Replay Attack Detection for Cyber-Physical Control Systems: A Dynamical Delay Estimation Method

Dong Zhao, *Senior Member, IEEE*, Bo Yang, *Member, IEEE*, Yueyang Li, *Member, IEEE*, and Hui Zhang, *Fellow, IEEE*

***Abstract*—Cyber attack detection plays a crucial role in ensuring cyber security of control systems. However, the reported methods for cyber attack detection often rely on large-scale upgrades of field modules or modification of specific monitoring signals. The resulting issues, such as impacts on system control performance or economic losses, make these methods challenging to be widely accepted in practice. To address this challenge, this article proposes a new replay attack detection scheme, combining dynamical system characteristics and data-driven implementation, while the scheme only needs to be implemented in the control and monitoring side. Specifically, considering the effect of replay attacks and the response of the control system, a dynamical delay estimation method is proposed for detecting replay attacks. Thus, the detection logic is established by characterizing and comparing the fluctuation ranges of dynamical delay. A sliding window is adopted for quantifying the dynamical delay between the input and output data of the cyber-physical control system. Moreover, the randomized algorithm method is introduced to determine the initial value of the delay estimation method, while the online updating of dynamical delay estimation is given for real-time replay attack detection. To highlight the replay attack effect, a novel window adaptive strategy is developed to achieve adaptive detection and tracking of replay attacks. Finally, the performance and effectiveness of the proposed detection method are verified through experiments on chlorosilane distillation.**

***Index Terms*—Cyber-physical systems, delay estimation, replay attack detection.**

I. Introduction

TRADITIONAL industrial control systems have been implemented with an isolated control network, where the information and network security of the industrial control systems has not been fully considered. With the technological explosion of intelligent manufacturing and industrial internet, industrial control systems are becoming more and more open. However, the openness brings not only convenience to the network but also significant cyber security risk, especially the threat of cyber attacks. Thus, cyber attack detection is vital for ensuring the cyber security of control systems and has attracted dramatic attention from both researchers and practitioners.

Generally, cyber attacks on control systems are classified as disclosure attacks, deception attacks, and disruption attacks [1], [2]. Replay attack, which is a combination of the disclosure attack and the deception attack and possesses high stealthiness while requiring little system knowledge for implementation, has became one of the most threatening cyber attacks for control systems, e.g., the well-known Stuxnet event. To deal with the replay attack, many elegant detection methods have been proposed, where the most widely reported ones are based on watermarking, moving target, and coding. The watermarking method was first proposed in [3] for replay attack detection, where the key is to check the distribution property of the detection residual that is characterized by the input noise. Later, the so-called dynamic watermarking method was proposed with the general control system signal distribution property checking [4], [5], [6], and the design of the watermarking signal was also proposed, e.g., the periodic watermarking [7] and the event-triggered watermarking [8]. As the watermarking signal disturbs the control system, the tradeoff between the attack detection and resilient control performance has been studied [9], [10]. The moving target method aims to add extra dynamics to the system being detected, where the information of the extra dynamics is independent of the potential attacker. Generally, stochastic or time-varying properties are introduced for the extra dynamics design to achieve a better defense and detection performance against replay attacks [11], [12], [13]. The coding and encryption scheme is widely adopted to protect communication data security. For control systems, coding and encryption scheme is still the most straightforward method for protecting the transmitted signal and works for replay attack detection.

Manuscript received 19 December 2023; revised 9 April 2024; accepted 9 May 2024. This work was supported in part by the Beijing Natural Science Foundation under Grant 4232047, in part by the National Natural Science Foundation of China under Grant 62273021, Grant 722010216, and Grant 61973135, and in part by A Project of Shandong Province Higher Educational Youth and Innovation Talent Introduction and Education Program. *(Corresponding author: Bo Yang.)*

Dong Zhao is with the School of Cyber Science and Technology, Beihang University, Beijing 100191, China (e-mail: dzhao@buaa.edu.cn).

Bo Yang is with the Department of Management Science and Data Science, Sichuan University, Chengdu 610065, China (e-mail: yangbo_1@scu.edu.cn).

Yueyang Li is with the School of Electrical Engineering, University of Jinan, Jinan 250022, China (e-mail: cse_liyy@ujn.edu.cn).

Hui Zhang is with the School of Transportation Science and Engineering, Beihang University, Beijing 102206, China (e-mail: huizhang285@buaa.edu.cn).

Digital Object Identifier 10.1109/TIE.2024.3406859

For control systems, there are different coding/encryption (and decoding/decryption) realization strategies that may utilize random sensor noise [14], output filtering [15], residual generator gain switching [16], system spectral estimation [17], disturbance compensation [18], and input signal coupling [19].

The reported replay attack detection methods are efficient under specific conditions; however, there are some practical problems to be solved.

1) Most of the industrial control systems are not intelligent. The local computing capability is generally too weak to realize even simple coding/encryption-based defense algorithms. Moreover, large-scale upgrade of the field devices, e.g., actuators and sensors, is not economically feasible.
2) The industrial control system configuration is fixed and the system dynamics modification is usually not allowed, especially for those safety-critical control systems (e.g., where reliability reassessing may be required).
3) Control performance is the core concern of the user, and thus the control performance loss due to the signal added for security monitoring may not be acceptable.

Concerning these practical challenges and the state of the art of the replay attack detection methods, a new replay attack detection perspective is needed.

Generally, data from industrial control systems are expressed as time series. Recalling the implementation of replay attack, where the historical data is used to replace the real-time data, the replay attack can be viewed as a special kind of time series anomaly and thus may be detected based on time-series-based anomaly detection. Traditional time-series-based anomaly detection mainly focuses on data stochastic property analysis, parameter estimation, and model prediction [20], [21]. Moreover, machine learning techniques have been widely used for anomaly detection of time series, for example, the support vector machine and isolation forest algorithm, as well as their improved versions [22], [23]. Regarding different types of anomalies, research effort has been given on time-series-based fault detection and classification [24], [25]. Moreover, the defense of adversarial attack and data poisoning of time series are reported as well [26], [27]. With the blooming development of deep learning, novel neural networks like graph neural networks, generative adversarial network, convolutional neural network, and long short-term memory network (LSTM) have been used for the time-series-based anomaly detection with sufficient applications, e.g., fraud detection [28], leak detection [29], process fault diagnosis [30], [31], and arrhythmia classification [32].

Although time-series-based anomaly detection has been reported with various applications, more effort is still needed regarding the dynamics of control systems and the replay attack detection. Different from the above-mentioned time-series-based anomaly detection tasks, a replay attack is highly stealthy (healthy data is used for replay), where the attack-induced change of data characteristics is not conspicuous enough for using the reported time-series-based anomaly detection method. Deep neural networks are powerful for mining the correlation of different variables/time series [33]; however, the high network training cost is not always tolerable, particularly for large-scale industrial control systems with heterogeneous dynamics. Most of the time-series-based anomaly detection methods focus on the single time series itself and ignore the correlation of different time series; however, characterized by the dynamics of the control loop, the correlation of the control system time series, e.g., input and output, is hard to fake by the attacker. Moreover, the data correlation can be quantified as a time delay when considering the order and dynamics of control systems, where it usually exists a delayed transitive relation between the input and output data. This delay is a comprehensive index that determined by both the detector quantification window and the system dynamics, which means that it is "encrypted" naturally against the attacker. Since the delay attack will induce the change of this delay, it is ideal for replay attack detection from the control and monitor side.

Keeping the implementation requirement of replay attack detection in mind, the time-series-based replay attack detection method is motivated; considering the state of the art of the time-series-based anomaly detection and the control system characteristic, a dynamical delay estimation (DDE) method, which is based on the data correlation characterized by the control system dynamics, is developed for replay attack detection. First, the delay range and data window length between system input and output data are quantified. Then, online delay estimation method is given, and an adaptive window length estimation algorithm is proposed. As the occurrence of a replay attack changes the delay of control system data, the replay attack detection can be achieved by referring to the quantified control system delay range. To enhance the detectability of the replay attack, an event-triggered strategy for determining the detection window length is proposed.

The contribution of this study is summarized as follows.

1) A new replay attack detection scheme, combining dynamical system characteristics and data-driven implementation, is first proposed, and the scheme requires to be implemented in the cyber-part of the control system only;
2) A new DDE method is proposed, where the delay quantification and adaptive attack effect tracking are given; and
3) A real experimental verification is given to show the performance and effectiveness of the proposed replay attack scheme.

Compared with the reported replay attack detection method, no control performance loss, local physical dynamics change, extra local computation, detailed model information, or extensive training are needed. All these properties ensure the feasibility and acceptability of the proposed method for industrial cyber-physical systems.

This article is organized as follows. The problem formulation is given in Section II, and the proposed detection scheme is proposed in Section III. In Section IV, an experimental example is used to illustrate the performance of the proposed scheme. Section V concludes this article.

II. Preliminaries

In this section, the system setting, the introduction of replay attack, and the problem to be solved are given.

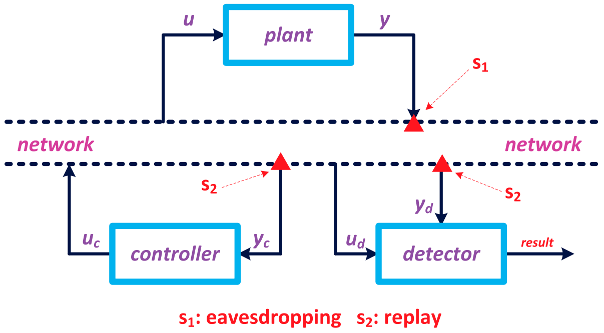

Fig. 1. Control system setting.

A. Cyber-Physical System Setting

The cyber-physical control system setting is given in Fig. 1, where the physical plant is connected with the remote control and monitoring modules via a network. The controller output u_c is transmitted to the physical plant as well as the detector, and the physical plant output y is transmitted to the controller and detector. To focus on the detection method, assume that the closed-loop control system is in a steady state. In the absence of a replay attack, one has

$$y(i) = y_d(i) = y_c(i),\ u(i) = u_d(i) = u_c(i) \tag{1}$$

where u_d and u_c are the received control effort for the detector and the controller output, respectively; y_d and y_c are the received output for the detector and controller, respectively.

In this article, the dynamical models of the considered plant and its adopted controller are not assumed, as a data-driven replay attack detection method is proposed.

B. Replay Attack

For replay attack, the attacker has access to the system output transmission and can modify the system input and output transmissions. Generally, a replay attack for cyber-physical systems posses a two-stage implementation, i.e., eavesdropping (s_1) and replay (s_2).

1) s_1: record sensor output transmission for a time period $[t_{e1}, t_{e2}]$, where $t_{e1}, t_{e2} \in \mathcal{N}^+$.
2) s_2: replace sensor output transmission since $i \geq T_0$ by the recorded data, where $T_0 \gg t_{e2}$ and $T_0 \in \mathcal{N}^+$.

In the presence of data replay, one has

$$y_d(i) = y_c(i) = y(\tau_i),\ u_d = u_c \approx u(\tau_i) \tag{2}$$

where $\tau_i \in \mathcal{N}^+$ and $t_{e1} < \tau_i \ll t_{e2}$. In the presence of replay attack, the controller will receive the delayed measurement. If a static controller is adopted, we will have $u_c = u(\tau_i)$; otherwise, $u_c \approx u(\tau_i)$. To increase the damage capacity of replay attack, an extra control input attack signal may be added to $u(i)$. The design of the control input attack signal is omitted here as this signal is not used for the replay attack detection.

C. Problem Formulation

Considering the control system illustrated in Fig. 1, with the data u_d and y_d, we will design an appropriate detection scheme against replay attack.

Due to the plant dynamics, the system output series is strongly related to the system input series, which is the intrinsic property of the dynamical control system. Specifically, the correlation between system input and output series is characterized by the detailed system dynamics (e.g., the output solution with respect to the system input). Note that it is hard to fake the correlation between system input and output series, because this correlation is characterized by system dynamics, varying with relative steps of series, and reflected by the stochastic properties instead of a specific value on y_d or u_d. After the occurrence of a replay attack, the correlation between system input and output series is changed, even if it seems that the pair (y_d, u_d) remains almost unchanged in value from the aspect of the detector. The almost unchanged value pair raises difficulties for replay attack detection and ensures stealthiness of attack implementation. Considering the replay attack detection requirement and the detector implementation limitation of cyber-physical systems, different from the reported watermarking, moving target, and coding methods, we try to detect replay attacks by measuring the correlation between input and output series. Replay attack will cause a delay in data, which leads to a change of the delay between input and output series. Under this circumstance, we turn to measure the dynamical delay between input and output series, which is a direct index of the mentioned correlation. As the replay attack detection scheme is based on the delay measuring, the key problems to be solved are as follows.

1) How to estimate the delay for the closed-loop control systems?
2) How to quantify the delay for replay attack detection?

To answer the above two questions, the DDE, dynamical window length determination, online delay feature capture before and after the occurrence of a replay attack, and delay feature enhancement strategies are studied.

III. Methodology

In this section, the offline learning of delay range and detection window length is given, and the online DDE method is proposed. Finally, the delay sensing alarm strategy is established.

A. Global Delay Estimation and Sequence Moving Range Determination

Time delay, in the context of time series analysis, refers to the temporal discrepancy between observed values. It serves as a crucial metric for characterizing temporal dependencies between two variables, e.g., y_d and u_d. Time delay is a direct reflection of the correlation between variables. To measure this impact, the global delay estimation method is frequently employed to rectify temporal misalignments in the data. Among these methods, one of the most effective and convenient approaches is based on the maximum cross-correlation function [34], [35].

Assuming that a and b are two time series variables with n observations. σ_a and μ_a are the standard deviation and expectation of a, respectively. σ_b and μ_b are the standard deviation and expectation of b, respectively. The association coefficient $\phi_{ab}(k)$ is given by the following equation:

$$\phi_{ab}(k) = \frac{E[(a_i - \mu_a)(b_{i+k} - \mu_b)]}{\sigma_a \sigma_b} \tag{3}$$
$$k = -n+1, \ldots, n-1, \quad n,\ i \in Z^+.$$

The correlation value is obtained by calculating the sample expectation

$$\hat{\phi}_{ab}(k) = \begin{cases} \frac{1}{n-k} \sum_{i=1}^{n-k} \frac{(a_i-\mu_a)(b_{i+k}-\mu_b)}{s_a s_b}, & k \geq 0 \\ \frac{1}{n+k} \sum_{i=1-k}^{n} \frac{(a_i-\mu_a)(b_{i+k}-\mu_b)}{s_a s_b}, & k < 0 \end{cases} \tag{4}$$

where s_a and s_b represent the sample standard deviations of a and b, respectively.

The idea behind the cross-correlation function method is to assume the existence of a specific global delay between each sequence, and the maximum absolute value $\hat{\phi}_{ab}(k)$ is identified as the correlation coefficient. The moment when the correlation coefficient reaches its maximum is the overall delay approximation between the sequences. The maximum and minimum value of the correlation coefficient can be calculated as follows:

$$\phi^{\max} = \max\{\phi_{ab}(k), 0\} \geq 0 \tag{5}$$
$$\phi^{\min} = \min\{\phi_{ab}(k), 0\} \leq 0. \tag{6}$$

Based on (5) and (6), the corresponding label $k^{\max}$ and $k^{\min}$ can be determined. Then, the delay between sequences is given by the following equation:

$$\lambda = \begin{cases} k^{\max}, & \phi^{\max} \geq -\phi^{\min} \\ k^{\min}, & \phi^{\max} < -\phi^{\min}. \end{cases} \tag{7}$$

The direction of the delay between variables a and b is determined by the sign of λ: when λ is greater than zero, the direction of the delay between variables a and b is from a to b; when λ is smaller than zero, the direction of the delay between variables a and b is from b to a. The real correlation coefficient with delay is $\phi_{ab}(\lambda) \in [-1, 1]$. The positive or negative definite of $\phi_{ab}(\lambda)$ determines whether the correlation between variables is positive or negative.

Theoretically, the absolute value range of parameter k in (3) is from 0 to $n-1$. However, as $|k|$ increases, the portion $l_{ab} = n - |k|$, to be calculated for both sequences, gradually decreases. When l_{ab} reduces to a certain extent, the length consistency of the sliding sequences cannot be guaranteed, and the correlation coefficient may become abnormal (e.g., unusually high), which is not consistent with the practical detection requirements and may affect subsequent computation results. Therefore, it is necessary to establish a suitable range for the sliding between sequences, where the maximum value of the sliding range should be less than the real maximum transmission delay size $T_{\max}$. If $T_{\max}$ cannot be directly obtained, an estimation is given: $\hat{T}_{\max} = \min\left(\tilde{T}_{\max}, n/m\right) \geq T_{\max}$, where $\tilde{T}_{\max}$ is the maximum transmission time estimation based on process knowledge, n is the sufficiently large sample length, and m can be determined based on the number of modes in which the system operates. In the absence of prior knowledge, m can be initialized to 2 and fine-tuned based on computational results.

B. Dynamical Delay Estimation

Algorithm 1 shows the process of estimating a dynamical delay based on the sliding window method. The detailed procedures and analysis are given as follows.

1) To acquire dynamical delay information between variables, it is first essential to establish a sliding window. The specific equation for the sliding correlation is as follows:

$$\hat{\phi}(t, k_t) = \frac{1}{l} \sum_{i=t-r}^{t+r} \left(a_i - \mu_a\right)\left(b_{i+k_t} - \mu_b\right)$$
$$\Big/ s_a s_b,\ k_t \in (0, \tilde{T}_{\max}),\ t \in (r, n-r) \tag{8}$$

where t represents the time instant, l represents the size of the sliding window, k_t denotes the displacement of b relative to a at time t, r represents the observation radius of the sliding window, and $l = 2r+1$; μ_a and μ_b, respectively represent the mean values of a and b, and s_a and s_b respectively represent the sample standard deviations. Note that these means and deviations can be time-varying.

2) In this context, we assume that the direction of information transmission remains constant within the same working mode, which is reasonable for control systems in a steady state (to ensure the stealthiness of replay attacks). The direction of k_t can be determined based on the overall delay direction. Therefore, the maximum value of the sliding correlation coefficient leads to the dynamical delay estimation, as expressed by the following equation:

$$\phi(t)^{\max} = \max_{k_t=0}^{\tilde{T}_{\max}} [\hat{\phi}(t, k_t)]. \tag{9}$$

At the same time, the corresponding label $k_t^{\max}$ is obtained, and the DDE $\hat{\lambda}(t)$= $k_t^{\max}$.

3) The initial size of the sliding window l can be determined based on the Randomized Algorithms [36]. First, the estimation function is defined as follows:

$$\hat{p}(\gamma) = \Pr(|\lambda(t)| \leq \gamma) \tag{10}$$

and

$$\Pr\left(\left|p(\gamma) - \hat{p}(\gamma)\right| < \varepsilon\right) \geq 1 - \delta \tag{11}$$

where $\gamma = \max_{t=r}^{n-r}[\hat{\lambda}(t)]$, $\varepsilon \in (0, 1)$ is the given accuracy requirement, and $1-\delta$ is the corresponding confidence level with $\delta \in (0, 1)$. According to the Hoeffding inequality, (11) is derived that

$$l \geq \frac{1}{2\varepsilon^2} \log \frac{1}{\delta} \Rightarrow \Pr\left(p(\gamma) < \hat{p}(\gamma) + \varepsilon\right) > 1 - \delta. \tag{12}$$

To enhance the offline computational efficiency, preset l based on (12) with ε and δ. Then, gradually increase l until

Algorithm 1: DDE offline

```
input  : two time series variables $a, b$ under normal
         conditions
output: the dynamical delay sequence $\lambda(t)$ between $a, b$
Initialization: $\hat{T}_{\max}$, $l$
for $t \leftarrow r$ to $n - r$ do
    for $k_t \leftarrow 1$ to $\hat{T}_{max}$ do
        for $l \leftarrow l_{\min}$ to $l_{\max}$ do
            calculate the value $\hat{\phi}(t, k_t)$ in (8);
            if $\phi(t)^{\max} > \phi_{ab}(\lambda)$ then  save the value
              $\phi(t)^{\max}$; $\hat{\lambda}(t)$= $k_t^{\max}$;
            break ; // Check the condition
              to break
            else $l++$// Increment $l$ ;
        end
    end
end
```

Algorithm 2: Dynamical delay sensing alarm strategy

```
Input  : two time series variables $a, b$
Output: detection flag function $J(t)$
Initialization: $\hat{T}_{\max}$, $l_{min}$, $l_{max}$, $\hat{\lambda}_{\min}$, $\hat{\lambda}_{\max}$
for $t \leftarrow r$ to $n - r$ do
    for $k_s \leftarrow 1$ to $\hat{T}_{max}$ do
        for $l \leftarrow l_{\min}$ to $l_{\max}$ do
            calculate $\hat{\phi}(t, k_t)$ in (8) and $J(t)$ in (13);
            if $\phi(t)^{\max} > \phi_{ab}(\lambda)$ then
                if $J(t) = 1$ and $l > l_{max}$ then
                    Output: $J(t) = 1$;
                end
                else if $J(t) = 1$ and $l \leq l_{max}$ then
                  $l++$ // increment $l$;
                else  Output: $J(t) = 0$;
                break// Break out of the
                  loop ;
            end
        end
    end
end
```

$\phi(t)^{\max} > \phi_{ab}(\lambda)$, where $\phi_{ab}(\lambda)$ is the overall correlation coefficient.

Remark 1: By introducing a dynamic time window updating mechanism and employing the RA method to optimize the determination of key hyperparameters, our approach effectively adapts to data changes and captures dynamic time delay information between variables, finding optimal parameter settings across various scenarios. These innovations not only enhance the accuracy and adaptability of time delay estimation and increase work efficiency, but also simplify the implementation process and expand the application range of the method.

C. *Dynamical Delay Sensing Alarm Strategy*

The dynamical delay sequence between the input and output of the control system is measured by the dynamical delay estimation $\hat{\lambda}(t)$. For replay attack detection, it is necessary to determine the delay range $\hat{\lambda} \in [\hat{\lambda}_{\min}, \hat{\lambda}_{\max}]$ and the size range $l \in [l_{\min}, l_{\max}]$ under the attack-free case.

It is known that a replay attack will change the delay between the input and output of control systems. Based on the obtained delay range under an attack-free case, a flag function J is introduced

$$J(t) = \begin{cases} 1, & \hat{\lambda}(t) \notin [\hat{\lambda}_{\min}, \hat{\lambda}_{\max}] \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

Thus, $J(t) = 1$ implies the existence of a replay attack, as the delay between the input and output of the considered control system falls outside of the delay range for the attack-free case.

Due to the feedback control property, i.e., u is generated by the controller based on y, the effect of the replay attack shows typical finite time characteristics. Thus, we propose an "event-triggered" varying scheme for the sliding window length. The initial size of the sliding window online l_o is set to $l_{\min}$, and the current moment's estimated value is calculated. If the estimated value falls within the normal range, subsequent time instants are processed. In the case where the estimated value exceeds the normal range, l_o gradually increases. This is for confirming the alarm information and establishing an adaptive attack feature tracking scheme. When $l_o > l_{\max}$ and $\hat{\lambda}(t)$ still falls outside the normal range, trigger an alarm of replay attack. The procedure for dynamical delay sensing is summarized in Algorithm 2.

Remark 2: We discuss the computational complexity of the proposed method. The input data is assumed to have a dimension of n. The time complexity of the proposed method is $O(n * \hat{T}_{\max} * (l_{\max} - l_{\min} + 1))$, and the space complexity is $O(1)$. This means that the execution time of the algorithm is mainly affected by the input data size n, $\hat{T}_{\max}$, and $l_{\max} - l_{\min}$, while the storage space required by the Algorithm is constant. To compare the time and space complexities of related methods: One-class support vector machine (SVM) operates at $O(n^3)$ for time complexity and $O(n * d)$ for space complexity. LSTM's time complexity is $O(n * d * h^2)$, with space complexity at $O(d^2 * h)$. Isolation Forest has a time complexity of $O(n * d * n_{\text{trees}} * \max_{\text{depth}})$ and space complexity of $O(n_{\text{trees}} * n)$. Local outlier factor (LOF) features a time complexity of $O(n^2)$ and space complexity of $O(n * d)$. Here, n represents the number of samples, d represents the dimensionality of the feature space, h denotes the number of hidden units, n_{trees} is the number of trees built, and $\max_{\text{depth}}$ is the maximum depth of the trees. It is evident that the proposed method requires less computations than the methods reported, especially when n is large.

IV. Experiments

A. *Experimental Setup*

The experimental system is from a real distillation column. The field diagram of the distillation column example is shown

Fig. 2. Field diagram of the distillation column.

in Fig. 2. The data were collected from the liquid level sensor LIC301 on the tower reactor kettle of the three-stage distillation column. The process flow of the rectification column is illustrated in Fig. 3. The data of 7000 time units were used for offline learning. As the attacked system output and input are required for online detection, we first record attack-free data from the tower reactor kettle liquid level LIC301.PV and the corresponding LIC301.MV starting from a setting time 0; for the replay attack injection, when $t \geq 12\,000$, we replace the real-time LIC301.PV by the recorded measurement data with a time lag of 3000 and send the replayed LIC301.PV to the controller device to get the corresponding LIC301.MV.

Some representative anomaly detection methods are used for comparison studies to demonstrate the effectiveness and superiority of the proposed method (DDE) in detecting replay attacks. One-class SVM: maps normal data points from their original space to a higher dimensional space and find a hyperplane to distinguish between normal and abnormal data points. Isolation Forest: isolates abnormal data points by building random trees. LSTM: predicts the new time series data and compare the error between the predicted and actual values. LOF: determines whether a point is an anomaly by comparing the density of each point with that of its neighbors.

One-class SVM, isolation forest, and LOF use the source program of the sklearn package, and LSTM uses the source program of TensorFlow. The parameter settings of the proposed method are shown in Table I.

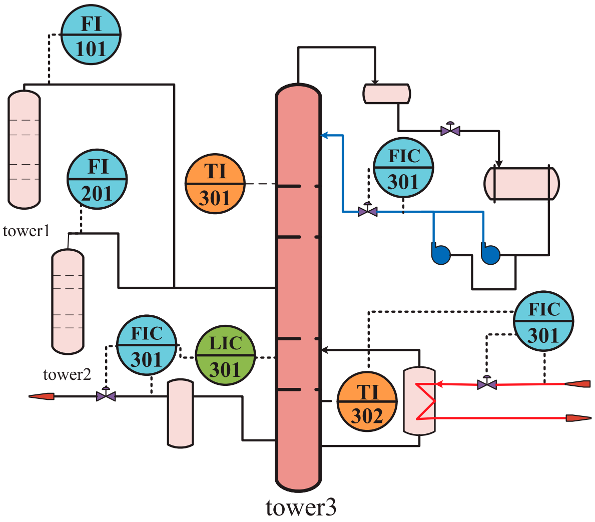

Fig. 3. Process flow chart of the distillation column.

TABLE I
HYPERPARAMETER SETTINGS FOR DDE

Hyperparameter	Value
Initial value of the mobile range k	0
Maximum value of the mobile range k	200
Initial value of the sliding window radius r	50
Maximum value of the sliding window radius r	300

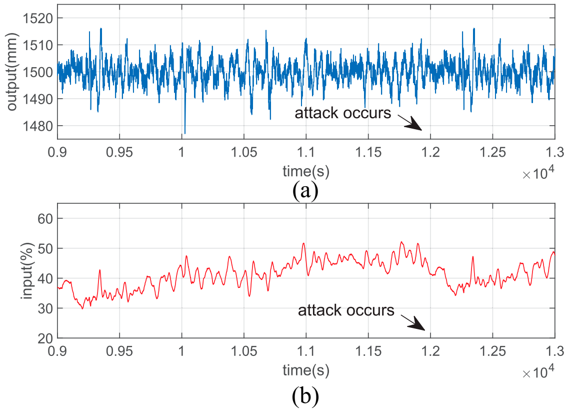

Fig. 4. Attacked control system input and output. (a) Output LIC301.PV; (b) Input LIC301.MV.

Four commonly used classification evaluation indicators (accuracy, precision, recall, and F1-score) are adopted to measure the comprehensive performance of the above-mentioned detection method. To test the robustness of the proposed method, the experiment is repeated 100 times with different random interference environments.

B. Performance Comparison

In Fig. 4, the attacked data of the system input and output are illustrated. The considered replay attack is stealthy, where it is hard to find the attack trace in the data directly. Fig. 5 shows

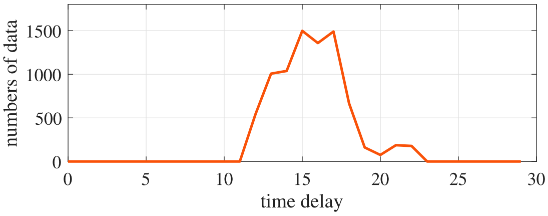

Fig. 5. Statistical results of normal delay range.

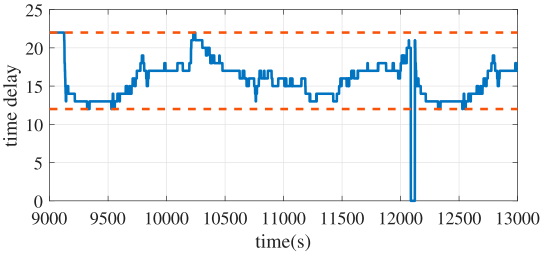

Fig. 6. Result of detecting replay attack. Vertical axis represents the information of the time delay sequence extracted based on DDE.

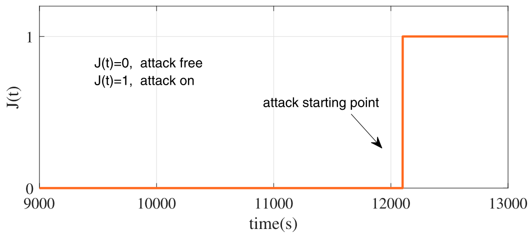

Fig. 7. Result of detecting replay attack.

the statistical (learning) results of the normal delay range from 11 to 23 s in the rectification tower experiment by the proposed DDE method. Figs. 6 and 7 illustrate the online estimation and the flag function response, respectively.

Based on Figs. 5 and 6, one can find that the testing estimation result (normal range $t < 12\,000$) in Fig. 6 matches with the result in Fig. 5. Due to the finite-time scheme and the closed-loop property that the control input is generated based on the output, the abnormal delay response induced by the replay attack exists only within a short time window; however, the proposed detection method captured this abnormal response successfully, as shown by Figs. 6 and 7. To confirm the existence of a replay attack, we adopted the sliding window scheme again and used the adaptive window length strategy. Finally, as can be seen from Fig. 7, the attack detection flag function illustrates a satisfactory attack tracking performance.

Moreover, from Figs. 6 to 7, one can also find the detection delay by comparing the timeline of the alarm response and the implemented replay attack. To tradeoff the precision and timeliness, a window adaptive amplification strategy has been incorporated into the identification strategy. The system can maintain high accuracy while minimizing latency as much as possible. Table II shows the indices of the comparative experiments, where traditional methods exhibit a low detection rate and a high number of false positives, which makes it difficult to effectively assess their detection latency. These experimental results demonstrate the effectiveness of the proposed method.

TABLE II
COMPARATIVE RESULTS WITH DIFFERENT DETECTION METHODS

Approaches	Accuracy	Precision	Recall	F1-Score
One-class SVM	0.372	0.357	0.741	0.482
LSTM	0.542	0.688	0.365	0.477
Isolation forest	0.489	0.420	0.722	0.531
LOF	0.401	0.412	0.806	0.572
DDE	0.907	0.998	0.839	0.912

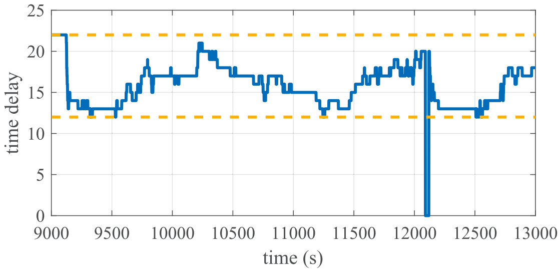

Fig. 8. Result of detecting replay attack with signal transmission noise.

Network-induced disruptions, e.g., disturbances or noise, are crucial issues that need attention. In our experimental study, we deliberately introduced signal transmission noise and presented the corresponding results in Fig. 8. The results indicate that disruptions caused by the network do not affect the effectiveness of the proposed identification method.

V. Conclusion

In this article, we propose a dynamic time-delay-based replay attack detection method. The method acquires input and output data from various consecutive time instants in a general industrial control system and initializes the moving range and sliding window size. Global correlation coefficients and delay estimations are computed for each data point. The maximum correlation coefficient is selected, and the dynamic delay sequence between the input and output data is calculated. Replay attacks are detected based on whether the desired delay falls within the normal delay range for input and output data. By gradually increasing the initial values of the sliding window, the attack characteristics are further enhanced to confirm alarm information. Through comparisons with other methods using the real distillation column process data set, the proposed method is demonstrated to be effective in identifying replay attacks. In the future, the tradeoff between the false alarm rate and detection rate can be further studied.

References

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

[2] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annu. Rev. Control*, vol. 47, pp. 394–411, 2019.

[3] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2009, pp. 911–918.

[4] B. Satchidanandan and P. R. Kumar, "Dynamic watermarking: Active defense of networked cyber-physical systems," *Proc. IEEE*, vol. 105, no. 2, pp. 219–240, Feb. 2017.

[5] M. Porter, P. Hespanhol, A. Aswani, M. Johnson-Roberson, and R. Vasudevan, "Detecting generalized replay attacks via time-varying dynamic watermarking," *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 3502–3517, Aug. 2021.

[6] L. Shangguan et al., "Dynamic watermarking for cybersecurity of autonomous vehicles," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 11 735–11 743, Nov. 2023.

[7] C. Fang, Y. Qi, P. Cheng, and W. X. Zheng, "Optimal periodic watermarking schedule for replay attack detection in cyber-physical systems," *Automatica*, vol. 112, 2020, Art. no.108698.

[8] D. Du, C. Zhang, X. Li, M. Fei, and H. Zhou, "Attack detection for networked control systems using event-triggered dynamic watermarking," *IEEE Trans. Ind. Inform.*, vol. 19, no. 1, pp. 351–361, Jan. 2023.

[9] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Proc. 52nd IEEE Conf. Decis. Control*, 2013, pp. 1854–1859.

[10] S. Weerakkody, Y. Mo, and B. Sinopoli, "Detecting integrity attacks on control systems using robust physical watermarking," in *Proc. 53rd IEEE Conf. Decis. Control*, 2014, pp. 3757–3764.

[11] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, 2015, pp. 5820–5826.

[12] P. Griffioen, S. Weerakkody, and B. Sinopoli, "An optimal design of a moving target defense for attack detection in control systems," in *Proc. Amer. Control Conf. (ACC)*, 2019, pp. 4527–4534.

[13] P. Griffioen, S. Weerakkody, and B. Sinopoli, "A moving target defense for securing cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 66, no. 5, pp. 2016–2031, May 2021.

[14] D. Ye, T.-Y. Zhang, and G. Guo, "Stochastic coding detection scheme in cyber-physical systems against replay attack," *Inf. Sci.*, vol. 481, pp. 432–444, May 2019.

[15] R. M. Ferrari and A. M. Teixeira, "A switching multiplicative watermarking scheme for detection of stealthy cyber-attacks," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2558–2573, Jun. 2021.

[16] S. X. Ding, L. Li, D. Zhao, C. Louen, and T. Liu, "Application of the unified control and detection framework to detecting stealthy integrity cyber-attacks on feedback control systems," *Automatica*, vol. 142, 2022, Art. no. 110352.

[17] B. Tang, L. D. Alvergue, and G. Gu, "Secure networked control systems against replay attacks without injecting authentication noise," in *Proc. Amer. Control Conf. (ACC)*, 2015, pp. 6028–6033.

[18] C. Trapiello, D. Rotondo, H. Sanchez, and V. Puig, "Detection of replay attacks in CPSs using observer-based signature compensation," in *Proc. 6th Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, 2019, pp. 1–6.

[19] H. Guo, Z.-H. Pang, J. Sun, and J. Li, "An output-coding-based detection scheme against replay attacks in cyber-physical systems," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 68, no. 10, pp. 3306–3310, Jun. 2021.

[20] Y. Liu, Z. Chen, L. Wei, X. Wang, and L. Li, "Braking sensor and actuator fault diagnosis with combined model-based and data-driven pressure estimation methods," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 11 639–11 648, Nov. 2023.

[21] Q. Xie, G. Tao, C. Xie, and Z. Wen, "Abnormal data detection based on adaptive sliding window and weighted multiscale local outlier factor for machinery health monitoring," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 11 725–11 734, Nov. 2023.

[22] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discovery Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 2012.

[23] X. Ni, D. Yang, H. Zhang, F. Qu, and J. Qin, "Time-series transfer learning: An early stage imbalance fault detection method based on feature enhancement and improved support vector data description," *IEEE Trans. Ind. Electron.*, vol. 70, no. 8, pp. 8488–8498, Aug. 2023.

[24] Z. Chen, R. Guo, Z. Lin, T. Peng, and X. Peng, "A data-driven health monitoring method using multiobjective optimization and stacked autoencoder based health indicator," *IEEE Trans. Ind. Inform.*, vol. 17, no. 9, pp. 6379–6389, Sep. 2021.

[25] Z. Chen, H. Ke, J. Xu, T. Peng, and C. Yang, "Multichannel domain adaptation graph convolutional networks-based fault diagnosis method and with its application," *IEEE Trans. Ind. Inform.*, vol. 19, no. 6, pp. 7790–7800, Jun. 2023.

[26] Y. Zhuo, Z. Yin, and Z. Ge, "Attack and defense: Adversarial security of data-driven FDC systems," *IEEE Trans. Ind. Inform.*, vol. 19, no. 1, pp. 5–19, Jan. 2023.

[27] Y. Chen, X. Zhu, X. Gong, X. Yi, and S. Li, "Data poisoning attacks in internet-of-vehicle networks: Taxonomy, state-of-the-art, and future directions," *IEEE Trans. Ind. Inform.*, vol. 19, no. 1, pp. 20–28, Jan. 2023.

[28] Z. Deng, G. Xin, Y. Liu, W. Wang, and B. Wang, "Contrastive graph neural network-based camouflaged fraud detector," *Inf. Sci.*, vol. 618, pp. 39–52, Dec. 2022.

[29] Z. Zuo, H. Zhang, L. Ma, T. Liu, and S. Liang, "Leak detection for natural gas gathering pipelines under multiple operating conditions using RP-1dConvLSTM-AE and multimodel decision," *IEEE Trans. Ind. Electron.*, vol. 71, no. 6, pp. 6263–6273, Jun. 2024.

[30] A. Yang, C. Lu, W. Yu, J. Hu, Y. Nakanishi, and M. Wu, "Data augmentation considering distribution discrepancy for fault diagnosis of drilling process with limited samples," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 11774–11783, Nov. 2023.

[31] W. Gong, Y. Wang, M. Zhang, E. Mihankhah, H. Chen, and D. Wang, "A fast anomaly diagnosis approach based on modified CNN and multisensor data fusion," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 13636–13646, Dec. 2022.

[32] P. Liu, X. Sun, Y. Han, Z. He, W. Zhang, and C. Wu, "Arrhythmia classification of LSTM autoencoder based on time series anomaly detection," *Biomed. Signal Process. Control*, vol. 71, 2022, Art. no. 103228.

[33] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Multireceptive field graph convolutional networks for machine fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12 739–12 749, Dec. 2021.

[34] F. Yang, S. Shah, D. Xiao, and T. Chen, "Improved correlation analysis and visualization of industrial alarm data," *ISA Trans.*, vol. 51, no. 4, pp. 499–506, 2012.

[35] B. Yang, H. Li, and B. Wen, "A dynamic time delay analysis approach for correlated process variables," *Chem. Eng. Res. Des.*, vol. 122, pp. 141–150, Jun. 2017.

[36] S. X. Ding, L. Li, and M. Kruger, "Application of randomized algorithms to assessment and design of observer-based fault detection systems," *Automatica*, vol. 107, pp. 175–182, Sep. 2019.

Dong Zhao (Senior Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in control science and engineering from Beijing University of Chemical Technology, Beijing, China, in 2011 and 2016, respectively.

From 2017 to 2018, and in 2021, he worked as a Postdoctoral Research Fellow with the Institute for Automatic Control and Complex Systems (AKS), University of Duisburg-Essen, Duisburg, Germany. From 2018 to 2020, he joined as a Postdoctoral Research Fellow with KIOS Research and Innovation Center of Excellence, the University of Cyprus, Nicosia, Cyprus. Since 2022, he has been a Professor with the School of Cyber Science and Technology, Beihang University, Beijing, China. His research interests include fault diagnosis, fault-tolerant control, cyber-physical systems, and cyber security.

Bo Yang (Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in control science and engineering from Beijing University of Chemical Technology, Beijing, China, in 2011 and 2018, respectively.

From 2021 to 2022, he was a Visiting Scholar with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada. He is a Lecturer with the Business School, Sichuan University, Sichuan, China. His research interests include process data analysis, intelligent decision-making, and energy demand forecasting.

Yueyang Li (Member, IEEE) received the B.Sc. and Ph.D. degrees in control theory and control engineering from Shandong University, Jinan, China, in 2006 and 2011, respectively.

In 2014, he was a Visiting Scholar with the Center for Robotics of Shandong University, Jinan, China. In 2015, he was a Visiting Scholar with the Institute for Automatic Control and Complex Systems (AKS), University of Duisburg-Essen, Duisburg, Germany. He is currently the Vice Dean and a Professor with the School of Electrical Engineering, University of Jinan, Jinan, China. His research interests include fault diagnosis for time-varying systems, robust filtering for stochastic systems, and its applications to robotics and multidimensional systems.

Hui Zhang (Fellow, IEEE) received the B.Sc. degree in mechanical design manufacturing and automation from Harbin Institute of Technology, Weihai, China, in 2006, the M.Sc. degree in automotive engineering from Jilin University, Changchun, China, in 2008, and the Ph.D. degree in mechanical engineering from the University of Victoria, Victoria, BC, Canada, in 2012. He is currently a Professor with the School of Transportation Science and Engineering, Beihang University, Beijing, China.

Dr. Zhang was an Associate Editor for IEEE TRANSACTIONS ON INTELLIGENT VEHICLES, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Journal of the Franklin Institute*, *SAE International Journal of Vehicle Dynamics, Stability, and NVH*, *SAE International Journal of Connected and Automated Vehicles*, and *ASME Transactions Journal of Dynamic Systems, Measurement and Control*, and the Board member of the *International Journal of Hybrid and Electric Vehicles, Mechanical Systems, and Signal Processing*. He was the recipient of 2017 IEEE Transactions on Fuzzy Systems Outstanding Paper Award, 2018 SAE Ralph R. Teetor Educational Award, IEEE Vehicular Technology Society 2019 Best Vehicular Electronics Paper Award, and 2019 SAE International Intelligent and Connected Vehicles Symposium Best Paper Award. He is a member of the SAE International and ASME.