Detection of Cyber Attacks in Encrypted Control Systems

Moritz Fauser* and Ping Zhang*

Abstract—Recently a resilient homomorphic encryption (RHE) scheme has been proposed in [1], which allows not only to carry out the evaluation process of an output feedback controller in an encrypted environment but also can neutralize the effect of additive attacks injected into the ciphertexts. However, the resilience to additive attacks has its limits. In this paper, at first it will be shown that the resilience range of the RHE scheme to additive attacks is indeed much larger than shown in [1]. Then, a detection approach is proposed to give a twofold protection to control systems encrypted by the RHE scheme. A warning signal is triggered as soon as an additive attack is injected into the ciphertexts transmitted over the network, while an alarm signal is triggered when the attack is outside of the resilience range. This is achieved by exploiting the symmetric property of the inner product. Therefore, the RHE scheme can be combined with the proposed detection approach to ensure the integrity of the signals obtained after decryption in case of additive attacks. A simulation example of the well-established quadruple-tank benchmark process is used to demonstrate the proposed detection approach for encrypted control systems.

I. INTRODUCTION

Industrial plants are connected more and more over communication networks and are the so-called cyber-physical systems (CPS) [2]. A CPS in an industrial environment is, for instance, the industrial control system (ICS). Due to the remote access to an ICS, criminals can execute a cyber attack over the network structure of a CPS, whose goal is to manipulate the physical system or obtain information. Therefore, the interest in cyber security of control systems has increased greatly in the last decade [3]. If a CPS maintains an acceptable level of performance in the face of cyber attacks, the CPS is resilient to such attacks.

A well-known concept in the information security is the CIA triad, which defines three primary properties, namely, confidentiality, integrity and availability [4]. Confidentiality refers to the ability to protect the data from those who are not authorized to view it. Integrity relates to the validity of data from being changed in an unauthorized manner. Availability refers to the ability to access data when needed.

Homomorphic encryption (HE) schemes can not only ensure the confidentiality of data sent over the network but also enable arithmetic operations on encrypted data. A HE scheme can be implemented in a CPS structure or cloud architecture to operate with ciphertexts (see, for instance, [5]–[9]). A HE scheme which allows an infinite number of additions and multiplications on encrypted data is a so-called fully homomorphic encryption (FHE) scheme [10]. To improve the computational efficiency and meet the requirement of real-time implementation, somewhat homomorphic encryption (SWHE) schemes have been developed to allow a limited number of additions and multiplications on encrypted data [11].

Some approaches which ensure both the confidentiality and the integrity of signals transmitted over the network have been proposed in [12], [13]. In these schemes, if a manipulation of the plaintext obtained after decryption is detected, the message will be rejected. But the rejection of messages contradicts the requirement of a control system because it needs real-time feedback information. Therefore, the resilience of signals sent over the network is important to ensure the availability of data.

To solve this problem, the resilient homomorphic encryption (RHE) scheme has been developed in [1] and can be used to encrypt controllers based on the matrix-vector product. The RHE scheme can neutralize the effect of an attack injected into the signals encrypted by the RHE scheme so that the controller can get the true sensor information and the actuators can get the true control input signals. The question now is how to increase the resilience range of the RHE scheme and how to detect the attack, if the ciphertext transmitted over the network is manipulated. Especially, if the attack is outside of the resilience range, how to generate an alarm to alert plant operators.

A detection approach for a control system encrypted by ElGamal [14] is introduced in [15]. By exploiting the sensitivity of the ElGamal HE scheme to additive attacks, the attack injected into the ciphertexts strongly influences the control input signals got after the decryption. Then the attack is detected by comparing the control input with a threshold.

In this paper, an extended resilience range of the RHE scheme [1] against additive attacks will be shown, which increases the attractiveness of the RHE scheme for the encryption of control systems. Then, the detection approach is presented that reveals attacks injected into the signals encrypted by the RHE scheme. By exploiting the symmetric property of the inner product, two residual signals can be formulated. In case that the attack is inside of the resilience range, the first residual signal deviates from zero and a warning is generated. As soon as the attack is outside of the resilience range, the second residual signal deviates from zero and an alarm is triggered. That means, we go one step further to prevent attacks that want to break the resilience of the RHE scheme by using the proposed detection approach.

This paper is organized as follows. Section II introduces some preliminary knowledge. The extended resilience range of the RHE scheme is shown in Section III. Section IV
presents the proposed attack detection approach. Section V shows an example of the quadruple-tank benchmark process.

II. PRELIMINARIES
A. System description
Fig. 1 shows the typical structure of a CPS. The plant is a discrete linear time-invariant (LTI) process described by
\[ x(k + 1) = Ax(k) + Bu(k) + Ed(k) \]
\[ y(k) = Cx(k) + Du(k) + F_d d(k) \]
where \( x(k) \in \mathbb{R}^n \) is the state vector, \( u(k) \in \mathbb{R}^m \) is the control input vector, \( y(k) \in \mathbb{R}^b \) is the measured output vector, \( d(k) \in \mathbb{R}^{m_4} \) is the unknown disturbance vector, \( A, B, C, D, E_d \) and \( F_d \) are matrices of compatible dimensions. Assume that the plant in (1) communicates over the network with a dynamic feedback controller described by
\[ x_s(k + 1) = Ax_s(k) + B_s y(k) \]
\[ u(k) = C_s x_s(k) + D_s y(k) \]
where \( x_s(k) \in \mathbb{R}^{n_1} \) is the state vector of the controller, \( A_s, B_s, C_s \) and \( D_s \) are matrices of appropriate dimensions. The controller (2) can be equivalently re-written as
\[ \begin{bmatrix} x_s(k + 1) \\ u(k) \end{bmatrix} = \begin{bmatrix} A_s & B_s \\ C_s & D_s \end{bmatrix} \begin{bmatrix} x_s(k) \\ y(k) \end{bmatrix} = T g(k), \]
where \( T \in \mathbb{R}^{(n_s + m) \times (m + b)}, g(k) \in \mathbb{R}^{m + b}, \alpha = n_s + m \) and \( \beta = n_s + b \).
B. Resilient homomorphic encryption scheme
The RHE scheme in [1] allows the calculation process of feedback controllers with ciphertexts. Different from other homomorphic encryption schemes, the RHE scheme is able to neutralize the effect of attacks injected into the ciphertexts.

The RHE scheme [1] is a kind of symmetric encryption scheme and uses the same secret key for both encryption and decryption. It consists of four functions \( E = (\text{KeyGen}, \text{Enc}, \text{Dec}, \text{Eval}) \).

Key Generation (KeyGen): Select the prime numbers \( p \) and \( q \) in a way that an adversary cannot factorize \( pq \) to get the prime numbers \( p, q \) in polynomial time. Moreover, a security parameter \( \theta \) has to be chosen to decrypt later correctly.

Encryption (Enc): A matrix \( F \in \mathbb{Z}^{n \times \beta} \) is encrypted by \( F_c = \text{Enc}(F, \theta, p) = (S_f + F \theta + R_f p) \mod pq \)
\[ = \begin{bmatrix} s_{11} + f_{11} \theta + r_{11} p \\ \vdots \\ s_{1\beta} + f_{1\beta} \theta + r_{1\beta} p \\ \vdots \\ s_{\alpha1} + f_{\alpha1} \theta + r_{\alpha1} p \\ \vdots \\ s_{\alpha\beta} + f_{\alpha\beta} \theta + r_{\alpha\beta} p \end{bmatrix} \mod pq \]
and a vector \( w \in \mathbb{Z}^{\beta \times 1} \) is encrypted by \( w_c = \text{Enc}(w, \theta, p) = (S_w + w \theta + r_w p) \mod pq \)
\[ = \begin{bmatrix} s_1 + w_1 \theta + r_1 p \\ \vdots \\ s_\beta + w_\beta \theta + r_\beta p \end{bmatrix} \mod pq \]
where \( f_{ij}, w_i \in \{0, 1, \cdots, M\} \) are the plaintexts, \( r_{ij}, r_i \in \{1, 2, \cdots, q - 1\}, s_{ij}, s_j \in \{0, 1, \cdots, S\} \) are the noises, \( i = 1, 2, \cdots, \alpha, j = 1, 2, \cdots, \beta, S \) and \( M \) are known integers.

Decryption (Dec): The decryption of an encrypted vector \( c_x \in \mathbb{Z}^{n \times 1} \) is given by \( \nu = \text{Dec}(c_x, \theta, p) \)
\[ = (c_x \mod p - (c_x \mod p) \mod \theta^2) / \theta^2 \]
Evaluation (Eval): The matrix-vector product of the encrypted matrix \( F_c \) and the encrypted vector \( w_c \) is
\[ c_x = F_c w_c \mod pq \]
whose decryption is equal to \( F w, \)
\[ \|S_f s_w + \theta(S_f w + F s_w)\|_\infty < \theta^2, \]
\[ \|S_f s_w + \theta(S_f w + F s_w) + \theta^2 F w\|_\infty < p, \]
where \( \|w\|_\infty = \max_i |w_i| \) denotes the infinity norm.

C. Mapping function
The RHE scheme [1] allows only positive integer plaintexts from the interval \( \{0, 1, \cdots, M\} \). However, the signals in the CPS take often real values. Therefore, it is necessary to transform real values into positive integer values.

Let \( \lambda_1, \lambda_2 \in \mathbb{N} \) represent the range and the resolution of the quantization. Then negative and positive numbers \( \zeta \in [-2^{\lambda_1}, 2^{\lambda_1}] \) can be mapped into the set \( Q = \{-2^{\lambda_1}, -2^{\lambda_1 + 2}, \cdots, 2^{\lambda_1} - 2^{\lambda_2}\} \) by the following quantization function
\[ Q(\zeta) = \begin{cases} \max(\zeta) \in Q, \zeta \leq \zeta \zeta, & \text{if } \zeta \geq 0 \\
\min(\zeta) \in Q, \zeta \geq \zeta \zeta, & \text{if } \zeta < 0 \end{cases} \]
Let \( \mu = 2^{2(\lambda_1 + \lambda_2) + 1} + 1 \) and the bound \( M = 2^{\lambda_1 + \lambda_2} + 1 \). The quantized value \( \zeta \) can be mapped to an integer by
\[ \Gamma(\zeta) = 2^{\lambda_2} \zeta \mod \mu \]
To obtain the quantized value \( \Gamma(\zeta) \) corresponding to an integer \( \eta \) obtained after the matrix-vector product, the inverse mapping function \( \Gamma^{-1}(\eta) \) is given by
\[ \Gamma^{-1}(\eta) = \begin{cases} \frac{\eta}{2^{\lambda_2}} (\eta \mod \mu - \mu / 2), & \text{if } \eta \mod \mu > \mu / 2 \\
\frac{1}{2^{\lambda_2}} (\mu - \eta \mod \mu), & \text{if } \eta \mod \mu \leq \mu / 2 \end{cases} \]
It can be shown that the relation \( \Gamma^{-1}(\Gamma(T)) \Gamma(\Gamma(g(k))) = \Gamma(\tilde{g}(k)) \) holds, where \( \tilde{T} \) and \( g(k) \) are obtained, respectively, by the quantization of \( T \) and \( g(k) \), i.e. \( T = Q(T), g(k) = Q(g(k)) \). The details of the derivation can be found in [16].

D. Encrypted controller scheme
The structure of control systems encrypted with the RHE scheme in [1] is shown in Fig. 2. The controller parameters \( A_s, B_s, C_s \) and \( D_s \) are collected in the matrix \( T \) by (3) and quantized element-wise by (10) as \( T = Q(T) \). Then, the quantized matrix \( T \) is mapped element-wise by the mapping function \( \Gamma(\tilde{T}) \) in (11) and encrypted by (4), which gives \( F_c \). Instead of sending the sensor outputs \( y(k) \)
and the state vector $x_s(k)$ directly over the communication network, $y(k)$ and $x_s(k)$ are stacked together by (3) and then mapped with, respectively, the quantization function (10) and the mapping function in (11) into the integer vector

$$w(k) = [T^T(Q(x_s)) T^T(Q(y))].$$

Before $w(k)$ is sent over the network to the controller, $w(k)$ is encrypted to $w_e(k)$ in (5). On the controller side, $w_e^*(k)$ is processed by the evaluation procedure $c_x^*(k) = F_c w_e^*(k) \mod pq$ in (7). The ciphertext $c_x^*(k)$ arriving at the plant side will be decrypted by (6) and then the inverse mapping function in (12) is applied, which gives the control input signal $u(k)$ and the state vector $x_s(k+1)$. For the next controller cycle, the state vector $x_s(k+1)$ is set to $x_s(k)$.

The quantization function introduced in Section II-C causes a quantization error. But if the parameters of the quantization are appropriately selected, the quantization error is bounded and

$$2^{-\lambda_2} \max_{k=0} \sum_{j=1} \|((A \Sigma - B \Sigma C \Sigma) B \Sigma) T e_j\| < 2^{\lambda_1},$$

then the closed-loop system is stable, where $\|v\|_1 = \sum_{j=1} |v_j|$ denotes the 1-norm of the vector $v \in \mathbb{R}^{\beta \times 1}$ and $e_j \in \mathbb{R}^{\beta \times 1}$ is the unit vector and

$$A \Sigma = \begin{bmatrix} O & O \\ O & A \end{bmatrix}, \quad B \Sigma = \begin{bmatrix} I & O \\ I & B \end{bmatrix}, \quad C \Sigma = \begin{bmatrix} I & O \\ O & C \end{bmatrix}.$$ (14)

The matrix $I$ and the matrix $O$ are, respectively, an identity matrix and a zero matrix of compatible dimensions. If $A \Sigma - C \Sigma B \Sigma$ is Schur stable and $\lambda_1, \lambda_2$ are selected to be large enough, the closed-loop system is stable [16]. Note that the summation term in (13) can be calculated with the methods provided by [17].

III. RESILIENCE RANGE OF THE RHE SCHEME

Let $a_1$ and $a_2$ denote the attacks injected into the signals encrypted by the RHE scheme, as shown in Fig. 3. The previous analysis in [1] has shown that the RHE scheme is resilient to additive attacks satisfying

$$-h \leq G a_1 + a_2 < \theta^2 1 - h.$$ (15)

In this section, we shall show that the resilience range of the RHE scheme is indeed much larger than (15), which significantly increases the attractiveness of the RHE scheme for the encryption of control systems.

**Theorem 1:** Assume that the matrix $F \in \mathbb{N}^{\alpha \times \beta}$ and a vector $w \in \mathbb{N}^{\beta \times 1}$ are encrypted by, respectively, (4) and (5), which gives

$$F_c = Enc(F, \theta, p) = (S_f + F \theta + R f p) \mod pq,$$

$$w_c = Enc(w, \theta, p) = (s_w + w \theta + r_w p) \mod pq.$$ (16)

Let $a_1 \in \mathbb{N}^{\beta \times 1}$ and $a_2 \in \mathbb{N}^{\alpha \times 1}$ denote additive attacks on, respectively, the ciphertexts $w_c$ and $c_x^*$, i.e.,

$$w_c^* = w_c + a_1, \quad c_x^* = c_x^* + a_2,$$ (17)

where $c_x^* = (F_c w_c^*) \mod pq$. If $\|F_c w_c^* (s_w + w \theta + r_w p + a_1)\|_\infty < pq$, then the result $\nu$ obtained by decryption in (6) is

$$\nu = Dec(c_x^*, \theta, p) = F w_c,$$ (18)

as long as

$$zp - h \leq G a_1 + a_2 < zp + \theta^2 1 - h.$$ (19)

where $h = S_f s_w + \theta (S_f w + F s_w), G = S_f + \theta F, 1_{a \times 1}$ is a vector of ones and $z \in \mathbb{N}^{\alpha \times 1}$ whose entries are $z_i \in \{0, 1, \cdots, q - 1\}$.

**Proof:** According to (5), the encryption of $w$ gives $w_c = Enc(w, \theta, p) = (s_w + w \theta + r_w p) \mod pq$, (20) where $\theta$ is the security parameter and $p, q$ are prime numbers. Due to the additive attack $a_1$,

$$w_c^* = Enc(w, \theta, p) + a_1$$ (21)

If $\|(S_f + F \theta + R f p)(s_w + w \theta + r_w p + a_1)\|_\infty < pq$, then the matrix-vector product in (7) $c_x^*$ gives

$$c_x^* = F_c w_c^* \mod pq$$ (22)

Because of the attack $a_2$, there is $c_x = c_x^* + a_2$. Note that $c_x^* \mod p = ((S_f + F \theta + R f p)(s_w + w \theta + r_w p + a_1) + a_2) \mod p$

$$= (S_f + \theta F) a_1 + s_f s_w + \theta (S_f w + F s_w) + \theta^2 F w$$

$$+ p(S_f r_w + R f (s_w + a_1) + \theta (F r_w + R f w) + p R f r_w + a_2) \mod p$$ (23)

If the attacks $a_1$ and $a_2$ satisfy (19), then $zp \leq G a_1 + h + a_2 < zp + \theta^2 1$. Let

$$G a_1 + h + a_2 = zp + \Delta,$$ (24)

where $\Delta \in \mathbb{N}^{\alpha \times 1}$ whose entries are $\Delta_i \in \{0, 1, \cdots, \theta^2 - 1\}$. Thus,

$$c_x^* \mod p = (G a_1 + h + a_2 + \theta^2 F w) \mod p$$ (25)

$$= (zp + \Delta + \theta^2 F w) \mod p = \Delta + \theta^2 F w.$$ (26)

Due to the decryption (6), it yields

$$\nu = (c_x^* \mod p - (c_x \mod p) \mod \theta^2) \mod \theta^2$$ (27)

$$= (\Delta + \theta^2 F w - (\Delta + \theta^2 F w) \mod \theta^2) \mod \theta^2 = F w,$$

i.e., the vector $\nu$ is still the correct result of $F w$.

By comparing (15) with (19), it can be seen that the resilience range against additive attacks is indeed much larger than shown in [1]. The resilience range in (19) varies with time, because the resilience range involves the vector $h(k)$ that depends on the noise vector $s_w(k)$ whose entries change at each time instant.

IV. DETECTION APPROACH

To detect an attack that breaks the resilience range of the RHE scheme, a detection approach is proposed to give a double protection to the control system encrypted with the RHE scheme. As soon as an additive attack is injected into the ciphertexts sent over the network, a warning will be generated. If the attack breaks the resilience range, an alarm will be triggered. The basic idea of the detection process is schematically illustrated in Fig. 4. The attack detection system is composed of a warning generator and an alarm generator.
A. Detection of the presence of attacks

The warning signal $\phi_1(k)$ aims to detect the existence of an attack injected into the ciphertext over the network. It is generated based on the noise $s_w(k)$ used in the encryption (5) and the received ciphertext $c_x(k)$ by

$$\hat{s}(k) = (c_x(k) \mod p) \mod \theta$$

(27)

e_1(k) = \langle s_w(k), S_f \cdot s_{pre} \rangle - \langle s_{pre}, \hat{s}(k) \rangle (28)

where $e_1(k)$ is the residual signal, $\langle \cdot \rangle$ is the inner product and $s_{pre} \in \mathbb{N}^{\alpha \times 1}$ is a predefined vector that has to be kept secret whose entries are $s_{pre,i} \neq 0, i \in \{1, 2, \cdots, \alpha\}$. A warning is triggered based on the following decision logic

$$e_1(k) = 0 \Rightarrow \text{CPS is attack-free, } \phi_1(k) = 0$$

$$e_1(k) \neq 0 \Rightarrow \phi_1(k) = 1,$$

(29)

where $\phi_1(k)$ is the warning signal.

B. Detection of attacks outside of resilience range

The alarm signal $\phi_2(k)$ aims to reveal attacks that break the resilience range. It is generated based on the vector $w(k) = [x_s^T(k) y^T(k)]^T$ and the vector $\nu(k)$ got after the decryption in (6) by

$$e_2(k) = \langle w(k), F^T s_{pre} \rangle - \langle s_{pre}, \nu(k) \rangle (30)$$

where $\nu(k)$ is the result after decryption (6). An alarm is triggered based on the following decision logic

$$e_2(k) = 0 \Rightarrow \nu(k) \text{ is not influenced, } \phi_2(k) = 0$$

$$e_2(k) \neq 0 \Rightarrow \nu(k) \text{ is influenced, } \phi_2(k) = 1,$$

(31)

where $\phi_2(k)$ is the alarm signal.

C. Analysis of behavior of $\phi_1(k)$ and $\phi_2(k)$

Three cases can be differentiated.

Case I: No attacks in the CPS, i.e. $a_1(k) = 0, a_2(k) = 0$. In this case, $w^*_s(k) = w_s(k)$ and $c_x(k) = c^*_x(k)$. If $\|S_f s_w\|_\infty < \theta$, substituting (7) into (27) gives

$$\hat{s} = (c_x \mod p) \mod \theta = \langle (F \cdot w_s \mod p \cdot q) \mod \theta = \langle S_f s_w + \theta(F w_s + F s_w) + \theta^2 F w_s \rangle$$

(32)

$$+ p(S_f r_w + R_f s_w + \theta(F r_w + R_f w))$$

$$+ p(R_f r_w) \mod p) \mod \theta = S_f s_w$$

As a result,

$$e_1(k) = \langle s_w(k), S_f^T s_{pre} \rangle - \langle s_{pre}, \hat{s}(k) \rangle$$

(33)

Due to the symmetric property of the inner product, it holds

$$\langle s_w(k), S_f^T s_{pre} \rangle = \langle s_{pre}, S_f s_w(k) \rangle$$

(34)

Therefore,

$$e_1(k) = 0 \text{ and } \phi_1(k) = 0$$

(35)

According to (6), the vector $\nu(k)$ is given by

$$\nu = (c_x \mod p - (c_x \mod p) \mod \theta^2) / \theta^2 = F w$$

(36)

From (30) and the decision logic (31), we get

$$e_2(k) = \langle w(k), F^T s_{pre} \rangle - \langle s_{pre}, \nu(k) \rangle = 0$$

$$\langle w(k), F^T s_{pre} \rangle - \langle s_{pre}, F w(k) \rangle = 0$$

(37)

$$\phi_2(k) = 0$$

Case II: The attacks $a_1(k), a_2(k)$ are inside of the resilience range, i.e. $G a_1(k) + h(k) + a_2(k) \in \{z, zp, zp + \theta^2\}$. As shown in Section III, in this case, the influence of the attacks $a_1, a_2$ is cancelled out of the vector $\nu(k)$ obtained after decryption, i.e. $\nu(k) = F w(k)$. Let $G a_1(k) + h(k) + a_2(k) = z(k)p + \Delta(k)$

and $\Delta \in \mathbb{N}^{\alpha \times 1}$ whose entries are $\Delta_i \in \{0, 1, \cdots, \theta^2 - 1\}$. Substituting (25) into (27) yields

$$\hat{s} = (c_x \mod p) \mod \theta = (\Delta + \theta^2 F w) \mod \theta = \Delta_0$$

(39)

where $\Delta_0 \in \mathbb{N}^{\alpha \times 1}$ whose entries are $\Delta_0 \in \{0, 1, \cdots, \theta - 1\}$. Taking into account (28), we get

$$e_1(k) = \langle s_w(k), S_f^T s_{pre} \rangle - \langle s_{pre}, \Delta_0 \rangle$$

(40)

Therefore,

$$e_1(k) \neq 0 \text{ and } \phi_1(k) = 1,$$

(41)

as long as $\Delta_0 \neq S_f s_w(k) + \Delta_0$, where $\langle s_{pre}, \Delta_0 \rangle = 0$. Recall that the entries of the vector $s_w(k)$ in (40) are randomly chosen integers and change with time, its impossible for the attacker to reach $\Delta_0 = S_f s_w(k) + \Delta_0$ at each time step. Since $\nu(k) = F w(k)$, there is

$$e_2(k) = \langle w(k), F^T s_{pre} \rangle - \langle s_{pre}, \nu(k) \rangle$$

$$\langle w(k), F^T s_{pre} \rangle - \langle s_{pre}, F w(k) \rangle = 0$$

(42)

$$\phi_2(k) = 0$$

Case III: The attacks $a_1(k), a_2(k)$ are outside of the resilience range, i.e. $G a_1(k) + h(k) + a_2(k) \notin \{z, zp, zp + \theta^2\}$. Let

$$G a_1(k) + h(k) + a_2(k) = z(k)p + \Theta(k)$$

(43)

where $l(k) \in \mathbb{N}^{\alpha \times 1}$ whose entries are $l_i(k) \in [-\nu_1(k), \frac{\nu}{\theta}]$, $l_i(k)$, $\nu_1(k)$ are the entries of the vector $\nu(k) = F w(k)$ and $\nu$ is the operator that rounds to the nearest integer toward negative infinity. Because of the attacks $a_1(k)$ and $a_2(k)$, it holds

$$c_x \mod p = \langle G a_1 + h + a_2 + \theta^2 F w \rangle \mod p$$

$$\nu = (z + \theta^2 + \Delta + \theta^2 F w) \mod p$$

(44)

$$\nu = \Delta + \theta^2 (F w + l),$$

where $||l^2 + \theta^2 F w||_\infty < p$. By taking into account (27), it yields

$$\hat{s} = (c_x \mod p) \mod \theta$$

(45)

$$e_1(k) = \langle s_{\nu}(k), S_f^T s_{pre} \rangle - \langle s_{pre}, \Delta_\nu \rangle$$

where $\Delta_\nu(k) \in \mathbb{N}^{\alpha \times 1}$ whose entries are $\Delta_\nu(k) \in \{0, 1, \cdots, \theta - 1\}$. If $\Delta_\nu(k) \neq S_f s_w(k) + \Delta_\nu$, where $\langle s_{pre}, \Delta_\nu \rangle = 0$, then

$$e_1(k) \neq 0 \text{ and } \phi_1(k) = 1.$$
Due to (44), the decryption of $c(x)$ by (6) gives
\begin{align*}
\nu &= Dec(c(x), \theta, p) = (c(x) \mod p - (c(x) \mod p) \mod \theta^2) / \theta^2 \\
&= (\Delta + \theta^2(Fw + l) - \Delta) / \theta^2 = Fw + l
\end{align*}
(47)
From (9) and (30), it yields
\begin{align*}
e_2(k) &= \langle w(k), F^T s_{pre} \rangle - \langle s_{pre}, \nu(k) \rangle \\
&= \langle w(k), F^T s_{pre} \rangle - \langle s_{pre}, Fw(k) + l(k) \rangle \\
&= \langle s_{pre}, l(k) \rangle
\end{align*}
(48)
Thus, if \( \langle s_{pre}, l(k) \rangle \neq 0 \), then
\begin{equation}
e_2(k) \neq 0 \quad \text{and} \quad \phi_2(k) = 1. \tag{49}
\end{equation}

\section*{D. Influence of attacks outside of the resilience range}
As can be seen from (47), if the attacks $a_1(k)$, $a_2(k)$ break the resilience range, then the vector $\nu(k)$ got by the decryption is not equal to the true value of the vector $Fw(k)$ any more. Now we check how big $l(k)$ (i.e. the difference between $\nu(k)$ and the true value of $Fw(k)$) is.

If $zp + 2\theta^2 - 1 - h \leq Ga_1 + a_2 < zp + 2\theta^2 - 1$, then $l = 1$.

If $zp - \theta^2 - 1 - h \leq Ga_1 + a_2 < zp - h$, then $l = -1$.
That means, the entries of the vector $\nu(k)$ obtained by decryption differ from the true values by 1.

According to (43), there is
\begin{equation}
l = (Ga_1 + a_2 + h - \Delta - zp) / \theta^2 \tag{50}
\end{equation}
Recall that $\theta$ is the security parameter presupposed during the key generation and is usually a large number, $0 \leq \Delta \leq \theta^2 - 1$ and $\|h\|_{\infty} = \|Sf_s w + \theta(Sf w + Fs w)\|_{\infty} \leq \theta^2$. It holds
\begin{equation}
-\theta^2 - 1 \leq h - \Delta \leq \theta^2 - 1 \quad \Leftrightarrow -1 < \langle \nu - Fw \rangle / \theta^2 < 1 \tag{51}
\end{equation}
If $Ga_1 + a_2 \geq zp + 2\theta^2 - 1 - h$, then $l$ is a vector with positive integer entries that always satisfies
\begin{equation}
l \geq (Ga_1 + a_2 - zp) / \theta^2 + 1 \leq Ga_1 + a_2 \tag{52}
\end{equation}
If $Ga_1 + a_2 \leq zp - 1 - \theta^2 - h$, then $l$ is a vector with negative integer entries and satisfy
\begin{equation}
l \leq (Ga_1 + a_2 - zp) / \theta^2 - 1 \tag{53}
\end{equation}
In these two scenarios, the difference between the vector $\nu$ got after the decryption and the true value of $Fw(k)$ is $l = \nu - Fw$. Because $\theta$ is a large number, $|l| \ll |Ga_1 + a_2|$. That means, the influence of attacks outside of the resilience range will be significantly reduced.

In summary, as soon as the attacks $a_1(k)$ and $a_2(k)$ are injected into the encrypted signals, the residual signal $e_1(k)$ got by (28) will deviate from zero and the warning signal $\phi_1(k)$ will be triggered to signify the existence of the attacks. If the attacks $a_1(k)$, $a_2(k)$ are outside of the resilience range (i.e. $Ga_1(k) + h(k) + a_2(k) \notin [zp, zp + \theta^2]$), the residual signal $e_2(k)$ obtained by (30) will deviate from zero and triggers the alarm signal $\phi_2(k)$. Note that, because $|l(k)| \ll |Ga_1(k) + a_2(k)|$, the difference between $\nu(k)$ got after the decryption and the true value of $Fw(k)$ caused by the attacks $a_1(k)$ and $a_2(k)$ is significantly reduced. As a result, the control system can still keep much of the control performance.

\section*{V. EXAMPLE}
In this section, an example is given to illustrate the proposed detection approach in the control system encrypted by the RHE scheme as shown in Fig. 4.
Consider the quadruple-tank system in [18], which is a nonlinear continuous-time plant model and linearized at the working point $[x_1 \ x_2 \ x_3 \ x_4]^T = [12.6 \ 13 \ 4.8 \ 4.9]^T cm$.
The states in the model are the liquid levels in the tanks, the control inputs are the pressure in the pumps. The sampling time is $T = 1$ second. The plant is described by (1) with
\begin{equation}
A = \begin{bmatrix}
0.9843 & 0 & 0.0251 & 0 \\
0 & 0.9891 & 0 & 0.0176 \\
0 & 0 & 0.9747 & 0 \\
0 & 0 & 0 & 0.9823
\end{bmatrix}, \quad B = \begin{bmatrix}
0.0478 & 0.0010 \\
0.0005 & 0.0348 \\
0 & 0.0776 \\
0.0554 & 0
\end{bmatrix}, \quad C = \begin{bmatrix}
0.5 & 0 & 0 & 0 \\
0 & 0.5 & 0 & 0
\end{bmatrix} \quad D = O, \quad E_u = \begin{bmatrix} B \ O \end{bmatrix}, \quad F_d = \begin{bmatrix} O \ I \end{bmatrix} \tag{54}
\end{equation}
To stabilize the system, the parameters of the controller defined in (3) are selected as
\begin{equation}
A_s = \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}, \quad B_s = \begin{bmatrix}
0.0625 & 0 \\
0 & 0.01563
\end{bmatrix}, \quad C_s = \begin{bmatrix}
0.032 & 0 & -0.0192 \\
0 & 1 & 0
\end{bmatrix}, \quad D_s = \begin{bmatrix}
1 & 0 \\
0 & -0.3
\end{bmatrix} \tag{55}
\end{equation}
It can be checked that $A_{\Sigma2} - C_S B_{\Sigma2}$ is Schur stable.

The maximal signal value in the system is smaller than 15. Thus, select $\lambda_1 = 4$. In order to satisfy condition (13), the integer $\lambda_2$ is chosen as $\lambda_2 = 16$, which gives $\mu = \beta2^{(\lambda_1+\lambda_2)+1} + 1 = 4 \times 2^{16} + 1$. As a result, the bound $M$ of the plaintext is $M = 2^{\lambda_1+\lambda_2+1} = 221$. The integer value $S$ defines the interval of the entries of the vector $s_w(k)$ and the matrix $S_f$ and is chosen as $S = 10^7$. The prime numbers $p$, $q$ and the security parameter $\theta$ are chosen from the interval $p \in (2^{189}, 2^{190})$, $q \in (2^{579}, 2^{580})$ and $\theta \in (2^{50}, 2^{51})$.

The parameters of the controller in (55) are mapped by (10), (11) and then encrypted by (4). Due to the limitation of space, the ciphertexts of the controller are omitted here.

The entries of the vector $s_{pre}$ of the residual signals (28), (30) are randomly chosen from the interval $\{0, 1, \cdots, S\}$.

During the simulation, an attack is imposed on the control input channels at $k = 150s$ (see Fig. 5e), while there is no attack on the sensor output channels. The adversary is not able to estimate the resilience range of the RHE scheme, because only ciphertexts are sent over the network. In order to show the effect of attacks inside of the resilience range and attacks outside of the resilience range, the attack signals $a_2(k) = [a_{21}(k), a_{22}(k)]^T$ are shown in Fig. 5e.

In the encrypted control system, the sensor output signal $y(k)$ (see Fig. 5a) and the state vector $x_s(k)$ of the controller are stacked together, mapped into the integer vector $w(k)$ and then encrypted by the RHE scheme which leads to the ciphertext $w_c(k)$. Due to space limitation, the ciphertext $w_c(k)$ and the ciphertext $c_s(k)$ got after the evaluation process of the controller are omitted here. Fig. 5b shows the control input signal $u(k)$ received by the actuators got after the decryption process. For comparison, the control input signal $u^*(k)$ generated by the classical feedback controller in the form of (2) and (55) is shown in Fig. 5c which is got based on the unencrypted signals (i.e. the plaintexts of controller parameters and outputs). Before the attack happens at $k = 150s$, no difference between $u^*(k)$ and $u(k)$ can be observed. Both the warning signal and the alarm signal keep
to be 0, i.e. $\phi_1(k) = 0$, $\phi_2(k) = 0$ (see Fig. 5d). Now we check the behaviour of the system after the attack happens at $k = 150s$. During the time interval from $k = 150s$ to $k = 195s$, the attack signal $a_2(k)$ is inside of the resilience range, because $a_2(k) < \theta^2 1 - h(k)$, where $h(k) = S_f s_u(k) + \theta(S_f w(k) + F s_w(k))$. Due to the neutralization effect of the RHE scheme, the attack signal has no influence of the control input signal $u(k)$ got after the decryption (see Fig. 5b). That means, the true values of the control input signal $u(k)$ are available to the actuators. Furthermore, the warning signal $\phi_2(k)$ changes from 0 to 1 at $k = 150s$, which signifies the presence of the attack signal $a_2(k)$. Beginning from $k = 195s$, the attack signal $a_2(k) \geq \theta^2 1 - h(k)$ is outside of the resilience range and influences the control input signal $u(k)$ obtained by decryption. The alarm signal $\phi_2(k)$ changes from 0 to 1 at $k = 195s$ and reveals that the resilience range is broken by the attack signal $a_2(k)$. Fortunately, the influence of the attack signal $a_2(k)$ on the control input signal $u(k)$ is strongly reduced by the factor $\frac{1}{2^{2s^2} \theta^2} > \frac{1}{2^{2s^2} 2^{2s^2}}$. Hence, the influence of the attack signal $a_2(k)$ is almost not visible in Fig. 5b.

We have also carried out simulations for the attack scenarios $a_1 \neq 0$, $a_2 = 0$ and $a_1 \neq 0$, $a_2 \neq 0$, which shows similar phenomenon as shown above. Due to space limitation, the simulation results cannot be included here.

VI. CONCLUSION

In this paper, at first the resilience range of the recently proposed RHE scheme is analyzed. Though the RHE scheme is able to neutralize the effect of additive attacks injected into the encrypted control system, the resilience range has still some limits. Therefore, a detection approach is proposed to protect the control systems encrypted with the RHE scheme in two levels. Two residual signals are generated by exploiting the symmetric property of the inner product.

If the additive attack injected into the signals sent over the network is inside of the resilience range, the control input signals obtained after the decryption are completely trusted and only a warning signal will be delivered. As soon as the attack breaks the resilience range and influences the control input signals got after the decryption, an alarm is triggered and countermeasures should be taken.

A CPS encrypted by the RHE scheme and equipped with proposed detection approach can not only make the CPS be resilient to additive attacks in a large resilience range, but also protect the control system from attacks that break the resilience range by giving alarms to alert plant operators.

REFERENCES

[1] M. Fauser and P. Zhang, “Resilient Homomorphic Encryption Scheme for Cyber-Physical System,” Accepted by the 60th IEEE Conf. on Decision and Control, Austin, USA, 2021.
[2] E. A. Lee, “Cyber-physical systems - are computing foundations adequate?” in Proc. Workshop Cyber-Physical Syst.: Research Motivation, Techniques and Roadmap, 2006, pp. 1–9.
[3] M. S. Chong, H. Sandberg, and A. M. Teixeira, “A tutorial introduction to security and privacy for cyber-physical systems,” in Proc. of the 18th European Control Conf., Naples, Italy, 2019, pp. 968–978.
[4] J. Andress, The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice. Syngress, New York, 2014.
[5] K. Kogiso and T. Fujita, “Cyber-security enhancement of networked control systems using homomorphic encryption,” in Proc. of 54th IEEE Conf. on Decision and Control, Osaka, Japan, 2015, pp. 6836–6843.
[6] F. Farokhi, I. Shames, and N. Bhatteham, “Secure and private control using semi-homomorphic encryption,” Control Engineering Practice, vol. 67, pp. 13–20, 2017.
[7] J. H. Cheon, K. Han, S. M. Hong, H. J. Kim, J. Kim, S. Kim, H. Seo, H. Shim, and Y. Song, “Toward a secure drone system: Flying with real-time homomorphic authenticated encryption,” IEEE Access, vol. 6, pp. 24325–24339, 2018.
[8] M. S. Darup, A. Redder, and D. E. Quevedo, “Encrypted cooperative control based on structured feedback,” IEEE Contr. Syst. Lett., vol. 3, no. 1, pp. 37–42, 2018.
[9] M. Ruan, H. Gao, and Y. Wang, “Secure and privacy-preserving consensus,” IEEE Trans. on Automatic Control, vol. 64, no. 10, pp. 4035–4049, 2019.
[10] C. Gentry, “A Fully Homomorphic Encryption Scheme,” Ph.D. dissertation, Stanford University, 2009.
[11] J. Dyer, M. Dyer, and J. Xu, “Practical homomorphic encryption over the integers for secure computation in the cloud,” International Journal of Information Security, vol. 18, no. 5, pp. 549–579, 2019.
[12] M. Abdalla, M. Bellare, and P. Rogaway, “The oracle Diffie-Hellman assumptions and an analysis of DHIES,” in Naccache D. (Ed.) Topics in Cryptology–CT-RSA, 2001, pp. 143–158.
[13] N. P. Smart, “The exact security of ECIES in the generic group model,” in Honary B. (Ed.) Cryptography and Coding, 2001, pp. 73–84.
[14] T. Elgamal, “A public key cryptosystem and a signature scheme based on discrete logarithms,” IEEE Trans. on Information Theory, vol. 31, no. 4, pp. 469–472, 1985.
[15] R. Baba, K. Kogiso, and M. Kishida, “Detection method of controller falsification attacks against encrypted control system,” in Proc. of the SICE Annual Conf., Nara, Japan, 2018, pp. 244–248.
[16] M. Fauser and P. Zhang, “Detection of Cyber Attacks in Encrypted Control Systems,” TechRxiv. https://doi.org/10.36223/techrxiv.19071125.2022.
[17] S. V. Rakovic, E. C. Kerrigan, K. I. Kouramas, and D. Q. Mayne, “Invariant approximations of the minimal robust positively invariant set,” IEEE Trans. on Automatic Control, vol. 50, no. 3, pp. 406–410, 2005.
[18] K. H. Johansson, “The quadruple-tank process: A multivariable laboratory process with an adjustable zero,” IEEE Trans. Control Syst. Technol., vol. 8, no. 3, pp. 456–465, 2000.