2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Effects of Quantization on Zero-Dynamics Attacks to Closed-loop Sampled-data Control Systems

Xile Kang and Hideaki Ishii

Abstract— **This paper focuses on cyber-security issues of networked control systems in closed-loop forms from the perspective of quantized sampled-data systems. As sampling can introduce non-minimum phase zeros in discretized systems, we consider zero dynamics attacks, which is a class of false data injection attacks. Quantization of control inputs disables such attacks to be made exactly, resulting in certain errors in the system output. Specifically, we characterize a trade-off relation between attack performance and stealthiness, and then show that the attacker can reduce the output error with a modified approach by considering the quantization error of the attack signal. We provide a numerical example to demonstrate the effectiveness of the proposed approaches.**

I. Introduction

Networked control systems play a central role in various industries. By reducing wiring and maintenance costs and enhancing system flexibility and efficiency, they are critical in large-scale control systems for power grid, mining facilities, automation, and so on. In recent years, advances in electronics and communications have increased the use of wireless networks, enabling further connectivity and integration with the Internet of Things (IoT). As a consequence, small and medium size businesses also benefit from wider level of field automation through digital transformation.

On the other hand, cyber-security issues arise together with networked control (see, e.g., [8]). Well-known cyber-attack incidents include Stuxnet attacks on nuclear facilities in Iran and ransom attacks on the Colonial Pipeline [2]. Malicious malwares like Triton can also disable control operations running on industrial networks [1]. Devices connected to the Internet require extensive studies on both possible attack approaches and defensive countermeasures. Typical classes of cyber-attacks are denial-of-services attacks, false data injections, and replay attacks [13], [16].

This paper studies a type of false data injection attacks known as zero-dynamics attacks, which require full knowledge of plant models (see, e.g., [8], [16]). Such attack signals are generated by a special dynamics based on unstable zeros of the plant. They cause the plant state to converge to the attack dynamics while maintaining the system output to be normal at sampling instances. As the attack dynamics diverge at the rate of unstable plant zeros, such attacks can cause physical damages by forcing the plant state to become large while remaining undetected by conventional safety countermeasures.

X. Kang and H. Ishii are with the Department of Computer Science, Tokyo Institute of Technology, Yokohama, Japan. E-mails: kang.x.aa@m.titech.ac.jp, ishii@c.titech.ac.jp

This work was supported in part by JSPS under Grant-in-Aid for Scientific Research Grant No. 22H01508.

We focus on the effects of quantization in a sampled-data control systems. Such a system comprises a continuous-time plant and a digital controller whose output is quantized and transmitted via a network, which can be compromised by an attacker. We aim to formulate a feasible approach from the attacker's perspective on this type of systems.

In sampled-data control systems, sampling can introduce unstable zeros known as sampling zeros [3], [19]. Even if the continuous-time plant is minimum phase, sampling zeros can be unstable if the relative degree of the plant is larger than or equal to 3 [19]. As a result, zero-dynamics attacks can be a threat to real systems and require our research attention. A new version of quantized zero-dynamics attacks called ϵ-stealthy attacks has been introduced in [10]. This attack method shares a concept similar to dynamical quantization from [4] to compensate quantization errors. Such an error and system output variation can be designed based on the attacker's needs, and thus may be more difficult to detect.

We however must note that the study in [10] is limited to open-loop systems, and quantized zero-dynamics attacks to closed-loop sampled-data systems still lack research. We will analyze how quantization error in attack signals can affect the control loop. Since such attacks are visible from the system output, a certain modification will be employed in a closed-loop environment.

This paper is organized as follows. Section II describes the sample-data control system and zero-dynamics attacks. In Section III, we evaluate the plant output to characterize the effects of quantization on attack signals. Section IV explores a new variation of zero-dynamics attacks called ϵ-stealthy attacks. Section V presents a numerical example illustrating the effectiveness of the proposed attack approaches. In Section VI, we provide some concluding remarks.

Notation: We denote by $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{Z}_+$ the sets of real numbers, integers, and nonnegative integers, respectively, and $\mathbb{R}^{n\times m}$ is the set of $n \times m$ matrices. Let $d\mathbb{Z}$ be the set of numbers which can be written as dz with $z \in \mathbb{Z}$. The absolute value and 2-norm are given by $|\cdot|$ and $\|\cdot\|$, respectively. Moreover, for a given matrix $A = [a_{ij}]$, let $\mathrm{abs}(A) \coloneqq [|a_{ij}|]$. The null space of matrix C is written as $\ker C$.

II. Problem Formulation

A. System Setup

Consider the networked control system shown in Fig. 1. The plant is a single-input single-output linear time-invariant (LTI) system. The controller receives the plant output y_b and the reference input r, estimates the plant state as $\hat{x}_p$, and calculates the control input u by a certain control law. This

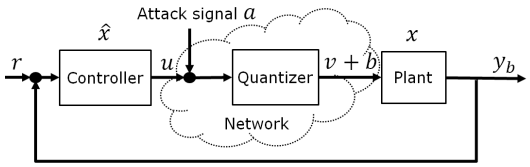


Fig. 1: Closed-loop Networked Control Systems

input u is quantized to v by the quantizer, which outputs a value from a discrete set for transmission over a digital channel. However, this channel is subject to cyber-attacks that falsify the transmitted input with an additive attack signal a. Note that this attack signal is also quantized, resulting in a quantized attack signal denoted by b. Throughout this paper, the subscripts a and b indicate signals and systems affected by the original and the quantized attack signals, respectively.

More specifically, the plant P is assumed to be a discretized system of a continuous-time n-dimensional LTI system. Its state-space equation is expressed as

$$P : \begin{cases} x(k+1) = A_p x(k) + B_p \left(u(k) + a(k)\right), \\ \quad y(k) = C_p x(k), \end{cases} \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the plant state, $u(k) \in \mathbb{R}$ is the control input from the controller, $a(k) \in \mathbb{R}$ is the attack signal, and $y(k) \in \mathbb{R}$ is the output at time $k \in \mathbb{Z}_+$. The system matrices are $A_p \in \mathbb{R}^{n\times n}$, $B_p \in \mathbb{R}^{n\times 1}$, and $C_p \in \mathbb{R}^{1\times n}$. Here, we assume that the plant is controllable and $C_p B_p \neq 0$.

As of the controller C, we estimate the plant state by the plant output and calculate control input $u(k)$ based on the estimated state as

$$C : \begin{cases} \hat{x}(k+1) = A_K \hat{x}(k) + B_{1K} r(k) + B_{2K} y(k), \\ \quad u(k) = C_K \hat{x}(k) + D_{1K} r(k), \end{cases} \tag{2}$$

where $A_K \in \mathbb{R}^{n\times n}$, B_{1K} and $B_{2K} \in \mathbb{R}^{n\times 1}$, $C_K \in \mathbb{R}^{1\times n}$, and $D_{1K} \in \mathbb{R}$. We design the observer gain B_{2K} such that $\hat{x}(k)$ asymptotically converges to $x(k)$, and the state feedback gain C_K to make $A_p + B_p C_K$ stable.

For the quantizer, we utilize the static uniform quantizer $q : \mathbb{R} \to d\mathbb{Z}$ with the step $d > 0$. We follow the nearest neighbor quantization towards $-\infty$: For any value μ, the quantized value $q(\mu)$ satisfies $|q(\mu) - \mu| \leq \frac{d}{2}$. This is a round-to-nearest-d quantizer in real implementation. The controller output $u(k)$ is quantized into $v(k)$:

$$v(k) = q(u(k)). \tag{3}$$

This will be applied to the plant in (1) instead of $u(k)$.

B. Zero-Dynamics Attacks

We now look at how the control system can be attacked by zero-dynamics attacks [16]. It is well known that when discretizing a linear system, if the sampling period is small enough, unstable sampling zeros might appear [3]. To this end, we assume that the plant (1) is non-minimum phase, i.e., it has unstable sampling zeros. In a sampled-data controller, the plant output during the intervals between sampling instances is not measured nor processed, which can be exploited by attackers. Generally, under zero-dynamics attacks, the plant state may diverge in both discrete time and continuous time, but the plant output at each sampling instance remains as if only the original input $u(k)$ is applied. So it is difficult to detect such attacks from the plant output.

Zero-dynamics attacks are a kind of false injection attacks to the signals in control systems. Here, we consider the case, where the control input $v(k)$ is attacked via the attack signal $a(k)$. However, since $v(k)$ takes a quantized value, the attack signal must be quantized to the same value set $d\mathbb{Z}$ as well. The effects of quantization have been studied in an open-loop system setting in [10]. In our paper, we extend the analysis to the closed-loop case. This is not straightforward as in the non-quantized case [16] since the error introduced by quantization in the attack signal will remain inside the system in the closed-loop case.

We employ the class of quantized attacks from [10]. This can be generated by the system given by

$$\begin{aligned} h(k+1) &= \left(A_p - \frac{B_p C_p A_p}{C_p B_p}\right) h(k), \\ a(k) &= -\frac{C_p A_p}{C_p B_p} h(k), \end{aligned} \tag{4}$$

where $h(k) \in \mathbb{R}^n$ and $h(0) \in \ker C_p$ is small enough. The matrix $A_p - \frac{B_p C_p A_p}{C_p B_p}$ is unstable as it has the plant zeros as its eigenvalues. If the attack signal $a(k)$ is added to the control input $u(k)$ in the plant (1), it will force the plant state to converge to $h(k)$ asymptotically. Note that $C_p \left(A_p - \frac{B_p C_p A_p}{C_p B_p}\right) = 0$, so $h(k)$ stays in $\ker C_p$ and thus is undetectable from the plant output.

Since the control input is quantized, the attack signal $a(k)$ must also quantized by q to $b(k)$ as

$$b(k) = q(a(k)) \tag{5}$$

and then applied to the plant (1). This means that the attack may not be as effective as in the original unquantized case because the attack signals are modified by quantization. To quantify the effectiveness of the attacks, we follow the definition in [10] as follows:

Definition 2.1: For a given $\epsilon > 0$, the attack signal is said to be ϵ-stealthy if the following condition is satisfied:

$$|y_b(k) - y(k)| \leq \epsilon, \tag{6}$$

where $y_b(k)$ denotes the system output excited by the quantized attack signal $b(k)$, and $y(k)$ is the system output from the attack-free system.

In this paper, we study the problem of quantized zero-dynamics attacks in the closed-loop setting, which can be stated as follow: Consider the quantized networked system in Fig. 1. Given a positive scalar ϵ, design the quantized attack signal $b(k)$ which causes the state $x(k)$ of the plant to diverge while the ϵ-stealthy condition in (6) is satisfied. Note that without quantization, the original attack $a(k)$ from (4) can achieve 0-stealthiness.

At this point, we write the closed-loop system under quantized attacks as follows. The subscript b in the state, the

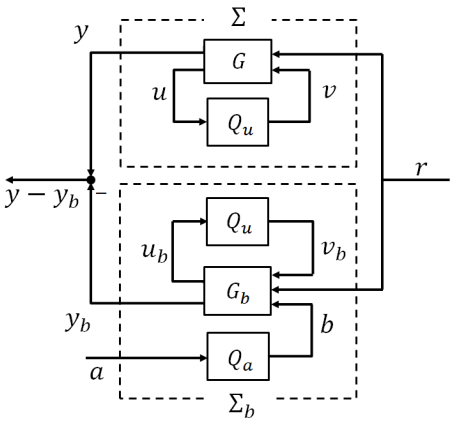


Fig. 2: Error system for Σ and Σ_b

input, and the output refers to the presence of the attacks.

$$G_b : \begin{cases} \overline{x}(k+1) = A\overline{x}(k) + B_1 r(k) + B_2 v_b(k) + B_3 b(k), \\ u_b(k) = C_2\overline{x}(k) + D_2 r(k), \\ v_b(k) = q(u_b(k)), \\ b(k) = q(a(k)), \\ y_b(k) = C_1\overline{x}(k). \end{cases}$$

Here $\overline{x}(k) \coloneqq \left[x^T(k)\ \hat{x}^T(k)\right]^T$ is the combined state of the plant and the controller. The matrices A, B_1, B_2, B_3, C_1, and C_2 are given by

$$A \coloneqq \begin{bmatrix} A_p & 0 \\ B_{2K} & A_K \end{bmatrix},\ B_1 \coloneqq \begin{bmatrix} 0 \\ B_{1K} \end{bmatrix},\ B_2 = B_3 \coloneqq \begin{bmatrix} B_p \\ 0 \end{bmatrix},$$
$$C_1 \coloneqq \begin{bmatrix} C_p & 0 \end{bmatrix},\ C_2 \coloneqq \begin{bmatrix} 0 & C_K \end{bmatrix},\ D_2 \coloneqq D_{1K},$$

with the closed-loop system matrix $\widetilde{A} \coloneqq A + B_2 C_2$.

III. Quantized Zero-Dynamics Attacks

Although zero-dynamics attacks are known to be 0-stealthy for non-quantized systems in the closed-loop setting of Fig. 1, when quantization is introduced in the control signals, concerns about the effectiveness of such attacks arise. In this section, we characterize the effects of quantization in attack signals by utilizing the approach from [6].

A. Dynamic Quantization and Its Effect on Attacks

We use the symbol $y(k, x_0, r, b)$ to denote the plant output at sample time k under the reference input sequence $r := \{r(0), r(1), \ldots\}$ of arbitrary finite values, the initial state $\overline{x}(0) = x_0$, and the attack signal b. For the attack-free system, the output can be expressed as $y(k, x_0, r, 0)$. To characterize the quantization effects in the attacks, we define the performance index to measure the difference:

$$E(b) \coloneqq \sup_{k,r} |y(k, x_0, r, 0) - y(k, x_0, r, b)|. \tag{7}$$

Now, we introduce the approach of dynamic quantization [6]. Such quantization is used for the control input and is represented by the dynamical system with the internal state $\xi_u(k) \in \mathbb{R}^{2n}$ whose output is quantized as

$$Q_u : \begin{cases} \xi_u(k+1) = F_u\xi_u(k) + J_{u1}u(k) + J_{u2}v(k), \\ v(k) = q\left(H_u\xi_u(k) + u(k)\right), \end{cases} \tag{8}$$

where q is the quantizer in (3) with quantization step d. The optimal quantizer to minimize $E(b)$ is $F_u = \overline{A}$, $J_{u1} = -J_{u2} = -B_2$ and $H_u = -\frac{C_1\overline{A}}{C_1 B_2}$, according to [6]. In this case, let the error due to quantization in $u(k)$ be

$$e_u(k) \coloneqq v(k) - \left(H_u\xi_u(k) + u(k)\right).$$

This is bounded as $e_u(k) \in (-d/2, d/2]$. The purpose of dynamic quantization is to minimize the system output variation due to quantization by recording such errors in the internal dynamics ξ_u and adjusting the quantized control input $v(k)$ accordingly. Furthermore, the static quantizer can be regarded as a special case with $H_u = 0$. We must note that dynamic quantization is useful to reduce the impact of quantization in the system.

In this paper, we extend dynamic quantization for the attack signal. In particular, from the perspective of the attacker, we would like to analyze the effectiveness of such quantization in decreasing the error in the plant output so as to reduce the chance to be detected. To this end, the dynamic quantizer Q_a for $a(k)$ is given in the same form and matrices as the one in (8) for the control input as

$$Q_a : \begin{cases} \xi_a(k+1) = F_a\xi_a(k) + J_{a1}a(k) + J_{a2}b(k), \\ b(k) = q\left(H_a\xi_a(k) + a(k)\right). \end{cases} \tag{9}$$

Here, let the quantization error be $e_a(k) := b(k) - \left(H_a\xi_a(k) + a(k)\right)$. Clearly, we have $e_a(k) \in (-d/2, d/2]$. Then, we can combine the closed-loop system G with Q_u and Q_a together as Σ_b (see Fig. 2). We define

$$e_{u,b}(k) \coloneqq v_b(k) - \left(H_u\xi_{u,b}(k) + u_b(k)\right)$$

as the quantization error of $u_b(k)$ in the attacked system Σ_b.

B. Quantized Attacks via Static Quantization

We first present the result for the case of static quantizers in both Q_u and Q_a.

Theorem 3.1: For the quantized system G under zero-dynamics attacks, suppose that the control input and attack signals are quantized as (3) and (5). Then, the performance index in (7) is upper bounded by

$$E_b(Q) \le 3\left\|\sum_{i=0}^{\infty} \mathrm{abs}\left(C_1\overline{A}^i B_2\right)\right\|\frac{d}{2}. \tag{10}$$

Proof: Detailed proof is omitted to save space. ■

When we compare the closed-loop system under attack to the attack-free one, the term $e_{u,b}(k)$ cannot be canceled with $e_u(k)$ because they are independent. The control input $v_b(k)$ is perturbed by $e_a(k)$ and can take a value different from $v(k)$. Hence, the worst-case upper bound of the plant output difference is influenced by both $e_{u,b}(k)$ and $e_u(k)$, and then added up with the influence of $e_a(k)$. We will illustrate such an effect in the next section.

For the attack-free system case, when only the control input is quantized, the performance index becomes

$$E(Q) = \left\|\sum_{i=0}^{\infty} \mathrm{abs}\left(C_1\overline{A}^i B_2\right)\right\|\frac{d}{2} \tag{11}$$

by [6]. This implies that detection of zero-dynamics attacks may be possible by observing the closed-loop system output.

C. Quantized Attacks via Dynamic Quantization

Our next objective is to reduce the output error of a system by utilizing a dynamic quantizer for the attack signal $a(k)$, while keeping the quantizer for the control input $u(k)$ static. We consider the system G, which is identical to G_b but without attack signal $b(k)$. For non-minimum phase plants, we employ the serial decomposition approach proposed by [12]. This involves decomposing plant P in (1) into $P = P_s \cdot P_u$, where P_s represents the minimum phase component and P_u contains the unstable zeros. The realization of P_s is given by $(A_s, B_s, C_s, 0)$ and $C_s B_s \neq 0$. The matrices for the optimal Q_a in (9) are $F_a = A_s$, $J_{a1} = -J_{a2} = -B_s$, and $H_a = -\frac{C_s A_s}{C_s B_s}$, which are shown in [10]. Also, let $\|P_u\|_{i_\infty}$ be the induced l_∞-norm of P_u given by $\|P_u\|_{i_\infty} := \sup_{u \in l_\infty, u \neq 0} \frac{\|P_u u\|_\infty}{\|u\|_\infty}$. We characterize $E_b(Q)$ for the case of dynamic quantization for the attack signal as a corollary of Theorem 3.1.

Corollary 3.1: For the quantized system Σ under zero-dynamics attacks, suppose that the control input and attack signals are quantized, respectively, by (3) and (9) based on the above choice of matrices. Then, the performance index in (7) is upper bounded by

$$E_b(Q) \leq 2 \left\| \sum_{i=0}^{\infty} \text{abs}\left(C_1 \overline{A}^i B_2 \right) \right\| \frac{d}{2} + \left\| \frac{1}{1+PC} \right\|_{i_\infty} \|P_u\|_{i_\infty} |C_s B_s| \frac{d}{2}.$$

For a plant with a single unstable zero, the optimal decomposition is to take P_u in its transfer function form as $P_u(z) = \frac{z - z_0}{z}$, where z_0 is the unstable zero [12]. In such a case, the l_∞-norm of P_u is $\|P_u\|_{i\infty} = 1 + |z_0|$. The performance index of the optimal dynamic quantizer for $a(k)$ of the closed-loop system is worse than that of the open-loop case obtained in [10]. We also note that a countermeasure based on dynamic quantization in the control input is possible but omitted from this paper for space reason.

Finally, we discuss the selection of $h(0)$. By [17], $h(0)$ should be a vector small enough in ker C_p. For an LTI system under attacks, we can separate it into attack-free system G excited by $r(k)$ and the same system under attack with $r(k) = 0$ and zero initial state as

$$\widetilde{G}_b : \begin{cases} \overline{x}(k+1) = \overline{A}\overline{x}(k) + B_3 b(k), \\ y^b(k) = C_1 \overline{x}(k). \end{cases}$$

Then we augment $h(k)$ into $h^a(k)$ to have the same dimension as $\overline{x}$. The augmented zero-dynamics attacks are

$$h^a(k+1) = \left(\overline{A} - \frac{B_3 C_1 \overline{A}}{C_1 B_3} \right) h^a(k),$$
$$a(k) = -\frac{C_1 \overline{A}}{C_1 B_3} h^a(k).$$

Let the error between $\overline{x}(k)$ and $h^a(k)$ be $e^b(k) = \overline{x}(k) - h^a(k)$. We have

$$e^b(k+1) = \overline{x}(k+1) - h^b(k+1) = \overline{A} e^b(k) - B_3 e_a(k).$$

This indicates that the convergence rate of $e^b(k)$ depends on the closed-loop system matrix $\overline{A}$. So we need to select a small $\overline{x}(0)$ to make $e^b(0)$ invisible from $y^b(0)$. Otherwise, there will be a transient response from $\overline{x}(0)$, which will reveal the attack at the system output side.

IV. Quantized ϵ-stealthy Attacks to the Closed-loop System

In this section, we extend the ϵ-stealthy attacks proposed in [10] for open-loop systems to closed-loop systems. Compared to the traditional zero-dynamics attacks, ϵ-stealthy attacks take a small modification with the design parameter ϵ:

$$\begin{aligned} h(k+1) &= \left(A_p - \frac{B_p C_p A_p}{C_p B_p} \right) h(k) + \frac{B_p}{C_p B_p} \epsilon, \\ a(k) &= -\frac{C_p A_p}{C_p B_p} h(k) + \frac{\epsilon}{C_p B_p} \end{aligned} \tag{12}$$

with $h(0) = 0$. When applied to an open-loop system, it is straightforward that the plant state at sample time $k = 1$ differs from the attacked case by $x^b(1) - x(1) = h(0) = \epsilon$. Moreover, we can show by induction that $x^b(k) - x(k) = h(k)$ holds at every sample time k.

The quantized version of ϵ-stealthy attacks is

$$\begin{aligned} h(k+1) &= A_p h(k) + B_p b(k), \\ b(k) &= q\left(-\frac{C_p A_p}{C_p B_p} h(k) + \frac{\text{sgn}(C_p B_p)\epsilon}{C_p B_p} - \frac{d}{2} \right). \end{aligned} \tag{13}$$

Now, let

$$g(k) = \frac{d}{2} + e_a(k) = -\frac{C_p A_p}{C_p B_p} h(k) + \frac{\text{sgn}(C_p B_p)\epsilon}{C_p B_p} - b(k).$$

Clearly, we have $g(k) \in (0, d]$. The quantized ϵ-stealthy attacks dynamics becomes

$$\begin{aligned} h(k+1) =& \left(A_p - \frac{B_p C_p A_p}{C_p B_p} \right) h(k) \\ &+ B_p \left(\frac{\text{sgn}(C_p B_p)\epsilon}{C_p B_p} - g(k) \right), \\ b(k) =& -\frac{C_p A_p}{C_p B_p} h(k) + \frac{\text{sgn}(C_p B_p)\epsilon}{C_p B_p} - g(k). \end{aligned} \tag{14}$$

In contrast to the traditional zero-dynamics attacks, when ϵ-stealthy attacks are applied to the closed-loop system, any error at the plant output side will be fed into the controller. At a certain time, this difference may become visible from the output side. In our study, we focus on the plant state x^b under attacks. Its dynamics is given by

$$x(k) = A_p^k x^b(0) + \sum_{i=0}^{k-1} A_p^{k-1-i} B_p v_b(i) + h(k). \tag{15}$$

We can establish this expression by induction. Since $h(0) = 0$, we have $B_p b(0) = h(1)$. Then

$$x^b(1) = A_p x^b(0) + B_p v(0) + h(1).$$

Assume (15) holds for k. It is easy to verify that it also holds for $k+1$. The first two terms in (15) are the same as in the attack-free case. Therefore, the output difference at time k can be expressed as

$$y_b(k) - y(k) = C_p h(k) = \text{sgn}(C_p B_p)\epsilon - C_p B_p g(k).$$

Let $\hat{x}^b(k)$ be the estimator state for the attacked system. The estimator state under such output difference is

$$\begin{aligned}\hat{x}^b(k+1) &= A_K \hat{x}^b(k) + B_{1K} r(k) + B_{2K} C_p x^b(k) \\ &= A_K \hat{x}^b(k) + B_{1K} r(k) + B_{2K} \Bigg(C_p A_p^k x^b(0) \\ &\quad + \text{sgn}(C_p B_p)\epsilon - C_p B_p g(k) + \sum_{i=0}^{k-1} C_p A_p^{k-1-i} B_p v(i) \Bigg).\end{aligned}$$

The above equation demonstrates that the estimated plant state $\hat{x}^b(k)$ will deviate from the attack-free one $\hat{x}(k)$ due to the quantization error of $a(k)$. Once this difference accumulates sufficiently, the quantized control input $v_b(k)$ will land on a neighboring quantization level. This introduces a step input of amplitude d, causing a transient state in the plant output. Depending on the control system, such transient might be visible and reveal the attack.

To avoid this transient states, we propose a compensation method for ϵ-stealthy attacks. This method requires the knowledge of the initial value of both the plant and the controller, the controller matrices, and the reference $r(k)$ to predict the quantized control input $\hat{v}_b(k)$ for the attacked system and $\hat{v}(k)$ for the attack-free one:

$$\begin{aligned}\hat{x}(k+1) &= A\hat{x}(k) + B_1 r(k) + B_2 \hat{v}(k), \\ \hat{v}(k) &= q\left(C_2 \hat{x}(k) + D_1 r(k)\right), \\ \hat{x}_b(k+1) &= A\hat{x}_b(k) + B_1 r(k) + B_2 \hat{v}_b(k) + B_3 b(k), \\ \hat{v}_b(k) &= q\left(C_2 \hat{x}_b(k) + D_1 r(k)\right), \\ \Delta v(k) &= \hat{v}_b(k) - \hat{v}(k).\end{aligned} \tag{16}$$

Moreover, the control input difference $\Delta v(k)$ will be compensated at the attack signal side as

$$b(k) = -\frac{C_p A_p}{C_p B_p} h(k) + \frac{\text{sgn}(C_p B_p)\epsilon}{C_p B_p} - g(k) + \Delta v(k). \tag{17}$$

Here we cannot use $h(k+1) = \left(A_p - \frac{B_p C_p A_p}{C_p B_p}\right) h(k) + B_p b(k)$ in the compensated attack dynamics as (13) since if $\Delta v(k)$ is fed into $h(k)$, $C_p h(k)$ will be greater than ϵ, violating the definition of ϵ-stealthy attacks. We present our second main result as the following theorem:

Theorem 4.1: For the quantized system G under zero-dynamics attacks, suppose that the control input is quantized as (3) and $\epsilon \geq |C_p B_p| d$. Then, the quantized attack signal $b(k)$ generated by (17) is ϵ-stealthy.

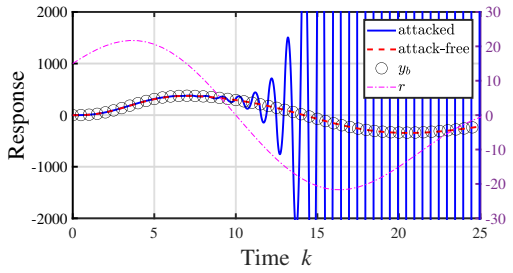


Fig. 3: Output of the closed-loop system under static quantized zero-dynamics attacks

To ensure $h(1) > 0$ and that the attack signal is not zero for $k \geq 1$, we need to have $\frac{\text{sgn}(C_p B_p)\epsilon}{C_p B_p} > d$. Since we choose $h(0) = 0$, this condition implies that $\epsilon \geq |C_p B_p| d$. Otherwise, $h(k)$ will remain zero forever, and there will be no attack signal.

V. Numerical Example

Consider the continuous-time plant with the transfer function given by

$$P(s) = \frac{s^2 + 5.12s + 6.804}{s^4 + 1.868s^3 + 0.645s^2 + 2.22s + 0.804}.$$

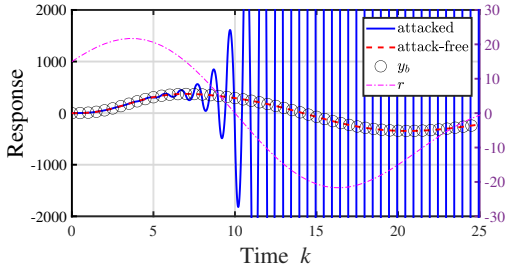
It has unstable poles $0.25 \pm 1.015i$ and stable zeros $-2.56 \pm 0.5i$. By discretizing it with the sampling period of $T_s = 0.5$, an unstable zero -1.608 will appear, and this can be the target of zero-dynamics attacks from (4) and (17). We construct the controller based on a Luenberger observer and estimated state feedback by designing C_K and B_{2K} to have $A_K = A_p + B_p C_K - B_{2K} C_p$. With this controller, the estimated plant state error matrix $A_p - B_{2K} C_p$ has poles $0.085, 0.14, -0.14 \pm 0.107i$, and the estimated state feedback matrix $A_p + B_p C_K$ has poles $0.9, 0.51, 0.83 \pm 0.35i$.

The control reference signal is given by $r(k) = 10\sin(0.1\pi k) + 15\cos(0.05\pi k)$. We select the initial state as $h(0) = [-11.7, 98.6, -6.97, -9.26]^T \times 10^{-4}$, which is a basis of ker C_p and the quantization level as $d = 1$.

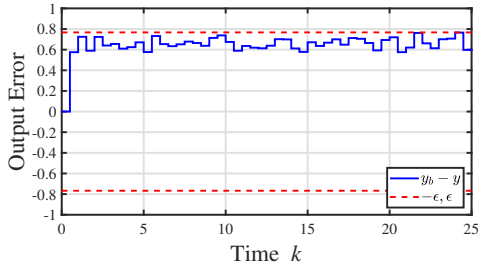
First, we demonstrate the closed-loop system output under zero-dynamics attacks with the static quantizer to both the control input and the attack signal in Fig. 3. The solid line represents the plant output under attacks in continuous time, while the dashed line is the output of attack-free case. The circles are the attacked output at sampling instances. The y-axis on the left side is for the plant output $y(k)$, while the one on the right side is for the reference signal $r(k)$. The output of the dynamic quantized attack signal is very similar to the static quantized case and is not shown here.

Next, we switch to ϵ-stealthy attacks with the same static quantizer and show the result in Fig. 4. The design parameter is selected as $\epsilon = 4 \times |C_p B_p| d = 0.767$. Under such attacks, the plant output is very close to the attack-free case, and we can verify that such variation is bounded by $\pm\epsilon$.

Fig. 5 (a) shows the norm of the plant state $\|x(k)\|$ under zero-dynamics attacks and ϵ-stealthy attacks. Generally, the plant state excited by ϵ-stealthy attacks is larger than that by the zero-dynamics attacks. Fig. 5 (b) demonstrates the output

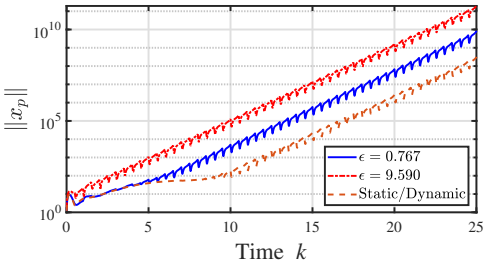


(a) System output in continuous time and discrete time

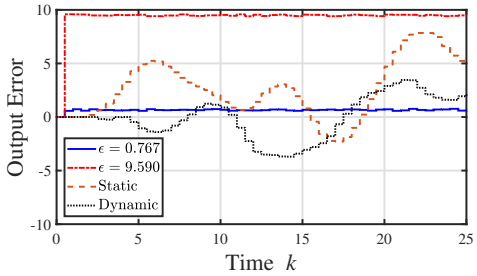


(b) System output error at each sampling time

Fig. 4: ϵ-stealthy attacks with $\epsilon = 0.767$



(a) Plant state norm by attack approaches



(b) Output difference by attack approaches

Fig. 5: Norm of the plant states and output error of zero-dynamics and ϵ-stealthy attacks

difference between static/dynamical quantized traditional and ϵ-stealthy methods with different ϵ. By (10), the upper bound for the output variation caused by static quantization of the attack signal is $E(b) = 3 \times 41.1 \times \frac{d}{2} = 61.6$. That of the dynamic quantized signal is $(2 \times 41.1 + 0.52) \times \frac{d}{2} = 41.3$.

A particular advantage of ϵ-stealthy attacks is that the plant state can reach the physical safety threshold earlier than the traditional zero-dynamics attacks. This is because the attack signal at $k = 0$ is $\frac{\mathrm{sgn}(C_p B_p)}{C_p B_p}\epsilon - g(0)$, which scales with ϵ. The norm of the plant states grows exponentially at the rate determined by the unstable sampling zero. The traditional attack approach needs some time for the plant states to converge to zero dynamics. On the other hand, there is a trade-off in selecting a larger ϵ, since the upper-bound on the variation also becomes more visible.

VI. Conclusion

We have demonstrated the effectiveness of zero-dynamics attacks in quantized closed-loop systems and presented bounds on the effects of quantization in the attack signals to the plant output. Moreover, we have explored ϵ-stealthy attacks on closed-loop systems, which are more difficult to detect from the output side. This approach requires extra knowledge of the initial states of the plant and the controller.

References

[1] National Cyber Security Center. "TRITON Malware Targeting Safety Controllers," 2017. https://www.ncsc.gov.uk/information/triton-malware-targeting-safety-controllers.

[2] Office of Cybersecurity, Energy Security, and Emergency Response. "Colonial Pipeline Cyber Incident," 2021. https://www.energy.gov/ceser/colonial-pipeline-cyber-incident.

[3] K. J. Åström and B. Wittenmark, *Computer Controlled System Theory and Design*, 3rd edition, Dover, 2011.

[4] S. Azuma, T. Sugie. "Optimal dynamic quantizers for discrete-valued input control". *Automatica*, 44, 396–406, 2008.

[5] S. Azuma and T. Sugie, "Synthesis of optimal dynamic quantizers for discrete-valued input control," *IEEE Trans. on Automatic Control*, vol. 53, no. 9, pp. 2064–2075, 2008.

[6] S. Azuma, Y. Minami and T. Sugie. "Optimal dynamic quantizers for feedback control with discrete-level actuators: Unified solution and experimental evaluation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 133, 2011.

[7] J. Back, J. Kim, C. Lee, G. Park and H. Shim, "Enhancement of security against zero dynamics attack via generalized hold," *Proc. IEEE Conference on Decision and Control (CDC)*, pp. 1350–1355, 2017.

[8] H. Ishii and Q. Zhu, Eds. *Security and Resilience of Control Systems, Lecture Notes in Control and Information Sciences*, vol 489. Springer, 2022.

[9] A. Isidori, *Nonlinear Control Systems*, 3rd ed. Springer, 1995.

[10] K. Kimura and H. Ishii, "Quantized zero dynamics attacks against sampled-data control systems," *Proc. IEEE Conference on Decision and Control (CDC)*, pp. 6140–6145, 2022.

[11] Y. Liu, M. Reiter and P. Ning, "False data injection attacks against state estimation in electric power grids," *ACM Trans. on Information and System Security*, vol. 14, pp. 21–32, 2009.

[12] Y. Minami and K. Kashima. "Dynamic quantizer design based on serial system decomposition," *Proc. 22nd Int. Symp. MTNS*, pages 577–579, 2016.

[13] Y. Mo, T. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-physical security of a smart grid infrastructure," *Proc. IEEE*, vol. 100, pp. 262–282, 2012.

[14] G. Park, H. Shim, C. Lee, Y. Eun and K. H. Johansson, "When adversary encounters uncertain cyber-physical systems: Robust zero-dynamics attack with disclosure resources," *Proc. IEEE Conference on Decision and Control (CDC)*, pp. 5085–5090, 2016.

[15] H. Shim, J. Back, Y. Eun, G. Park, J. Kim, "Zero-dynamics attack, variations, and countermeasures," in H. Ishii, Q. Zhu, Eds. *Security and Resilience of Control Systems, Lecture Notes in Control and Information Sciences*, vol 489. Springer, 2022.

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

[17] A. Teixeira, I. Shames, H. Sandberg and K. H. Johansson. "Revealing stealthy attacks in control systems," *Proc. Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, pp. 1806–1813, 2012.

[18] L. Xie, Y. Mo and B. Sinopoli, "False data injection attacks in electricity markets," *First IEEE International Conference on Smart Grid Communications*, Gaithersburg, MD, USA, pp. 226–231, 2010.

[19] J. Yuz and G. Goodwin, *Sampled-Data Models for Linear and Nonlinear Systems*. Springer, 2014.