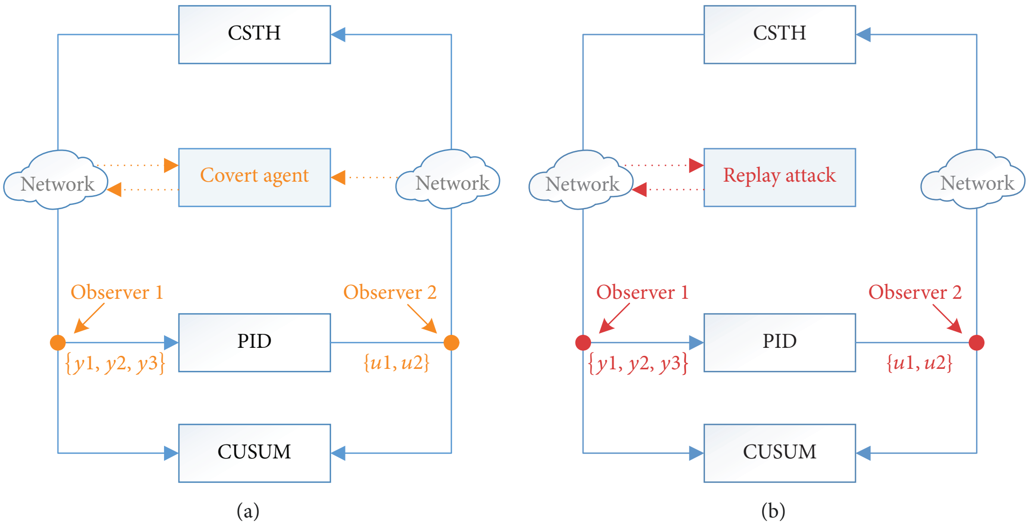


Figure 5: Setups of the experiments. (a) The proposed covert agent and (b) the replay attack.

next sample $\tilde{y}^j_{k+1}$ in a physical process. Therefore, the function estimation $\widehat{F}$ of F is given by

$$\widehat{F} = \left\{\widehat{f}^1, \dots, \widehat{f}^j, \dots, \widehat{f}^p\right\}, \qquad (17)$$

where

$$\widehat{f}^j = \text{LSSVR\_train}\left(\tilde{y}^j_{k+1}, \left[\tilde{y}^j_k, \mathbf{u}_k\right]\right). \qquad (18)$$

Then, the prediction $\widehat{\tilde{\mathbf{y}}}_{n+1}$ of $\tilde{\mathbf{y}}_{n+1}$ can be expressed as

$$\widehat{\tilde{\mathbf{y}}}_{n+1} = \left\{\widehat{\tilde{y}}^1_{n+1}, \dots, \widehat{\tilde{y}}^j_{n+1}, \dots, \widehat{\tilde{y}}^p_{n+1}\right\}, \qquad (19)$$

where

$$\widehat{\tilde{y}}^j_{n+1} = \begin{cases} \widehat{f}^j\left(\tilde{y}^j_{t_s}, \mathbf{u}_{t_s}\right), & n = t_s, \\ \widehat{f}^j\left(\widehat{\tilde{y}}^j_n, \mathbf{u}_n\right), & n > t_s, \end{cases} \qquad (20)$$

where t_s is the start time when the physical plant is covered by the covert agent. From (12) and (13), we have the output of the covert agent, which is the estimation $\widehat{\mathbf{y}}_{n+1}$ of $\mathbf{y}_{n+1}$; that is,

$$\widehat{\mathbf{y}}_{n+1} = \widehat{\tilde{\mathbf{y}}}_{n+1} + \widehat{\mathbf{w}}_{n+1}. \qquad (21)$$

3.2. Procedure of Training the Covert Agent Model. The training of the covert agent model consists of three phases: (1) data recording phase, (2) model training phase, and (3) output predicting phase. In the first phase, the control and sensory data are recorded to generate a training dataset Y and U, which will be used to train the covert plant model in the second phase. As shown in Figure 2, the dataset Y is firstly preprocessed to generate the required data for training each LSSVR model. Then, optimal parameters γ and c for each LSSVR are obtained through the automated grid search with n-fold cross-validation [30] on the training data. Finally, the outputs of this phase are p LSSVR models; that is, there is a LSSVR model for each output variable of the physical plant. In the third phase, as described in the previous subsection, the predictions $\widehat{\mathbf{y}}$ are generated based on the LSSVR models, and they are fed back to the controller to cover the real outputs of the physical plant.

4. Experiment Overview

The covert loop is illustrated by a case study of a continuous stirred tank heater (CSTH) pilot plant. In this section, the CSTH Simulink platform is briefly introduced, and the experiment setup is presented. Moreover, the assessment method used to evaluate the experimental results is also presented.

4.1. The CSTH Simulink Platform. The configuration of the CSTH plant is shown in Figure 3. Hot water and cold water are mixed in a stirred tank, heated by steam through a heating coil, and drained from the tank through a long pipe. A more detailed description of the CSTH model can be found in [31].

Our experiment is based on the CSTH Simulink model with closed-loop control, which is provided by Thornhill et al. (http://personal-pages.ps.ic.ac.uk/~nina/CSTHSimulation/index.htm). Under the closed-loop control, the CSTH model runs to a steady state from a nonsteady initial condition. The steady-state valve positions and instrument conditions in this experimental case are shown in Table 1 [31]. The simulation input and output represent electronic signals on 4–20 mA scale. The inputs to the CSTH are control signals of the cold water and steam valves. The outputs are electronic measurements from the temperature, level, and cold water flow.

Based on the CSTH basic Simulink model, Gaussian noises are added to the three outputs of the CSTH. Figure 4 shows the normal control signals and measurements under the closed-loop control. The default simulation time is 1000 s, and the default sampling rate is 3600 samples per hour

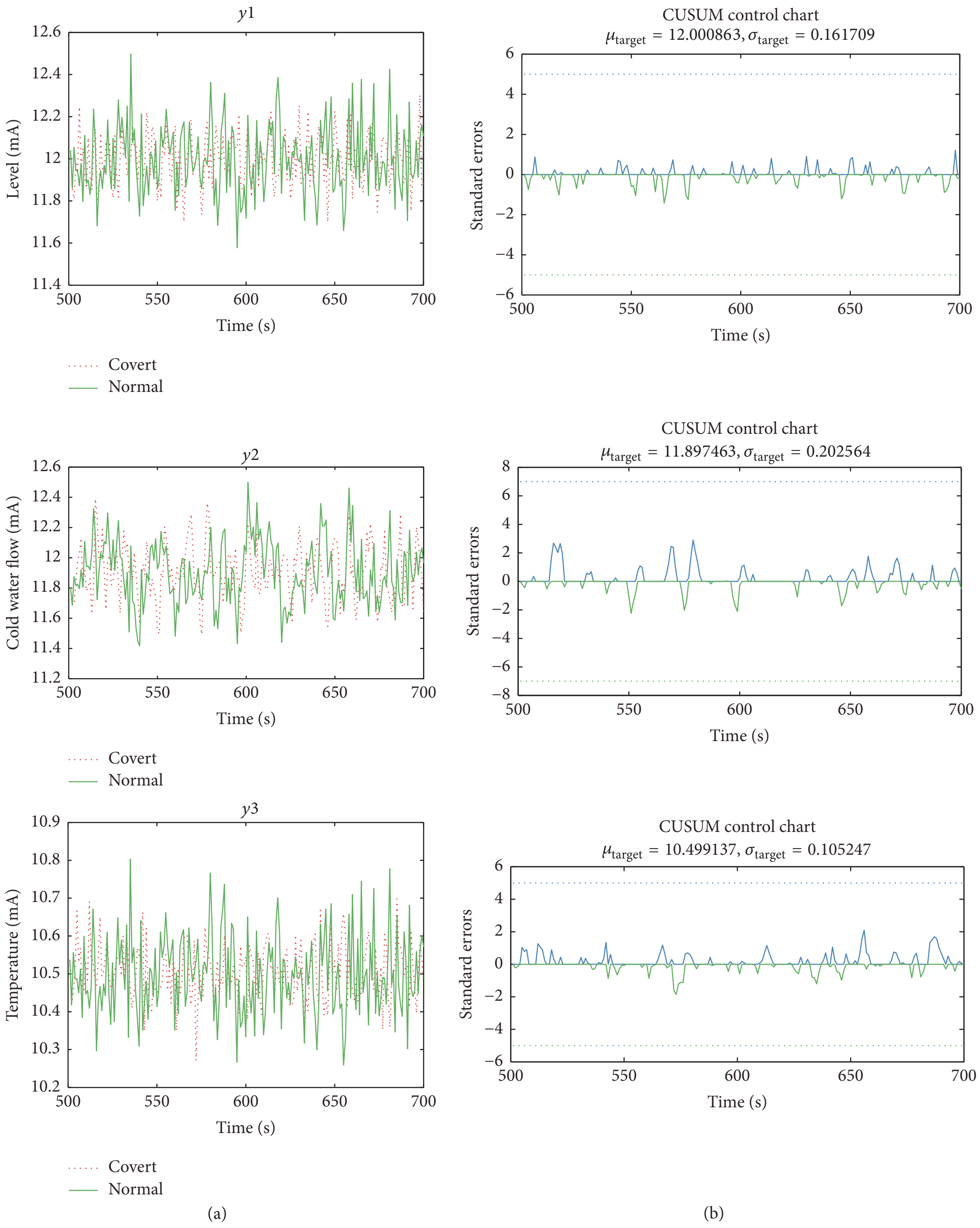
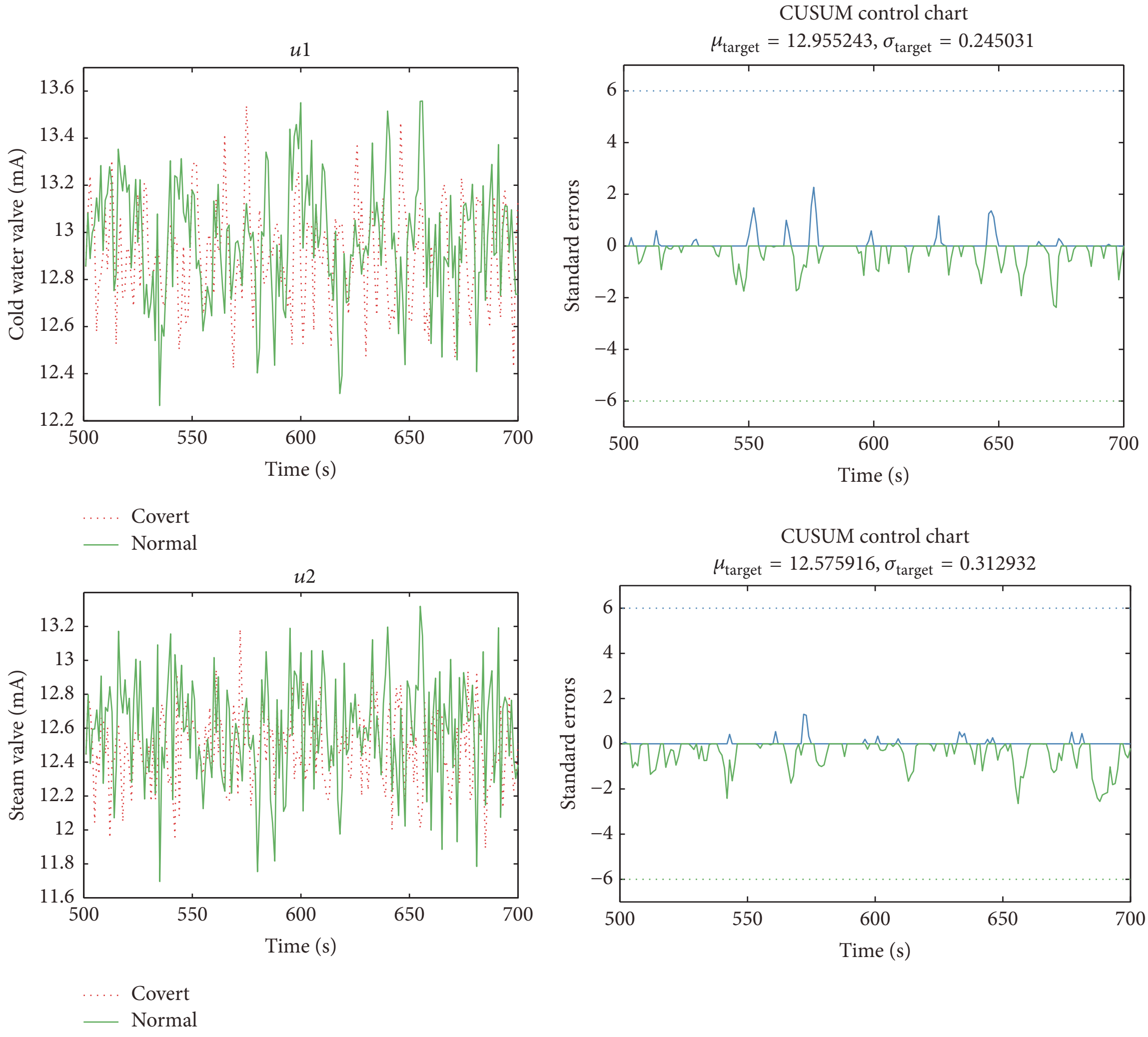


Figure 6: Data from Observer 1 in the covert agent experiment.

(s/h). The "nonsteady" initial phase of the CSTH plant lasts for about 150 seconds (s) and is excluded from all the experiments in this paper.

4.2. Experiment Setup. The CSTH plant depicted in Figure 3 is simulated in Matlab/Simulink, and its execution starts with the predefined base values. The covert agent is constructed based on the LSSVR method, which is available in the free LS-SVMlab toolbox (http://www.esat.kuleuven.be/sista/lssvmlab). In addition, the cumulative sum (CUSUM) algorithm is used to evaluate the stealthy time, which will be introduced in the next subsection. The setups of the



(a)

(b)

Figure 7: Data from Observer 2 in the covert agent experiment.

Table 1: Standard operating conditions.

Variables	Operating conditions (mA)	Variable in simulations
Cold water valve	12.96	$u1$
Cold water flow	11.89	$y2$
Steam valve	12.57	$u2$
Level	12.00	$y1$
Temperature	10.50	$y3$

experiments are illustrated in Figure 5. In order to better assess the stealthiness of the covert agent, we use the replay attack as a comparison and set up two observers to get the experimental data in the simulation. Observer 1 is used to capture the sensor data (i.e., $y1$, $y2$, and $y3$), and Observer 2 is used to capture the output of the controller (i.e., $u1$ and $u2$).

4.3. Assessment Method. In order to evaluate experimental results, the stealthy time τ is used, and it is defined as

$$\tau = t_e - t_s, \tag{22}$$

where t_s is the start time of the covert agent or the replay attack and t_e is the time when an anomaly is detected. A longer stealthy time is favorable to the attackers, as they can have more time to make the physical plant go into an unsafe state while remaining stealthy with respect to the anomaly detectors. In this paper, the anomaly detector is designed based on the CUSUM algorithm, which is one of the most commonly used algorithms for change detection problems [14]. Mathematical details of the CUSUM method can be found in [32].

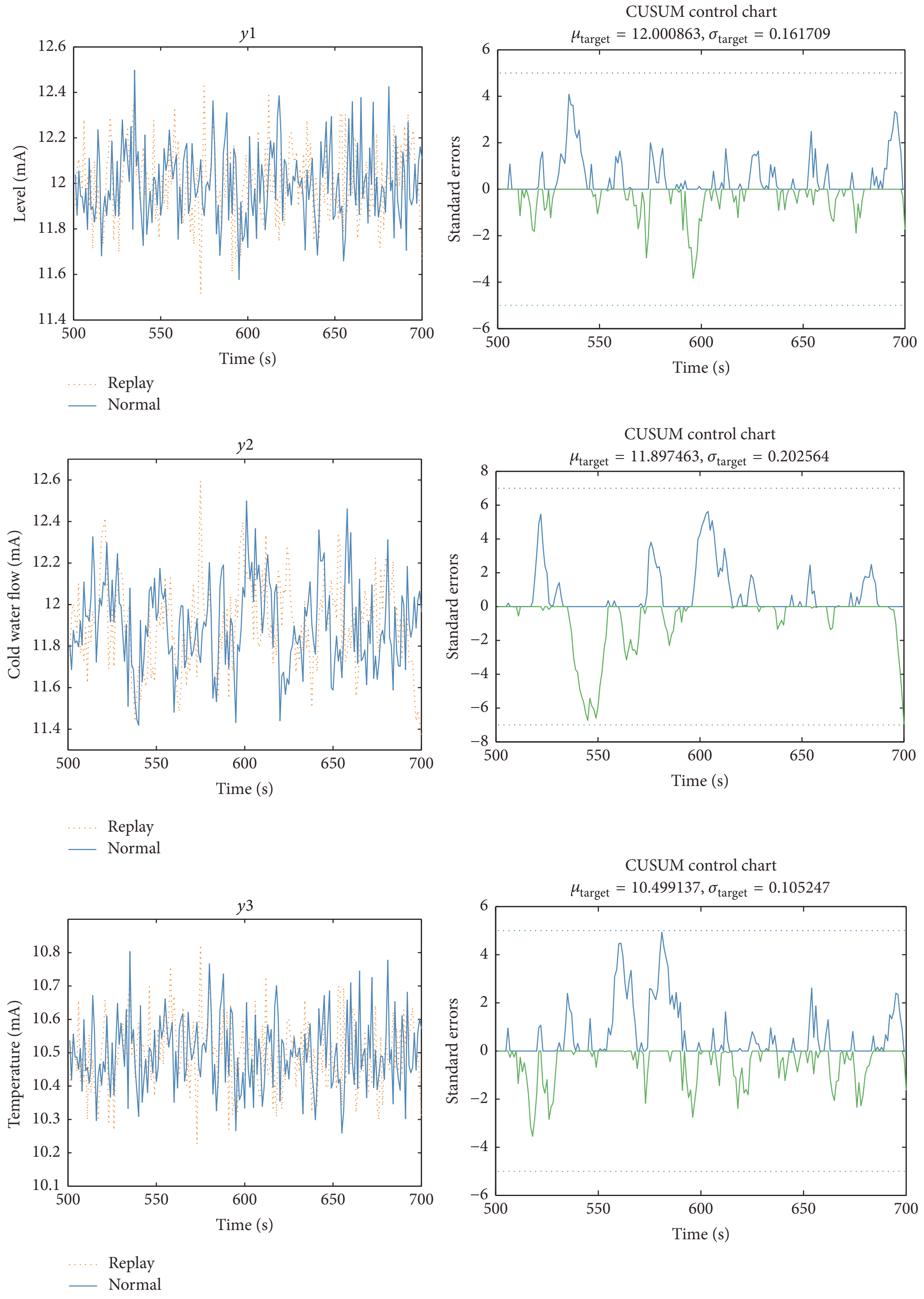


Figure 8: Data from Observer 1 in the replay attack experiment.

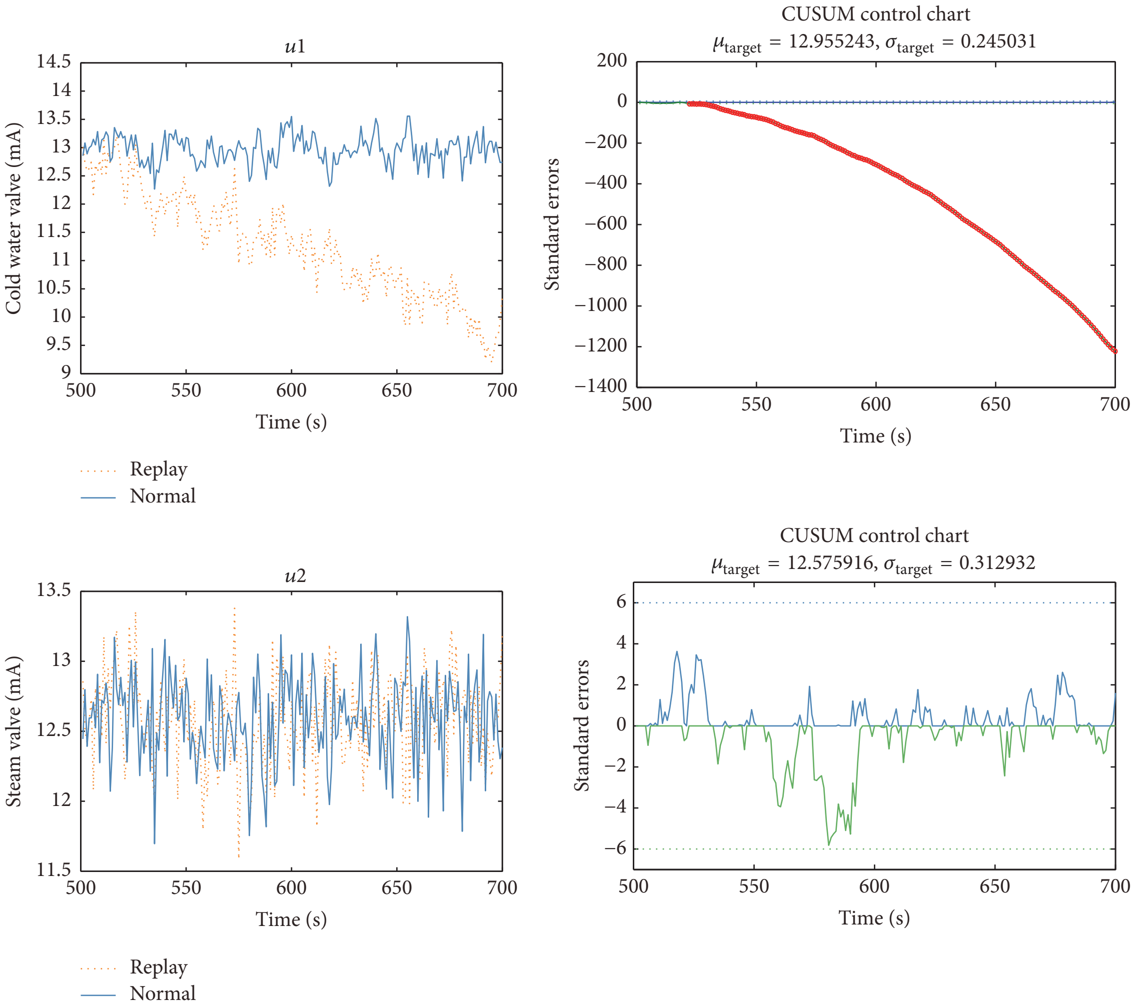


Figure 9: Data from Observer 2 in the replay attack experiment.

5. Experimental Results

In this study, we capture data from the two observers within the time window [201 s, 400 s] in a normal process and use them as the training or replaying data in the experiments. In order to get the statistical results, we run 100 simulations for the covert agent and the replay attack, respectively. In each individual simulation run, the covert agent or the replay attack starts at a random time t, where $500\text{ s} \leqslant t \leqslant 800\text{ s}$ (time is discrete), and persists for 200 seconds.

To get the corresponding stealthy time, the CUSUM algorithm is applied to the data that are obtained from the two observers in the simulations. The thresholds in CUSUM algorithm are determined based on the normal data in the time window ranging from 200 s to 1000 s, and each threshold is selected under the condition that it will not cause any false alarm on the normal data. In this section, we first introduce a covert agent experiment and a replay attack experiment. Then, we give the statistical results of all the experimental tests.

5.1. The Covert Agent and Replay Attack Experiments. In the two experiments, the covert agent and the replay attack are both started at the time $t = 501$ s. Figures 6 and 7 show the results of the covert agent experiment. Figures 6(a) and 7(a) show a comparison of data with and without a covert agent, and Figures 6(b) and 7(b) show the detection of the changes using the CUSUM algorithm. In comparison, Figures 8 and 9 show the results of the replay attack experiment.

From Figures 6 and 8, we can see that the covert agent is able to imitate the behaviors of the three output variables over a finite time window, just like the replay attack does. What is more, the peaks of the CUSUM standard errors in the covert agent experiment are smaller than the ones in the replay attack experiment, which means that the covert agent has better stealthiness and can avoid being detected by the CUSUM with a lower threshold. From Figures 7 and 9, we can see that the covert agent can also keep the control output stealthy, but the replay attack causes anomalies in the controller's output.

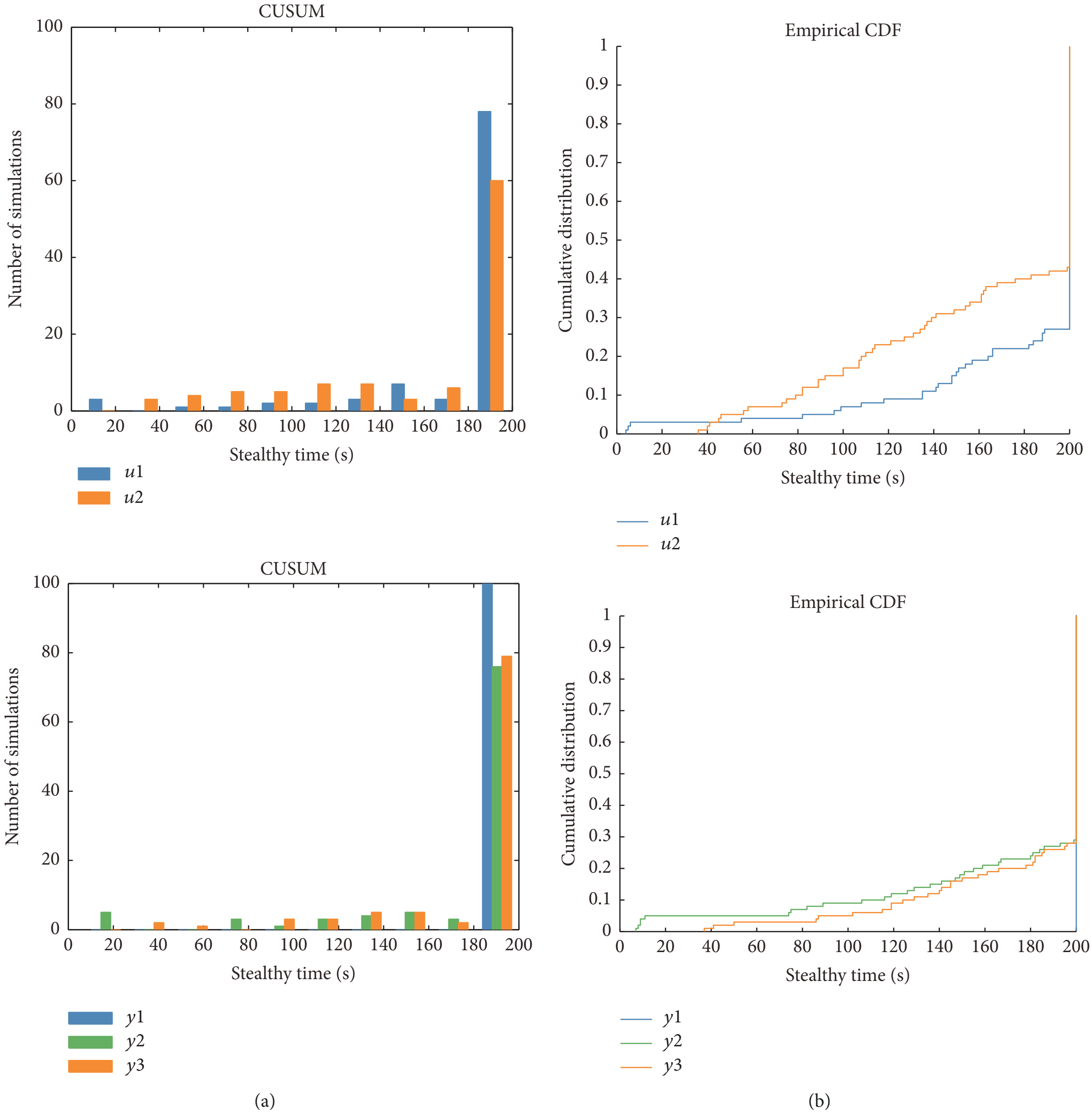


Figure 10: Statistical results of the covert agent experiments.

5.2. Statistical Results. Figure 10 shows the statistical results of the 100 simulations on the covert agent. Figure 10(a) provides the number distributions of the stealthy time by histograms, and Figure 10(b) gives the proportion distributions by the empirical cumulative distribution function (CDF). The empirical CDF $F(x)$ is defined as the proportion of the values less than or equal to x. As can be seen, the stealthy time is longer than 40 seconds in most of the covert agent simulations. Figure 11 shows the statistical results of the 100 simulations on the replay attack. Although the replayed sensor data can avoid being detected by the CUSUM detector, it is more likely to induce an abnormal behavior in the controller's output. More specifically, for the control variable $u1$, the stealthy time is no more than 40 seconds in all the replay attack simulations.

6. Conclusions and Future Work

This paper has investigated the design problem of machine learning based stealthy DI attacks on industrial control systems. A LSSVR-based covert agent has been presented to estimate the model of the physical system, by which attackers can carry out a stealthy DI attack without the need of prior model knowledge of the physical system. The experimental results demonstrate that the covert loop can keep the control output and sensor data both stealthy over a finite time

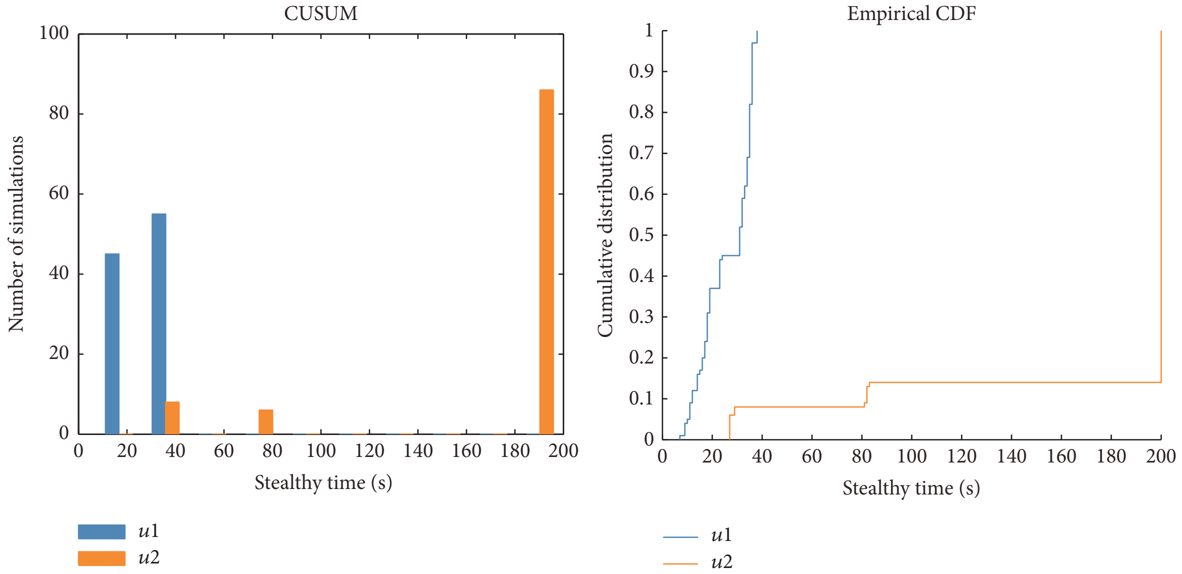


Figure 11: Statistical results of the replay attack experiments.

window. For future work, the proposed covert agent can be further extended to a two-loop covert structure, in which an attack agent can be added. In addition, it is also interesting to investigate the detecting methods of the LSSVR-based attacks.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Key Research and Development Program (no. 2016 YFB1001404) and the National Natural Science Foundation of China (Normal Projects no. 61672093 and no. 61432004).

References

[1] C. Zhou, S. Huang, N. Xiong et al., "Design and analysis of multimodel-based anomaly intrusion detection systems in industrial process automation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 10, pp. 1345–1360, 2015.

[2] T. Cruz, L. Rosa, J. Proenca et al., "A cybersecurity detection framework for supervisory control and data acquisition systems," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2236–2246, 2016.

[3] S. McLaughlin, C. Konstantinou, X. Wang et al., "The Cybersecurity Landscape in Industrial Control Systems," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1039–1057, 2016.

[4] M. Cheminod, L. Durante, and A. Valenzano, "Review of security issues in industrial networks," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 277–293, 2013.

[5] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—attacks, impacts, and defense: a survey," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411–423, 2017.

[6] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: a quantitative risk management approach," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 24–45, 2015.

[7] C. Kwon, W. Liu, and I. Hwang, "Security analysis for Cyber-Physical Systems against stealthy deception attacks," in *Proceedings of the 2013 1st American Control Conference, ACC 2013*, pp. 3344–3349, USA, June 2013.

[8] Z.-H. Pang, G.-P. Liu, D. Zhou, F. Hou, and D. Sun, "Two-channel false data injection attacks against output tracking control of networked systems," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3242–3251, 2016.

[9] A. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: The output-to-output l2-gain," in *Proceedings of the 54th IEEE Conference on Decision and Control, CDC 2015*, pp. 2582–2587, Japan, December 2015.

[10] H. Sedghi and E. Jonckheere, "Statistical structure learning to ensure data integrity in smart grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1924–1933, 2015.

[11] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using Kalman filter," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 370–379, 2014.

[12] A. Dutta and C. Langbort, "Stealthy output injection attacks on control systems with bounded variables," *International Journal of Control*, vol. 90, no. 7, pp. 1389–1402, 2017.

[13] R. S. Smith, "Covert misappropriation of networked control systems: presenting a feedback structure," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 82–92, 2015.

[14] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th International Symposium on Information, Computer and Communications Security (ASIACCS '11)*, pp. 355–366, Hong Kong, March 2011.

[15] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[16] D. I. Urbina, J. Giraldo, A. A. Cardenas et al., "Limiting the impact of stealthy attacks on Industrial Control Systems," in *Proceedings of the 23rd ACM Conference on Computer and Communications Security, CCS 2016*, pp. 1092–1105, Austria, October 2016.

[17] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2226–2238, 2013.

[18] Z. Yu and W. Chin, "Blind false data injection attack using pca approximation method in smart grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, 2015.

[19] A. Anwar and A. N. Mahmood, "Stealthy and blind false injection attacks on SCADA EMS in the presence of gross errors," in *Proceedings of the 2016 IEEE Power and Energy Society General Meeting, PESGM 2016*, USA, July 2016.

[20] A. Anwar, A. N. Mahmood, and M. Pickering, "Modeling and performance evaluation of stealthy false data injection attacks on smart grid in the presence of corrupted measurements," *Journal of Computer and System Sciences*, vol. 83, no. 1, pp. 58–72, 2017.

[21] Y. Yuan and Y. Mo, "Security in cyber-physical systems: Controller design against Known-Plaintext Attack," in *Proceedings of the 54th IEEE Conference on Decision and Control, CDC 2015*, pp. 5814–5819, Japan, December 2015.

[22] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: a data driven approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, 2015.

[23] X. Lu, W. Zou, and M. Huang, "A novel spatiotemporal LS-SVM method for complex distributed parameter systems with applications to curing thermal process," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1156–1165, 2016.

[24] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[25] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 431–438, 2015.

[26] H. C. Jung, J. S. Kim, and H. Heo, "Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach," *Energy and Buildings*, vol. 90, pp. 76–84, 2015.

[27] M. K. Goyal, B. Bharti, J. Quilty, J. Adamowski, and A. Pandey, "Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, fuzzy logic, and ANFIS," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5267–5276, 2014.

[28] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *Proceedings of the 2012 50th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2012*, pp. 1806–1813, USA, October 2012.

[29] G. Wu and J. Sun, "Optimal data integrity attack on actuators in Cyber-Physical Systems," in *Proceedings of the 2016 American Control Conference, American Automatic Control Council (AACC) (ACC '16)*, pp. 1160–1164, USA, July 2016.

[30] C. Staelin, "Parameter selection for support vector machines," Hewlett-Packard Company, 2003, Tech. Rep. HPL-2002-354R1.

[31] N. F. Thornhill, S. C. Patwardhan, and S. L. Shah, "A continuous stirred tank heater simulation model with applications," *Journal of Process Control*, vol. 18, no. 3-4, pp. 347–360, 2008.

[32] L. Koepcke, G. Ashida, and J. Kretzberg, "Single and multiple change point detection in spike trains: Comparison of different CUSUM methods," *Frontiers in Systems Neuroscience*, vol. 10, article no. 51, 2016.

Journal of Engineering
The Scientific World Journal
International Journal of Rotating Machinery
Journal of Sensors
Advances in Multimedia
Advances in Civil Engineering
Journal of Control Science and Engineering
Journal of Robotics
Journal of Electrical and Computer Engineering
Advances in OptoElectronics
VLSI Design
International Journal of Navigation and Observation
Modelling & Simulation in Engineering
International Journal of Aerospace Engineering
International Journal of Chemical Engineering
International Journal of Antennas and Propagation
Active and Passive Electronic Components
Shock and Vibration
Advances in Acoustics and Vibration