Detection of Cyber Attacks in Encrypted Control Systems

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY-NC-SA 4.0

SUBMISSION DATE / POSTED DATE

27-01-2022 / 01-02-2022

CITATION

Fausser, Moritz; Zhang, Ping (2022): Detection of Cyber Attacks in Encrypted Control Systems. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.19071125.v1>

DOI

[10.36227/techrxiv.19071125.v1](https://doi.org/10.36227/techrxiv.19071125.v1)

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

*The authors are with the Institute of Automatic Control, University of Kaiserslautern, Erwin-Schroedinger 12, 67653 Kaiserslautern, Germany (e-mail: fauser@eit.uni-kl.de; pzhang@eit.uni-kl.de)

Fig. 1. Classical structure of a CPS

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

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{E}_d\mathbf{d}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k) + \mathbf{F}_d\mathbf{d}(k) \end{aligned} \quad (1)$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state vector, $\mathbf{u}(k) \in \mathbb{R}^m$ is the control input vector, $\mathbf{y}(k) \in \mathbb{R}^b$ is the measured output vector, $\mathbf{d}(k) \in \mathbb{R}^{m_d}$ is the unknown disturbance vector, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}_d$ and $\mathbf{F}_d$ are matrices of compatible dimensions.

Assume that the plant in (1) communicates over the network with a dynamic feedback controller described by

$$\begin{aligned} \mathbf{x}_s(k+1) &= \mathbf{A}_s\mathbf{x}_s(k) + \mathbf{B}_s\mathbf{y}(k) \\ \mathbf{u}(k) &= \mathbf{C}_s\mathbf{x}_s(k) + \mathbf{D}_s\mathbf{y}(k) \end{aligned} \quad (2)$$

where $\mathbf{x}_s(k) \in \mathbb{R}^{n_s}$ is the state vector of the controller, $\mathbf{A}_s$, $\mathbf{B}_s$, $\mathbf{C}_s$ and $\mathbf{D}_s$ are matrices of appropriate dimensions. The controller (2) can be equivalently re-written as

$$\begin{bmatrix} \mathbf{x}_s(k+1) \\ \mathbf{u}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_s & \mathbf{B}_s \\ \mathbf{C}_s & \mathbf{D}_s \end{bmatrix} \begin{bmatrix} \mathbf{x}_s(k) \\ \mathbf{y}(k) \end{bmatrix} = \mathbf{T}\mathbf{g}(k), \quad (3)$$

where $\mathbf{T} \in \mathbb{R}^{\alpha \times \beta}$, $\mathbf{g}(k) \in \mathbb{R}^{\beta \times 1}$, $\alpha = n_s + m$ 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 $\mathcal{E} = (\text{KeyGen}, \text{Enc}, \text{Dec}, \text{Eval})$.

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

Encryption (Enc): A matrix $\mathbf{F} \in \mathbb{N}^{\alpha \times \beta}$ is encrypted by

$$\begin{aligned} \mathbf{F}_c &= \text{Enc}(\mathbf{F}, \theta, p) = (\mathbf{S}_f + \mathbf{F}\theta + \mathbf{R}_f p) \bmod pq \\ &= \begin{bmatrix} s_{11} + f_{11}\theta + r_{11}p & \cdots & s_{1\beta} + f_{1\beta}\theta + r_{1\beta}p \\ \vdots & \ddots & \vdots \\ s_{\alpha 1} + f_{\alpha 1}\theta + r_{\alpha 1}p & \cdots & s_{\alpha\beta} + f_{\alpha\beta}\theta + r_{\alpha\beta}p \end{bmatrix} \\ &\bmod pq \end{aligned} \quad (4)$$

and a vector $\mathbf{w} \in \mathbb{N}^{\beta \times 1}$ is encrypted by

$$\begin{aligned} \mathbf{w}_c &= \text{Enc}(\mathbf{w}, \theta, p) = (\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p) \bmod pq \\ &= \begin{bmatrix} s_1 + w_1\theta + r_1p \\ \vdots \\ s_\beta + w_\beta\theta + r_\beta p \end{bmatrix} \bmod pq \end{aligned} \quad (5)$$

where $f_{ij}, w_j \in \{0, 1, \dots, M\}$ are the plaintexts, $r_{ij}, r_j \in \{1, 2, \dots, q-1\}$, $s_{ij}, s_j \in \{0, 1, \dots, S\}$ are the noises, $i = 1, 2, \dots, \alpha$, $j = 1, 2, \dots, \beta$, S and M are known integers. Note that the prime number p should be much bigger than the bound M of the plaintexts.

Decryption (Dec): The decryption of an encrypted vector $\mathbf{c}_x \in \mathbb{N}^{\alpha \times 1}$ is given by

$$\begin{aligned} \mathbf{v} &= \text{Dec}(\mathbf{c}_x, \theta, p) \\ &= (\mathbf{c}_x \bmod p - (\mathbf{c}_x \bmod p) \bmod \theta^2) / \theta^2 \end{aligned} \quad (6)$$

Evaluation (Eval): The matrix-vector product of the encrypted matrix $\mathbf{F}_c$ and the encrypted vector $\mathbf{w}_c$ is

$$\mathbf{c}_x = \mathbf{F}_c \mathbf{w}_c \bmod pq \quad (7)$$

whose decryption is equal to $\mathbf{F}\mathbf{w}$, if

$$\|\mathbf{S}_f \mathbf{s}_w + \theta(\mathbf{S}_f \mathbf{w} + \mathbf{F} \mathbf{s}_w)\|_\infty < \theta^2, \quad (8)$$

$$\|\mathbf{S}_f \mathbf{s}_w + \theta(\mathbf{S}_f \mathbf{w} + \mathbf{F} \mathbf{s}_w) + \theta^2 \mathbf{F} \mathbf{w}\|_\infty < p, \quad (9)$$

where $\|\mathbf{v}\|_\infty = \max_i |v_i|$ denotes the infinity norm.

C. Mapping function

The RHE scheme [1] allows only positive integer plaintexts from the interval $\{0, 1, \dots, M\}$. However, the signals in the CPS take often real values. Therefore, it is necessary to transform real values into positive integer values. This can be achieved by a modified version of the mapping function given in [8].

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 $\mathcal{Q} = \{-2^{\lambda_1}, -2^{\lambda_1} + 2^{-\lambda_2}, \dots, 2^{\lambda_1} - 2^{-\lambda_2}\}$ by the following quantization function

$$Q(\zeta) = \begin{cases} \max\{\bar{\zeta} \in \mathcal{Q} | \bar{\zeta} \leq \zeta\}, & \text{if } \zeta \geq 0 \\ \min\{\bar{\zeta} \in \mathcal{Q} | \bar{\zeta} \geq \zeta\}, & \text{if } \zeta < 0 \end{cases} \quad (10)$$

Let $\mu = \beta 2^{2(\lambda_1 + \lambda_2) + 1} + 1$ and the bound M of the plaintexts is $M = 2^{\lambda_1 + \lambda_2 + 1}$. The quantized value $\bar{\zeta}$ can be mapped to an integer by

$$\Gamma(\bar{\zeta}) = 2^{\lambda_2} \bar{\zeta} \bmod \mu \quad (11)$$

To obtain the quantized value $\bar{\eta}$ corresponding to an integer η obtained after the matrix-vector product, the inverse mapping function $\Gamma^{-1}(\eta)$ is given by

$$\Gamma^{-1}(\eta) = \begin{cases} \frac{1}{2^{2\lambda_2}}(\eta \bmod \mu - \mu), & \text{if } \eta \bmod \mu > \mu/2 \\ \frac{1}{2^{2\lambda_2}}(\eta \bmod \mu), & \text{if } \eta \bmod \mu \leq \mu/2 \end{cases} \quad (12)$$

In the following it will be shown that the mapping procedure (11), (12) always delivers the correct result of a matrix-vector product, if the matrix and the vector involved in the matrix-vector product are element-wise quantized by (10).

Assume that the matrix $\mathbf{T}$ and the vector $\mathbf{g}(k)$ are element-wise quantized by (10), which yields

$$\bar{\mathbf{T}} = Q(\mathbf{T}) = \begin{bmatrix} Q(T_{11}) & \cdots & Q(T_{1\beta}) \\ \vdots & \ddots & \vdots \\ Q(T_{\alpha 1}) & \cdots & Q(T_{\alpha\beta}) \end{bmatrix} = \begin{bmatrix} \bar{T}_{11} & \cdots & \bar{T}_{1\beta} \\ \vdots & \ddots & \vdots \\ \bar{T}_{\alpha 1} & \cdots & \bar{T}_{\alpha\beta} \end{bmatrix} \quad (13)$$

and

$$\bar{\mathbf{g}}(k) = Q(\mathbf{g}(k)) = \begin{bmatrix} Q(g_1) \\ \vdots \\ Q(g_\beta) \end{bmatrix} = \begin{bmatrix} \bar{g}_1 \\ \bar{g}_2 \\ \vdots \\ \bar{g}_\beta \end{bmatrix} \quad (14)$$

Let the vector $\boldsymbol{\nu}(k)$ be defined by

$$\boldsymbol{\nu}(k) = \Gamma(\bar{\mathbf{T}})\Gamma(\bar{\mathbf{g}}(k)) \quad (15)$$

The i -th entry $\nu_i(k)$ of the vector $\boldsymbol{\nu}(k)$ is obtained by the matrix-vector product $\Gamma(\bar{\mathbf{T}})\Gamma(\bar{\mathbf{g}}(k))$ as

$$\nu_i(k) = \sum_{j=1}^{\beta} \Gamma(\bar{T}_{ij})\Gamma(\bar{g}_j(k)), \quad (16)$$

where $i = 1, 2, \dots, \alpha$, $\bar{T}_{ij}$ denotes the entry in the i -th row and the j -th column of the matrix $\bar{\mathbf{T}}$ and $\bar{g}_j(k)$ denotes the j -th entry in the column vector $\bar{\mathbf{g}}(k)$. Define a Heaviside function

$$\sigma(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases} \quad (17)$$

According to (11), we have

$$\begin{aligned} \nu_i(k) &= \sum_{j=1}^{\beta} \Gamma(\bar{T}_{ij})\Gamma(\bar{g}_j(k)) \\ &= \sum_{j=1}^{\beta} (2^{\lambda_2} \bar{T}_{ij} \bmod \mu) (2^{\lambda_2} \bar{g}_j(k) \bmod \mu) \end{aligned} \quad (18)$$

As $\bar{T}_{ij} \in \mathcal{Q}$ and $\bar{g}_j(k) \in \mathcal{Q}$, both $2^{\lambda_2} \bar{T}_{ij}$ and $2^{\lambda_2} \bar{g}_j(k)$ are integers in the interval $[-2^{\lambda_1+\lambda_2}, 2^{\lambda_1+\lambda_2}]$. Since $\mu = \beta 2^{2(\lambda_1+\lambda_2)+1} + 1$ and $\beta \geq 1$ is an integer, $\lambda_1 \in \mathbb{N}$, $\lambda_2 \in \mathbb{N}$, there is

$$-\mu < 2^{\lambda_2} \bar{T}_{ij} < \mu, \quad -\mu < 2^{\lambda_2} \bar{g}_j(k) < \mu \quad (19)$$

It gives

$$\begin{aligned} 2^{\lambda_2} \bar{T}_{ij} \bmod \mu &= \begin{cases} 2^{\lambda_2} \bar{T}_{ij}, & \text{if } \bar{T}_{ij} \geq 0 \\ \mu + 2^{\lambda_2} \bar{T}_{ij}, & \text{if } \bar{T}_{ij} < 0 \end{cases} \\ 2^{\lambda_2} \bar{g}_j(k) \bmod \mu &= \begin{cases} 2^{\lambda_2} \bar{g}_j(k), & \text{if } \bar{g}_j(k) \geq 0 \\ \mu + 2^{\lambda_2} \bar{g}_j(k), & \text{if } \bar{g}_j(k) < 0 \end{cases} \end{aligned} \quad (20)$$

With the help of the Heaviside function σ defined in (17), we have

$$\begin{aligned} 2^{\lambda_2} \bar{T}_{ij} \bmod \mu &= (1 - \sigma(\bar{T}_{ij}))\mu + 2^{\lambda_2} \bar{T}_{ij} \\ 2^{\lambda_2} \bar{g}_j(k) \bmod \mu &= (1 - \sigma(\bar{g}_j(k)))\mu + 2^{\lambda_2} \bar{g}_j(k) \end{aligned} \quad (21)$$

Substituting (21) into (18), we get

$$\begin{aligned} \nu_i(k) &= \sum_{j=1}^{\beta} [(1 - \sigma(\bar{T}_{ij}))\mu + 2^{\lambda_2} \bar{T}_{ij}] \\ &\quad \cdot [(1 - \sigma(\bar{g}_j(k)))\mu + 2^{\lambda_2} \bar{g}_j(k)] \\ &= \sum_{j=1}^{\beta} (1 - \sigma(\bar{T}_{ij})) (1 - \sigma(\bar{g}_j(k))) \mu^2 \\ &\quad + \sum_{j=1}^{\beta} (1 - \sigma(\bar{T}_{ij})) \mu 2^{\lambda_2} \bar{g}_j(k) \\ &\quad + \sum_{j=1}^{\beta} (1 - \sigma(\bar{g}_j(k))) \mu 2^{\lambda_2} \bar{T}_{ij} + \sum_{j=1}^{\beta} 2^{2\lambda_2} \bar{T}_{ij} \bar{g}_j(k) \end{aligned} \quad (22)$$

Then

$$\nu_i(k) \bmod \mu = \left(\sum_{j=1}^{\beta} 2^{2\lambda_2} \bar{T}_{ij} \bar{g}_j(k) \right) \bmod \mu \quad (23)$$

As mentioned before,

$$\begin{aligned} 2^{\lambda_2} \bar{T}_{ij} &\in [-2^{\lambda_1+\lambda_2}, 2^{\lambda_1+\lambda_2}] \\ 2^{\lambda_2} \bar{g}_j(k) &\in [-2^{\lambda_1+\lambda_2}, 2^{\lambda_1+\lambda_2}] \end{aligned} \quad (24)$$

it holds

$$2^{2\lambda_2} \bar{T}_{ij} \bar{g}_j(k) = 2^{\lambda_2} \bar{T}_{ij} 2^{\lambda_2} \bar{g}_j(k) \in (-2^{2(\lambda_1+\lambda_2)}, 2^{2(\lambda_1+\lambda_2)}] \quad (25)$$

and thus

$$\nu_i(k) = \sum_{j=1}^{\beta} 2^{2\lambda_2} \bar{T}_{ij} \bar{g}_j(k) \in (-\beta \cdot 2^{2(\lambda_1+\lambda_2)}, \beta \cdot 2^{2(\lambda_1+\lambda_2)}] \quad (26)$$

Recall that $\mu = \beta 2^{2(\lambda_1+\lambda_2)+1} + 1$ and $\beta \geq 1$ is an integer, $\lambda_1 \in \mathbb{N}$, $\lambda_2 \in \mathbb{N}$. Then $\frac{1}{2}\mu = \beta 2^{2(\lambda_1+\lambda_2)} + \frac{1}{2}$, from which we can see that

$$0 < \beta 2^{2(\lambda_1+\lambda_2)} < \frac{1}{2}\mu \quad (27)$$

From (23) and (27), we obtain

$$\begin{aligned} \nu_i(k) \bmod \mu &= \begin{cases} \mu + \nu_i(k) \in (\frac{1}{2}\mu, \mu), & \text{if } \nu_i(k) \in (-\beta \cdot 2^{2(\lambda_1+\lambda_2)}, 0) \\ \nu_i(k) \in [0, \frac{1}{2}\mu), & \text{if } \nu_i(k) \in [0, \beta \cdot 2^{2(\lambda_1+\lambda_2)}] \end{cases} \end{aligned} \quad (28)$$

Considering the inverse mapping function (12), it yields

$$\begin{aligned} \Gamma^{-1}(\nu_i(k)) &= \begin{cases} \frac{1}{2^{2\lambda_2}}(\nu_i(k) \bmod \mu - \mu), & \text{if } \nu_i(k) \bmod \mu > \frac{\mu}{2} \\ \frac{1}{2^{2\lambda_2}}(\nu_i(k) \bmod \mu), & \text{if } \nu_i(k) \bmod \mu \leq \frac{\mu}{2} \end{cases} \\ &= \begin{cases} \frac{1}{2^{2\lambda_2}}(\mu + \nu_i(k) - \mu), & \text{if } \nu_i(k) \bmod \mu > \frac{\mu}{2} \\ \frac{1}{2^{2\lambda_2}}\nu_i(k), & \text{if } \nu_i(k) \bmod \mu \leq \frac{\mu}{2} \end{cases} \\ &= \frac{1}{2^{2\lambda_2}}\nu_i(k) = \frac{1}{2^{2\lambda_2}} \sum_{j=1}^{\beta} 2^{2\lambda_2} \bar{T}_{ij} \bar{g}_j(k) = \sum_{j=1}^{\beta} \bar{T}_{ij} \bar{g}_j(k) \end{aligned} \quad (29)$$

As (29) holds for any $i = 1, 2, \dots, \alpha$, there is

$$\Gamma^{-1}(\boldsymbol{\nu}(k)) = \bar{\mathbf{T}}\bar{\mathbf{g}}(k) \quad (30)$$

Fig. 2. Encrypted controller of a CPS with the RHE scheme

That means, if the mapping function (11) and the inverse mapping function (12) are used, then

$$\Gamma^{-1}\left(\Gamma(\bar{T})\Gamma(\bar{g}(k))\right) = \bar{T}\bar{g}(k) \quad (31)$$

holds.

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 $\bar{T} = Q(T)$. Then, the quantized matrix $\bar{T}$ is mapped element-wise by the mapping function $\Gamma(\bar{T})$ 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 $w(k) = [\Gamma^T(Q(x_s)) \ \Gamma^T(Q(y))]^T$. Before $w(k)$ is sent over the network to the controller, $w(k)$ is encrypted to $w_c(k)$ in (5). On the controller side, $w_c(k)$ is processed by the evaluation procedure $c_x^*(k) = F_c w_c^*(k) \bmod pq$ 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, as shown below, if the parameters of the quantization are appropriately selected, the quantization error is bounded and

$$2^{-\lambda_2} \max_j \sum_{k=0}^{\infty} \|((A_\Sigma - B_\Sigma C_\Sigma)^k B_\Sigma)^T e_j\|_1 < 2^{\lambda_1}, \quad (32)$$

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

$$A_\Sigma = \begin{bmatrix} O & O \\ O & A \end{bmatrix}, B_\Sigma = \begin{bmatrix} I & O \\ O & B \end{bmatrix} \bar{T}, C_\Sigma = \begin{bmatrix} I & O \\ O & C \end{bmatrix} \quad (33)$$

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 λ_1, λ_2 are selected to be large enough, the closed-loop system is stable [8].

In order to illustrate the basic idea and for the sake of clarity, in the following derivation it is assumed that the plant in (1) is simplified by $d(k) = 0$ and $D = O$.

The quantized controller can be written as

$$\begin{bmatrix} x_s(k+1) \\ u(k) \end{bmatrix} = \begin{bmatrix} \bar{A}_s & \bar{B}_s \\ \bar{C}_s & \bar{D}_s \end{bmatrix} \bar{g}(k) \quad (34)$$

Let

$$x_\Sigma(k) = \begin{bmatrix} x_s(k) \\ x(k) \end{bmatrix} \quad (35)$$

Due to the plant in (1) and the quantized controller in (34), the dynamics of the closed-loop system is governed by

$$\begin{aligned} x_\Sigma(k+1) &= \begin{bmatrix} O & O \\ O & A \end{bmatrix} x_\Sigma(k) + \begin{bmatrix} I & O \\ O & B \end{bmatrix} \begin{bmatrix} x_s(k+1) \\ u(k) \end{bmatrix} \\ \begin{bmatrix} x_s(k) \\ y(k) \end{bmatrix} &= \begin{bmatrix} I & O \\ O & C \end{bmatrix} x_\Sigma(k) \end{aligned} \quad (36)$$

The closed-loop system in (36) can be equivalently re-written as

$$\begin{aligned} x_\Sigma(k+1) &= A_\Sigma x_\Sigma(k) + B_\Sigma (\Delta_g(k) + C_\Sigma x_\Sigma(k)) \\ &= (A_\Sigma - B_\Sigma C_\Sigma) x_\Sigma(k) + B_\Sigma \Delta_g(k), \end{aligned} \quad (37)$$

where $\Delta_g(k)$ denotes the quantization error given by

$$\Delta_g(k) = \bar{g}(k) - g(k) = \bar{g}(k) - C_\Sigma x_\Sigma(k) \quad (38)$$

Now, the closed-loop system in (37) is in a form to exploit set invariance theory that provides stability conditions for control systems with bounded disturbances (see, for instance, [16], [17]). Recall that after the quantization the signals take value in a finite set $\mathcal{Q} = \{-2^{\lambda_1}, -2^{\lambda_1} + 2^{-\lambda_2}, \dots, 2^{\lambda_1} - 2^{-\lambda_2}\}$, where λ_1 and λ_2 represents, respectively, the range and the resolution of the quantization. Let the set

$$\mathcal{G} = \{\Delta_g(k) \in \mathbb{R}^{\beta \times 1} \mid \|\Delta_g(k)\|_\infty \leq 2^{-\lambda_2}\}. \quad (39)$$

The maximal quantization error in (39) of the signals in the closed-loop system in (37) is equal to or smaller than the resolution $2^{-\lambda_2}$, if the state $x_\Sigma(k)$ is

$$x_\Sigma(k) \in \mathcal{X}_\Sigma = \{x_\Sigma(k) \in \mathbb{R}^{\beta \times 1} \mid \|x_\Sigma(k)\|_\infty \leq \chi\} \quad (40)$$

where $\chi \in \mathbb{R}$ denote the upper bound of $\|x_\Sigma(k)\|_\infty$, constrained by $0 < \chi < 2^{\lambda_1}$. To make sure that the state $x_\Sigma(k)$ is inside the set $\mathcal{X}_\Sigma$ at all time instants, the criterion given in [8] is used. If a non-empty robust positively invariant (RPI) set $\mathcal{K}$ exists and satisfies

$$\emptyset \subset \mathcal{K} \subseteq \mathcal{X}_\Sigma \text{ and } (A_\Sigma - B_\Sigma C_\Sigma)\mathcal{K} + B_\Sigma \mathcal{G} \subseteq \mathcal{K} \quad (41)$$

and the matrix

$$A_\Sigma - C_\Sigma B_\Sigma = \begin{bmatrix} -\bar{A}_s & -\bar{B}_s C \\ -\bar{B}_s \bar{C}_s & A - B \bar{D}_s C \end{bmatrix} \quad (42)$$

is Schur stable, then the closed-loop system in (37) is stable, because the quantization error $\Delta_g(k)$ (39) belongs to the set $\mathcal{G}$ in (39). Note that a RPI set formalizes the set of states that can be reached from the origin under a bounded state disturbance (see [16], [17] for details on RPI sets). Each trajectory that starts in the largest set $\mathcal{K}$ satisfying (41), converges to the minimal RPI set $\mathcal{S}_\infty \subseteq \mathcal{X}_\Sigma$ [16]. Thus,

$$\mathcal{S}_{k+1} = (A_\Sigma - B_\Sigma C_\Sigma)\mathcal{S}_k \oplus B_\Sigma \mathcal{G} \text{ with } \mathcal{S}_0 = \{0\}, \quad (43)$$

Fig. 3. Attacks on the signals encrypted by the RHE scheme

where $\mathcal{S}_{k+1}$ describes the sequence of the set $\mathcal{S}_\infty$ and $\mathcal{Y} \oplus \mathcal{Z}$ denotes the Minkowski sum that adds each vector in $\mathcal{Y}$ to each vector in $\mathcal{Z}$. Recall that χ is some upper bound of the state $\mathbf{x}_\Sigma(k) \in \mathcal{X}_\Sigma$. To make sure that the set $\mathcal{S}_\infty$ are entirely contained in $\mathcal{X}_\Sigma$, a hyperplane can be defined as follows.

$$h_s(\mathcal{S}_\infty, \mathbf{e}_j) = \sum_{k=0}^{\infty} h_s(\mathcal{G}, ((\mathbf{A}_\Sigma - \mathbf{B}_\Sigma \mathbf{C}_\Sigma)^k \mathbf{B}_\Sigma)^T \mathbf{e}_j) \leq \chi \quad (44)$$

where $h_s(\mathcal{S}_\infty, \mathbf{e}_j)$ is the so-called supporting function that describes the distance of the hyperplane of $\mathcal{S}_\infty$ from the origin and $\mathbf{e}_j \in \mathbb{R}^{\beta \times 1}$ is the unit vector, which denotes the j -th column of an identity matrix of dimensions $\beta \times \beta$. The equation (44) can be simplified, if the equation between (30) and (31) given in [8] is applied on (44), which gives

$$h_s(\mathcal{G}, ((\mathbf{A}_\Sigma - \mathbf{B}_\Sigma \mathbf{C}_\Sigma)^k \mathbf{B}_\Sigma)^T \mathbf{e}_j) = \sum_{k=0}^{\infty} \|((\mathbf{A}_\Sigma - \mathbf{B}_\Sigma \mathbf{C}_\Sigma)^k \mathbf{B}_\Sigma)^T \mathbf{e}_j\|_1, \quad (45)$$

where $\|\mathbf{v}\|_1 = \sum_{j=1}^{\beta} |v_j|$ denotes the 1-norm of the vector $\mathbf{v} = [v_1, v_2, \dots, v_\beta]^T$. Thus,

$$\begin{aligned} h_s(\mathcal{S}_\infty, \mathbf{e}_j) &= 2^{-\lambda_2} \max_{j \in \{1, 2, \dots, \beta\}} \sum_{k=0}^{\infty} \|((\mathbf{A}_\Sigma - \mathbf{B}_\Sigma \mathbf{C}_\Sigma)^k \mathbf{B}_\Sigma)^T \mathbf{e}_j\|_1 \\ &\leq \chi \end{aligned} \quad (46)$$

Note that the summation term in (46) can be calculated with the methods provided by [16], [17].

In summary, the closed-loop system in (37) influenced by the quantization error $\Delta_g(k)$ in (39) is stable, if the following conditions are fulfilled:

(Q1) λ_1 is selected so that the upper bound of the state $\mathbf{x}_\Sigma(k)$ satisfies $\chi < 2^{\lambda_1}$

(Q2) λ_2 is chosen so that condition (46) holds

(Q3) λ_1 and λ_2 are chosen so that $\mathbf{A}_\Sigma - \mathbf{C}_\Sigma \mathbf{B}_\Sigma$ is Schur

As stated in [8], the conditions (Q1)-(Q3) can be guaranteed, if λ_1 and λ_2 are large enough.

III. RESILIENCE RANGE OF THE RHE SCHEME

Let $\mathbf{a}_1$ and $\mathbf{a}_2$ 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

$$-\mathbf{h} \leq \mathbf{a}_1 \mathbf{G} + \mathbf{a}_2 < \theta^2 \mathbf{1} - \mathbf{h}. \quad (47)$$

In this section, we shall show that the resilience range of the RHE scheme is indeed much larger than (47), which significantly increases the attractiveness of the RHE scheme for the encryption of control systems.

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

$$\begin{aligned} \mathbf{F}_c &= \text{Enc}(\mathbf{F}, \theta, p) = (\mathbf{S}_f + \mathbf{F}\theta + \mathbf{R}_f p) \bmod pq \\ \mathbf{w}_c &= \text{Enc}(\mathbf{w}, \theta, p) = (\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p) \bmod pq, \end{aligned} \quad (48)$$

Let $\mathbf{a}_1 \in \mathbb{N}^{\beta \times 1}$ and $\mathbf{a}_2 \in \mathbb{N}^{\alpha \times 1}$ denote additive attacks on, respectively, the ciphertexts $\mathbf{w}_c$ and $\mathbf{c}_x^*$, i.e.

$$\mathbf{w}_c^* = \mathbf{w}_c + \mathbf{a}_1, \quad \mathbf{c}_x = \mathbf{c}_x^* + \mathbf{a}_2, \quad (49)$$

where $\mathbf{c}_x^* = (\mathbf{F}_c \mathbf{w}_c^*) \bmod pq$. If $\|(\mathbf{S}_f + \mathbf{F}\theta + \mathbf{R}_f p)(\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p + \mathbf{a}_1)\|_\infty < pq$, then the result $\mathbf{v}$ obtained by decryption in (6) is

$$\mathbf{v} = \text{Dec}(\mathbf{c}_x, \theta, p) = \mathbf{F}\mathbf{w}, \quad (50)$$

as long as

$$zp - \mathbf{h} \leq \mathbf{a}_1 \mathbf{G} + \mathbf{a}_2 < zp + \theta^2 \mathbf{1} - \mathbf{h}, \quad (51)$$

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

Proof. According to (5), the encryption of $\mathbf{w}$ gives

$$\mathbf{w}_c = \text{Enc}(\mathbf{w}, \theta, p) = (\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p) \bmod pq. \quad (52)$$

where θ is the security parameter and p, q are prime numbers. Due to the additive attack $\mathbf{a}_1$,

$$\begin{aligned} \mathbf{w}_c^* &= \text{Enc}(\mathbf{w}, \theta, p) + \mathbf{a}_1 \\ &= (\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p) \bmod pq + \mathbf{a}_1. \end{aligned} \quad (53)$$

If $\|(\mathbf{S}_f + \mathbf{F}\theta + \mathbf{R}_f p)(\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p + \mathbf{a}_1)\|_\infty < pq$, then the matrix-vector product in (7) $\mathbf{c}_x^*$ gives

$$\begin{aligned} \mathbf{c}_x^* &= \mathbf{F}_c \mathbf{w}_c^* \bmod pq \\ &= (\mathbf{S}_f + \mathbf{F}\theta + \mathbf{R}_f p)(\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p + \mathbf{a}_1) \end{aligned} \quad (54)$$

Because of the attack $\mathbf{a}_2$, there is $\mathbf{c}_x = \mathbf{c}_x^* + \mathbf{a}_2$. Note that

$$\begin{aligned} \mathbf{c}_x \bmod p &= ((\mathbf{S}_f + \mathbf{F}\theta + \mathbf{R}_f p)(\mathbf{s}_w + \mathbf{w}\theta + \mathbf{r}_w p + \mathbf{a}_1) \\ &\quad + \mathbf{a}_2) \bmod p \\ &= (\underbrace{\mathbf{a}_1 (\mathbf{S}_f + \theta \mathbf{F})}_{=\mathbf{G}} + \underbrace{\mathbf{S}_f \mathbf{s}_w + \theta(\mathbf{S}_f \mathbf{w} + \mathbf{F} \mathbf{s}_w)}_{=\mathbf{h}} + \theta^2 \mathbf{F} \mathbf{w} \\ &\quad + p(\mathbf{S}_f \mathbf{r}_w + \mathbf{R}_f (\mathbf{s}_w + \mathbf{a}_1)) + \theta(\mathbf{F} \mathbf{r}_w + \mathbf{R}_f \mathbf{w}) \\ &\quad + p \mathbf{R}_f \mathbf{r}_w + \mathbf{a}_2) \bmod p \end{aligned} \quad (55)$$

If the attacks $\mathbf{a}_1$ and $\mathbf{a}_2$ satisfy (51), then $zp \leq \mathbf{a}_1 \mathbf{G} + \mathbf{h} + \mathbf{a}_2 < zp + \theta^2 \mathbf{1}$. Let

$$\mathbf{a}_1 \mathbf{G} + \mathbf{h} + \mathbf{a}_2 = zp + \Delta, \quad (56)$$

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

$$\begin{aligned} \mathbf{c}_x \bmod p &= (\mathbf{a}_1 \mathbf{G} + \mathbf{h} + \mathbf{a}_2 + \theta^2 \mathbf{F} \mathbf{w}) \bmod p \\ &= (zp + \Delta + \theta^2 \mathbf{F} \mathbf{w}) \bmod p = \Delta + \theta^2 \mathbf{F} \mathbf{w}. \end{aligned} \quad (57)$$

Due to the decryption (6), it yields

$$\begin{aligned} \mathbf{v} &= (\mathbf{c}_x \bmod p - (\mathbf{c}_x \bmod p) \bmod \theta^2) / \theta^2 \\ &= (\Delta + \theta^2 \mathbf{F} \mathbf{w} - (\Delta + \theta^2 \mathbf{F} \mathbf{w}) \bmod \theta^2) / \theta^2 = \mathbf{F} \mathbf{w}, \end{aligned} \quad (58)$$

Therefore,

$$e_1(k) \neq 0 \text{ and } \phi_1 = 1, \quad (73)$$

as long as $\Delta_\theta(k) \neq \mathbf{S}_f \mathbf{s}_w(k) + \Delta_a(k)$, where $\langle \mathbf{s}_{pre}, \Delta_a(k) \rangle = 0$. Recall that the entries of the vector $\mathbf{s}_w(k)$ in (72) are randomly chosen integers and change with time, its impossible for the attacker to reach $\Delta_\theta(k) = \mathbf{S}_f \mathbf{s}_w(k) + \Delta_a(k)$ at each time step. Since $\boldsymbol{\nu}(k) = \mathbf{F}\mathbf{w}(k)$, there is

$$\begin{aligned} e_2(k) &= \langle \mathbf{w}(k), \mathbf{F}^T \mathbf{s}_{pre} \rangle - \langle \mathbf{s}_{pre}, \boldsymbol{\nu}(k) \rangle \\ &= \langle \mathbf{w}(k), \mathbf{F}^T \mathbf{s}_{pre} \rangle - \langle \mathbf{s}_{pre}, \mathbf{F}\mathbf{w}(k) \rangle = 0 \\ \phi_2(k) &= 0 \end{aligned} \quad (74)$$

Case III: The attacks $\mathbf{a}_1(k)$, $\mathbf{a}_2(k)$ are outside of the resilience range, i.e. $\mathbf{a}_1(k)\mathbf{G} + \mathbf{h}(k) + \mathbf{a}_2(k) \notin [zp, zp + \theta^2\mathbf{1}]$. Let

$$\mathbf{a}_1\mathbf{G} + \mathbf{h} + \mathbf{a}_2 = zp + l\theta^2 + \Delta \quad (75)$$

where $\mathbf{l}(k) \in \mathbb{N}^{\alpha \times 1}$ whose entries $l_i(k) \in [-\nu_i(k), \lfloor \frac{p}{\theta^2} \rfloor - \nu_i(k)]$, $\nu_i(k)$ are the entries of the vector $\boldsymbol{\nu}(k) = \mathbf{F}\mathbf{w}(k)$ and $\lfloor \cdot \rfloor$ is the operator that rounds to the nearest integer toward negative infinity. Because of the attacks $\mathbf{a}_1(k)$ and $\mathbf{a}_2(k)$, it holds

$$\begin{aligned} c_\times \bmod p &= (\underbrace{\mathbf{a}_1\mathbf{G} + \mathbf{h} + \mathbf{a}_2}_{\notin [zp, zp + \theta^2\mathbf{1}]} + \theta^2\mathbf{F}\mathbf{w}) \bmod p \\ &= (zp + l\theta^2 + \Delta + \theta^2\mathbf{F}\mathbf{w}) \bmod p \\ &= \Delta + \theta^2(\mathbf{F}\mathbf{w} + \mathbf{l}), \end{aligned} \quad (76)$$

where $\|l\theta^2 + \Delta + \theta^2\mathbf{F}\mathbf{w}\|_\infty < p$. By taking into account (59), it yields

$$\begin{aligned} \hat{\mathbf{s}} &= (c_\times \bmod p) \bmod \theta \\ &= (\Delta + \theta^2(\mathbf{F}\mathbf{w} + \mathbf{l})) \bmod \theta = \Delta_l \\ e_1(k) &= \langle \mathbf{s}_w(k), \mathbf{S}_f^T \mathbf{s}_{pre} \rangle - \langle \mathbf{s}_{pre}, \Delta_l(k) \rangle \end{aligned} \quad (77)$$

where $\Delta_l(k) \in \mathbb{N}^{\alpha \times 1}$ whose entries $\Delta_{li}(k) \in \{0, 1, \dots, \theta - 1\}$. If $\Delta_\theta(k) \neq \mathbf{S}_f \mathbf{s}_w(k) + \Delta_a(k)$, where $\langle \mathbf{s}_{pre}, \Delta_a(k) \rangle = 0$, then

$$e_1(k) \neq 0 \text{ and } \phi_1(k) = 1. \quad (78)$$

Due to (76), the decryption of $c_\times(k)$ by (6) gives

$$\begin{aligned} \boldsymbol{\nu} &= \text{Dec}(c_\times, \theta, p) = (c_\times \bmod p - (c_\times \bmod p) \bmod \theta^2) / \theta^2 \\ &= (\Delta + \theta^2(\mathbf{F}\mathbf{w} + \mathbf{l}) - (\Delta + \theta^2(\mathbf{F}\mathbf{w} + \mathbf{l})) \bmod \theta^2) / \theta^2 \\ &= (\Delta + \theta^2(\mathbf{F}\mathbf{w} + \mathbf{l}) - \Delta) / \theta^2 = \mathbf{F}\mathbf{w} + \mathbf{l} \end{aligned} \quad (79)$$

From (9) and (62), it yields

$$\begin{aligned} e_2(k) &= \langle \mathbf{w}(k), \mathbf{F}^T \mathbf{s}_{pre} \rangle - \langle \mathbf{s}_{pre}, \boldsymbol{\nu}(k) \rangle \\ &= \langle \mathbf{w}(k), \mathbf{F}^T \mathbf{s}_{pre} \rangle - \langle \mathbf{s}_{pre}, \mathbf{F}\mathbf{w}(k) + \mathbf{l}(k) \rangle \\ &= \langle \mathbf{s}_{pre}, \mathbf{l}(k) \rangle \end{aligned} \quad (80)$$

Thus, if $\langle \mathbf{s}_{pre}, \mathbf{l}(k) \rangle \neq 0$, then

$$e_2(k) \neq 0 \text{ and } \phi_2(k) = 1. \quad (81)$$

D. Influence of attacks outside of the resilience range

As can be seen from (79), if the attacks $\mathbf{a}_1(k)$, $\mathbf{a}_2(k)$ break the resilience range, then the vector $\boldsymbol{\nu}(k)$ got by the decryption is not equal to the true value of the vector $\mathbf{F}\mathbf{w}(k)$ any more. Now we check how big $\mathbf{l}(k)$ (i.e. the difference between $\boldsymbol{\nu}(k)$ and the true value of $\mathbf{F}\mathbf{w}(k)$) is.

If $zp + \theta^2\mathbf{1} - \mathbf{h} \leq \mathbf{a}_1\mathbf{G} + \mathbf{a}_2 < zp + 2\theta^2\mathbf{1} - \mathbf{h}$, then $\mathbf{l} = \mathbf{1}$. If $zp - \theta^2\mathbf{1} - \mathbf{h} \leq \mathbf{a}_1\mathbf{G} + \mathbf{a}_2 < zp - \mathbf{h}$, then $\mathbf{l} = -\mathbf{1}$. That means, the entries of the vector $\boldsymbol{\nu}$ obtained by decryption differ from the true values by 1.

According to (75), there is

$$\mathbf{l} = (\mathbf{a}_1\mathbf{G} + \mathbf{a}_2 + \mathbf{h} - \Delta - zp) / \theta^2 \quad (82)$$

Recall that θ is the security parameter prespecified during the key generation and is usually a large number, $0 \leq \Delta \leq \theta^2\mathbf{1} - \mathbf{1}$ and $\|\mathbf{h}\|_\infty = \|\mathbf{S}_f \mathbf{s}_w + \theta(\mathbf{S}_f \mathbf{w} + \mathbf{F}\mathbf{s}_w)\|_\infty < \theta^2$. It holds

$$\begin{aligned} -(\theta^2\mathbf{1} - \mathbf{1}) &\leq \mathbf{h} - \Delta \leq \theta^2\mathbf{1} - \mathbf{1} \\ \Leftrightarrow -\mathbf{1} &< (\mathbf{h} - \Delta) / \theta^2 < \mathbf{1} \end{aligned} \quad (83)$$

If $\mathbf{a}_1\mathbf{G} + \mathbf{a}_2 \geq zp + 2\theta^2\mathbf{1} - \mathbf{h}$, then $\mathbf{l}$ is a vector with positive integer entries that always satisfies

$$\mathbf{l} \leq (\mathbf{a}_1\mathbf{G} + \mathbf{a}_2 - zp) / \theta^2 + \mathbf{1} \ll \mathbf{a}_1\mathbf{G} + \mathbf{a}_2 \quad (84)$$

If $\mathbf{a}_1\mathbf{G} + \mathbf{a}_2 < zp - \theta^2\mathbf{1} - \mathbf{h}$, then $\mathbf{l}$ is a vector with negative integer entries and satisfy

$$\mathbf{l} \geq (\mathbf{a}_1\mathbf{G} + \mathbf{a}_2 - zp) / \theta^2 - \mathbf{1} \quad (85)$$

In these two scenarios, the difference between the value $\boldsymbol{\nu}$ got after the decryption and the true value of $\mathbf{F}\mathbf{w}$ is $\mathbf{l} = \boldsymbol{\nu} - \mathbf{F}\mathbf{w}$. Because θ is a large number, $|\mathbf{l}| \ll |\mathbf{a}_1\mathbf{G} + \mathbf{a}_2|$. That means, the influence of attacks outside of the resilience range will be significantly reduced.

In summary, as soon as the attacks $\mathbf{a}_1(k)$ and $\mathbf{a}_2(k)$ are injected into the encrypted signals, the residual signal $e_1(k)$ got by (60) will deviate from zero and the warning signal $\phi_1(k)$ will be triggered to signify the existence of the attacks. If the attacks $\mathbf{a}_1$, $\mathbf{a}_2$ are outside of the resilience range (i.e. $\mathbf{a}_1(k)\mathbf{G} + \mathbf{h}(k) + \mathbf{a}_2(k) \notin [zp, zp + \theta^2\mathbf{1}]$), the residual signal $e_2(k)$ obtained by (62) will deviate from zero and triggers the alarm signal $\phi_2(k)$. Note that, because $|\mathbf{l}(k)| \ll |\mathbf{a}_1(k)\mathbf{G} + \mathbf{a}_2(k)|$, the difference between $\boldsymbol{\nu}(k)$ got after the decryption and the true value of $\mathbf{F}\mathbf{w}(k)$ caused by the attacks $\mathbf{a}_1(k)$ and $\mathbf{a}_2(k)$ is significantly reduced. As a result, the control system can still keep much of the control performance.

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 $[x_1 \ x_2 \ x_3 \ x_4]^T = [12.6 \ 13 \ 4.8 \ 4.9]^T \text{ cm}$. 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

$$A = \begin{bmatrix} 0.9843 & 0 & 0.0251 & 0 \\ 0 & 0.9891 & 0 & 0.0176 \\ 0 & 0 & 0.9747 & 0 \\ 0 & 0 & 0 & 0.9823 \end{bmatrix}, B = \begin{bmatrix} 0.0478 & 0.0010 \\ 0.0005 & 0.0348 \\ 0 & 0.0776 \\ 0.0554 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \end{bmatrix}, D = O, E_d = [B \ O], F_d = [O \ I] \quad (86)$$

To stabilize the system, the parameters of the controller defined in (3) are selected as

$$A_s = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B_s = \begin{bmatrix} 0.0625 & 0 \\ 0 & 0.01563 \end{bmatrix}$$

$$C_s = \begin{bmatrix} 0.032 & 0 \\ 0 & -0.0192 \end{bmatrix}, D_s = \begin{bmatrix} 1 & 0 \\ 0 & -0.3 \end{bmatrix} \quad (87)$$

It can be checked that $A_\Sigma - C_\Sigma B_\Sigma$ is Schur stable.

The maximal signal value in the system is smaller than 15. Thus, select $\lambda_1 = 4$. In order to satisfy condition (46), the integer λ_2 is chosen as $\lambda_2 = 16$, which gives $\mu = \beta 2^{2(\lambda_1 + \lambda_2) + 1} + 1 = 4 \times 2^{41} + 1$. As a result, the bound M of the plaintext is $M = 2^{\lambda_1 + \lambda_2 + 1} = 2^{21}$. 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 θ are chosen from the interval $p \in (2^{189}, 2^{190})$, $q \in (2^{379}, 2^{380})$ and $\theta \in (2^{50}, 2^{51})$.

The parameters of the controller in (87) 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 (60), (62) are randomly chosen from the interval $\{0, 1, \dots, S\}$.

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 $c_x(k)$ 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 (87) 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 $k = 150s$

Fig. 5. Encrypted controller in the CPS with the RHE scheme

to $k = 195s$, the attack signal $a_2(k)$ is inside of the resilience range, because $a_2(k) < \theta^2 \mathbf{1} - \mathbf{h}(k)$, where $\mathbf{h}(k) = S_f s_w(k) + \theta(S_f w(k) + F s_w(k))$. 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 $\phi_1(k)$ 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 $a_2(k) \geq \theta^2 \mathbf{1} - \mathbf{h}(k)$ 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 $\frac{1}{2^{2\lambda_2} \theta^2} > \frac{1}{2^{32} \times 2^{100}}$. 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 trustable 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 to 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. Batterham, "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. 24 325–24 339, 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] 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.
- [17] C. Wang and C.-J. Ong, "Support function of minimal disturbance invariant set and its derivative: Application in designing feedback gain," in *Proc. of 8th Asian Control Conf.*, Kaohsiung, Taiwan, 2011, pp. 1000–1005.
- [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.