Zero Assignment via Generalized Sampler: A Countermeasure Against Zero-Dynamics Attack

DAEHAN KIM1, KUNHEE RYU1, (Graduate Student Member, IEEE), JUNG HOON KIM2, (Member, IEEE), AND JUHOON BACK1, (Member, IEEE)

1School of Robotics, Kwangwoon University, Seoul 01897, South Korea
2Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, South Korea

Corresponding author: Juhoon Back (backhoon@kw.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant through the Korea government (MSIT) under Grant 2019R1A4A1029003, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) through the Ministry of Education under Grant NRF-2018R1D1A1B07042833.

ABSTRACT Networked control systems have advantages such as flexibility, efficiency, but at the same time they are exposed to cyberattacks. Among many lethal attacks, the zero-dynamics attack is a model based attack and it is very hard to detect. In this paper, a new strategy for intrusion detection and defense against zero-dynamics attack is proposed, and it is based on the generalized sampler that takes a weighted average of multiple samples obtained during one sampling interval. By using the generalized sampler instead of the simple sampler, the zeros of the sampled-data system can be placed at arbitrary locations, and if all zeros are placed inside the unit circle, the attack signal becomes no longer effective. This strategy still works even if all the information is exposed to hackers and it is considerably insensitive to the shift of intrinsic zeros. A design procedure for the generalized sampler is provided under mild assumptions. Furthermore, optimal designs for the selection of desired zeros are formulated considering practical issues. Theoretical findings are validated through numerical simulations.

INDEX TERMS System security, sampled-data systems, zero assignment, zero-dynamics attack.

I. INTRODUCTION
Advanced communication technologies make it possible to control remote dynamic systems, monitor the status of geographically distributed systems such as power systems and smart factories, and make decisions for multi-agent systems. Although this high level of flexibility has been achieved, these systems are subject to cyber threats, as reported in real incidents such as StuxNet computer worm [1] and the attack on Ukrainian power plant [2].

Thus, ensuring security from cyber threats is a central issue these days and is gaining more and more attention. One of the main research streams on cyber threats is to model and analyze the cyber-attacks (see, e.g., [3]–[8] and references therein) and develop defense strategies against those attacks [3], [9]–[13].

Among many cyber-attacks, we focus on the zero-dynamics attack (ZDA) in this paper. It is one of the most dangerous cyber-attacks because it is stealthy and hardly possible to detect. The attack signal is constructed by using the information of system’s zero-dynamics so that the internal states corresponding to the zero-dynamics converge to that of ZDA while the effect of ZDA on the system output is almost negligible. Thus, ZDA is fatal to systems that have unstable zero-dynamics (i.e., non-minimum phase systems) since a properly designed ZDA can drive the internal states unbounded while unnoticed. It is emphasized that modern networked control systems are vulnerable to ZDA since in many cases the system can be modeled as a sampled-data system with unstable zero-dynamics. In fact, if a continuous-time system has relative degree greater than two and the sampling time is sufficiently small, the corresponding sampled-data system has at least one unstable sampling zero [14], regardless of the stability of the zero-dynamics of the continuous-time systems.

Several strategies for intrusion detection and defense against ZDA have been proposed. [9] proposed a detection method of ZDA by modifying the input, output, and dynamic characteristics of the system. [12] introduced a (constant or time-varying) modulation matrix to the path of the input.
channel to hide the actual input gain matrix. Both approaches can detect the intrusion of ZDA, but the information on modification or modulation matrix should be hidden. [15] proposed a dual-rate control; construct a lifted discrete-time system by collecting a sufficient number of output measurements during a single sampling interval. It is shown that the lifted system has no unstable zeros except 1. Although this approach can be used to detect the intrusion of ZDA, a collection of output measurements should be transmitted, which requires fairly large communication load. Recently, [13] proposed a strategy employing the generalized hold (GH) [14] instead of the zero-order hold (ZOH). If properly designed, the zeros of the sampled-data system can be placed in the stable region so that ZDA is no longer effective.

In this paper, we propose a new strategy for intrusion detection and defense against ZDA. The idea is to shift the zeros of the sampled-data system so that the system has stable zero-dynamics, and this is done by replacing the simple sampler (SS) with the generalized sampler (GS) [14]. GS is a signal processing device that can be represented by a continuous-time system having an impulse response and generates a discrete-time signal from a continuous-time signal [14]. If the delta function is used for the impulse response, GS becomes the conventional SS.

Motivated by the notion of GS, we consider a GS that takes several instantaneous samples during one sampling interval and generates a discrete-time output that is a weighted average of the samples. The number of samples (during one sampling interval) and the weights are the design parameters of GS. Thanks to the additional degree of freedom, the zeros can be placed anywhere desired under mild assumptions, and a constructive design procedure to choose the weights of GS is also presented. We emphasize that the proposed approach has several advantages over existing strategies against ZDA; first, no information needs to be hidden, second, the zeros can be arbitrarily assigned, and finally, there is no intersample behavior. Based on the weight design for a given set of desired zeros, we further investigate how to find optimal zeros. We propose several metrics for the design considering practical issues such as output deviation of the newly generated signal from the original sampled output and the effect of noise. With these metrics, we present optimal designs that can be solved by using well known numerical tools for convex optimization. Numerical simulations on a two-mass system are conducted for noisy/noise-free cases to validate theoretical findings.

The rest of this paper is organized as follows. In Section II, we briefly recall the concept of ZDA and several strategies against it. In Section III, the generalized sampler as a zero assignment tool is introduced in detail and compared with a recently developed assignment method. Designs of GS from optimization perspective are given in Section V. In Section IV, we applied the design of GS to neutralize ZDA and simulation results are included. Finally, Section VI concludes the paper.

Notation: \( \mathbb{R}^n \) denotes an \( n \)-dimensional Euclidean space. The set of natural numbers is denoted by \( \mathbb{N} \) and \( \mathbb{N}_0 := \mathbb{N} \cup \{0\} \). The set of complex numbers is denoted by \( \mathbb{C} \). For a given \( a \in \mathbb{C} \), \(|a|\) denotes the magnitude of \( a \). \( I_n \) is the \( n \times n \) identity matrix and \( 0_n \) denotes the \( n \times 1 \) zero vector.

II. PRELIMINARIES

A. ZERO-DYNAMICS ATTACK ON SAMPLED-DATA SYSTEM

We consider a continuous-time system controlled by a digital controller which is connected through a communication network. Suppose that the network has been compromised so that a malicious attack signal can be injected in the control signal. Precisely, the control system under attack is described by

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) + a(t) \\
y(t) &= Cx(t)
\end{align*}
\]

(1)

where \( x(t) \in \mathbb{R}^n \) is the state vector, \( u(t) \in \mathbb{R} \) is the control input, \( a(t) \in \mathbb{R} \) is the attack signal, and \( y(t) \in \mathbb{R} \) is the output. \( A, B \) and \( C \) are constant matrices with appropriate dimensions.

Remark 1: ZDA is classified as an actuator attack that the attack signal is applied to an input channel of a system. As a counterpart of the actuator attack, there is a sensor attack, which applies an attack to sensor measurements transmitted over the network. Pole-dynamics attack is one of them and it is a dual of ZDA [16].

In most networked control systems, ZOH and SS are used to interface the system (1) with a digital controller that is connected through communication network as depicted in Fig. 1. ZOH located in the input side converts the discrete-time input signal \( u_k := u(kT_s), \forall k \in \mathbb{N}_0 \) coming from the controller into \( u(t) = u_k, kT_s \leq t < (k+1)T_s \), where \( T_s \) is the sampling time, and SS converts the continuous-time signal \( y(t) \) into a discrete-time signal \( y_k := y(kT_s), \forall k \in \mathbb{N}_0 \) at each sampling time.

Suppose that a malicious attack signal intrudes into the compromised network at each sampling time so that the signal transmitted to ZOH becomes \( u_k + a_k \). Thus, we have \( u(t) + a(t) = u_k + a_k, kT_s \leq t < (k+1)T_s \). Then, it follows from the theory of linear system [17] that

\[
x((k+1)T_s) = e^{AT_s}x(kT_s) + \int_{kT_s}^{(k+1)T_s} e^{A((k+1)T_s-\tau)}B(a(\tau) + a(\bar{\tau}))d\tau
\]
where the dynamics of $\xi$ is given by (2) is the sampled-data system (2) [18].

The system (2) can be rewritten in the normal form [18] given by

$$
\begin{align*}
\eta_{k+1} &= S_d \eta_k + P_d C_\xi \xi_k \\
\xi_{k+1} &= A_\xi \xi_k + B_\xi (\phi_d^T \xi_k + \psi_d^T \eta_k + g_d(u_k + a_k)) \\
y_k &= C_\xi \xi_k
\end{align*}
$$

(3)

where the dynamics of $\xi$ explains the relation between the input and output and that of $\eta$ describes the internal behavior. The dynamics $\eta_{k+1} = S_d \eta_k$ is called the zero-dynamics, where the eigenvalues of $S_d$ correspond to the zeros of the sampled-data system (2) [18].

Let us briefly explain how (3) can be derived. See [18, Chapter 13] for more details. Suppose that the transfer function of (2) is given by

$$G(z) = K \frac{N(z)}{D(z)}$$

where $K \in \mathbb{R}$, $D(z)$ and $N(z)$ are monic polynomials whose degrees are $n$ and $n - \rho$, respectively, and $\rho$ is the relative degree. Let $Q(z)$ and $R(z)$ be the polynomials of degree $\rho$ and $R_L \leq \rho - 1$, respectively, such that $D(z) = N(z)Q(z) + R(z)$. Then, $G(z)$ can be written as

$$G(z) = K \frac{N(z)}{N(z)Q(z) + R(z)} = K \frac{1}{1 + \frac{R_L}{N(z)}}$$

It can be seen that $G(z)$ is a negative feedback system composed of $1 / (Q(z))$ in the forward loop (hence its output is $y_k$) and $R_L / N(z)$ (with $y_k$ being its input) in the feedback loop, and $K$ is the input gain. Let a state space realization of $R_L / N(z)$ be given by

$$
\begin{align*}
\eta_{k+1} &= S_d \eta_k + P_d C_\xi \xi_k \\
\eta_k &= \xi_k
\end{align*}
$$

where $S_d$, $P_d$, and $C_\xi$ are monic polynomials whose degrees are $n$, $n - \rho$, and $\rho$, respectively.

It is noted that $G(z)$ is a negative feedback system composed of $1 / (Q(z))$ in the forward loop (hence its output is $y_k$) and $R_L / N(z)$ (with $y_k$ being its input) in the feedback loop, and $K$ is the input gain. Let a state space realization of $R_L / N(z)$ be given by

$$
\begin{align*}
\eta_{k+1} &= S_d \eta_k + P_d C_\xi \xi_k \\
\eta_k &= \xi_k
\end{align*}
$$

(4)

which means that $det(zI - S_d) = N(z)$ and $q_d^T (zI - S_d)^{-1} P_d = R_L / N(z)$. In addition, $1 / (Q(z))$ can be realized in the control canonical form given by

$$
\begin{align*}
\xi_{k+1} &= (A_\xi + B_\xi \phi_d^T) \xi_k + B_\xi (K(u_k + a_k) - u_{\text{d}b,k}) \\
y_k &= C_\xi \xi_k
\end{align*}
$$

(5)

where $\xi, \phi_d, g_d, u_{\text{d}b,k}$ are determined from $Q(z)$, and the matrices and vectors are given by

$$A_\xi = \begin{bmatrix} 0_{\rho-1} & I_{\rho-1} \end{bmatrix}, \quad B_\xi = \begin{bmatrix} 0_{\rho-1} \end{bmatrix}, \quad C_\xi = \begin{bmatrix} 1_{\rho-1} \end{bmatrix}.$$
only on $\xi_k$ which converges to zero. Therefore, for non-minimum phase systems, the internal variable $\eta_k$ becomes unbounded whenever $z_k$ is excited by unstable modes of $S_d$, while $y_k$ converges to zero so that the intrusion of attack cannot be monitored from $y_k$.

The preceding discussion established that the system (1) is vulnerable to ZDA if the sampled-data system has at least one unstable zero and only may think that if the original continuous-time system is of minimum phase, then the system is safe from ZDA. Unfortunately, this is not true because the sampled-data system may have unstable zeros appearing from the sampling procedure. In fact, when the continuous-time system (1) has a relative degree greater than two and the sampling time is sufficiently small, it is inevitable that the sampled-data system has unstable zero-dynamics because at least one of the sampling zeros lies outside the unit circle [14]. Hence, the networked control system is vulnerable to ZDA if there are unstable zeros that come from the unstable zero of the continuous-time system or those emerge from the sampling procedure. This is illustrated in the following example.

Example 1: In this example, we consider that a malicious attacker carries out ZDA to a sampled-data system whose continuous-time system is given by

$$G(s) = \frac{s - 1}{(s + 2)(s + 3)}. \quad (8)$$

Suppose that ZOH and SS are used as sample and hold devices, and $T_s = 0.1$. Then, the zero of the sampled-data system is $z = 1.22$, which comes from the unstable zero of the continuous-time system. In addition, the parameters for ZDA are $g_d = 0.08$, $\psi_d = -0.13$, and $\mu$ is chosen as 0.01.

Fig. 2 shows the result of executing ZDA to the system. The attack is initiated at $t = 2$, and $y(t)$ denotes the continuous-time output of the system and $y_{k,SS}$ denotes the sampled output of $y(t)$ by SS. It can be clearly seen from Fig. 2 that $y(t)$ diverges due to ZDA but, there is no significant change in $y_{k,SS}$ which is transmitted over network. Therefore, it is difficult for the control and monitoring system to recognize that the system is being attacked. This shows the stealthy nature of ZDA.

Example 2: Consider a two-mass system (see Fig. 3) taken from [13], where $m_1 = m_2 = 1$kg, $k_1 = k_2 = 1$N/m and $b_2 = 1$Ns/m. We assume for the time being that $a = 0$. The transfer function from $u$ to $y$ becomes

$$G(s) = \frac{s + 1}{s^4 + 2s^3 + 3s^2 + s + 1}$$

and it is noted that the system is of minimum phase (one zero at $-1$). However, since the relative degree is 3, the sampled-data system under ZOH and SS will become a non-minimum phase system for a sufficiently small sampling time. Fig. 4 shows how the zeros of the sampled-data system vary as $T_s$ changes. As $T_s$ decreases, $z_1$ moves out of the unit circle.

Let $T_s = 0.1$s. Then, the zeros of the sampled-data system are $z_1 = -3.63$, $z_2 = -0.26$, and $z_3 = 0.90$, and the eigenvector corresponding $z_1$ is given by $v_1 = [-0.07, 0.26, -0.96]^T$. From the normal form of the system, the parameters for ZDA are obtained as $g_d = 1.62 \times 10^{-4}$ and $\psi_d = [5.02, 20.04, -28.28]^T$. Fig. 5 shows the outputs of the two-mass system under ZDA constructed by setting $z_{us} = z_1$, $v_{us} = v_1$, and $\mu = 10^{-10}$. Under the attack signal, $y(t)$ becomes unbounded, but the attack can be hardly observed from $y_{k,SS}$. 

![FIGURE 3. Two-mass system under attack a.](Image)

![FIGURE 4. Zeros of the two-mass system for $T_s$.](Image)

![FIGURE 5. Zero-dynamics attack on a minimum phase continuous-time system.](Image)
**B. EXISTING INTRUSION DETECTION AND DEFENSE STRATEGIES**

To enhance security against ZDA, several strategies have been developed. [9] investigated how the system structure affects the stealthiness property of ZDA and proposed a strategy that involves the modification of system structure to reveal the attack. [12] introduced a modulation matrix in the input channel so that actual input gain matrix is hidden from hackers. An optimization based design is proposed and time-varying (periodic) modulation matrix is also considered. Although these approaches can reveal ZDA, they have a drawback that information on modification or modulation matrix should be hidden. Instead of modifying the internal structure, [15] proposed to use dual-rate control. The idea is to obtain a sufficiently large number of measurements during a single sampling interval and consider the collection of the measurements as a new output. They proved that the system with new output has no unstable zeros except 1. It is the main advantage that it is not necessary to hide any information from hackers. However, a large amount of information should be transmitted.

Recently, a new strategy employing the GH [14] has been introduced in [13]. They applied the fact that GH can change the system zeros [14], and suggested to shift all the zeros into the stable region so that the sampled-data system becomes of minimum phase, which makes ZDA ineffective. GH involves a function \( h_k(t) \) so-called hold function that is defined as a piecewise continuous function. One candidate of hold function is a piecewise constant function given by \( h_k(t) = h_i, \frac{(i-1)T_s}{N} \leq t < \frac{iT_s}{N}, \ i = 1, \ldots, N \), where \( h_i \) are constant gains and \( N \) is the number of subintervals. They presented optimal designs of hold function so that the difference between GH and ZOH is reduced as small as possible. Although it has several advantages over other strategies, this approach inherits one drawback of GH that undesirable inter-sample behaviors [14] can be induced. Common instances of the undesirable inter-sample behaviors are overshooting and undershooting of the continuous-time output [19]. The overshooting (or undershooting) between sampling times, a phenomenon in which the continuous-time output between consecutive sampled outputs fluctuates significantly, may bring system damage, such as wearing of a bearing, etc. The level of inter-sample behavior depends on the hold function that is closely related to the choice of desired zeros [14].

**C. GENERALIZED SAMPLER**

GS is basically a signal processing device that converts a continuous-time signal into a discrete-time sequence. We consider the GS introduced in [14], which can be represented by a linear system with an impulse response. By denoting the impulse response as \( h(t) \), the output of GS is given by

\[
\tilde{y}_k := \hat{y}(kT_s) = \int_{(k-1)T_s}^{kT_s} y(\tau) h(kT_s - \tau) d\tau
\]

where \( \tilde{y}_k \) is the output sample generated by GS. GS is naturally a generalization of SS (the one with \( h(t) = \delta(t) \)).

In [14], the author proposed two samplers by specifying the impulse responses of generalized samplers; Piecewise Constant GS (PCGS) and Sinusoidal GS (SGS). In addition, they showed that the PCGS and SGS can move the zeros with an example of a second-order integrator system. For further details, see [14].

**III. WEIGHTED AVERAGING GENERALIZED SAMPLER AND ZERO ASSIGNMENT**

In this section, we introduce a particular generalized sampler called weighted averaging generalized sampler with which the zeros of the sampled-data system can be placed at any desired locations. The proposed generalized sampler takes several measurements during one sampling interval and computes a weighted average. It will be seen that this special structure facilitates the design of generalized sampler from a given set of desired zeros.

**A. WEIGHTED AVERAGING GENERALIZED SAMPLER**

Consider a GS (see Fig. 6) that takes \( N \) measurements \( y(\frac{i}{N}T_s + (k-1)T_s), y(\frac{i}{N}T_s + (k-1)T_s), \ldots, y(kT_s) \) during the sampling interval \( (k-1)T_s, kT_s \), and generates \( \tilde{y}_k \) given by

\[
\tilde{y}_k = \sum_{i=1}^{N} w_i y(\frac{i}{N}T_s + (k-1)T_s)
\]

where \( w_1, \ldots, w_N \) are the weights for GS, i.e., \( h(t) = \sum_{i=1}^{N} w_i \delta(t + \frac{i}{N}T_s - T_s) \). SS can be represented by (9) with \( w_1 = w_2 = \cdots = w_{N-1} = 0 \) and \( w_N = 1 \).

To proceed, we would like to find a sampled-data system whose output is \( \tilde{y}_k \). Since the state vector \( x \) of the linear system (1) (with \( a(t) = 0 \) for simplicity) at time \( t, (k-1)T_s \leq t \leq kT_s \), can be computed as

\[
x(t) = e^{A(t-(k-1)T_s)} x_{k-1} + \int_{(k-1)T_s}^{t} e^{A(t-\tau)} Bu(\tau) d\tau
\]

\[
= e^{A(t-(k-1)T_s)} x_{k-1} + \int_{(k-1)T_s}^{t} e^{A(t-\tau)} Bd\tau u_{k-1}.
\]

We have, with \( t = \frac{i}{N}T_s + (k-1)T_s \),

\[
x(\frac{i}{N}T_s + (k-1)T_s)
\]
\[ w = [\tilde{C}_d, \tilde{D}_d]M^+ \] (15)

where \( M^+ \) is the pseudo-inverse of \( M \).

**Proof:** Let \( \tilde{G}_d^*(z) = \frac{p^*(z)}{q^*(z)} \), and suppose that \( p^*(z) \) and \( q^*(z) \) are given by
\[
p^*(z) = k_d \left( z^n + (\frac{\tau_{n-1}}{k_d} + d_{n-1})z^{n-1} + \cdots + (\frac{\tau_0}{k_d} + d_0) \right)
\]
\[
q^*(z) = z^n + d_{n-1}z^{n-1} + \cdots + d_0. \quad (16)
\]

Then, \( \tilde{G}_d^*(z) \) can be realized in the control canonical form [20] given by
\[
x_k = A_{con}x_{k-1} + B_{con}u_{k-1} \]
\[
y_k = C_{con}x_k + D_{con}u_k \]

where
\[
A_{con} = \begin{bmatrix}
    0_{n-1} & I_{n-1} & \cdots & 0 \\
    -d_0 & -d_{n-1} & \cdots & 0
\end{bmatrix}, 
B_{con} = \begin{bmatrix} 0_{n-1} \\ 1 \end{bmatrix}, 
C_{con} = \begin{bmatrix} c_0 & \cdots & c_{n-2} & c_{n-1} \end{bmatrix}, 
D_{con} = k_d.
\]

Noting that \( G_d^*(z) = z^{-1}G_d^*(z) \), we can realize \( G_d^*(z) \) as
\[
x_k = A_{con}x_{k-1} + B_{con}u_{k-1} \]
\[
y_k = C_{con}x_k + D_{con}u_k \]

It is obvious that \( G_d(z) \) is identical to \( G_d^*(z) \) if and only if \( \tilde{D}_d = D_{con} \) and \( \tilde{C}_d A_d^{k-1}B_d = C_{con}A_{con}^{k-1}B_{con} \), \( k = 1, \ldots, n \) [20]. Since the pair \( (A_d, B_d) \) is controllable, this relation is equivalent to
\[
\tilde{C}_d = C_{con}C_{con} \quad (17)
\]

where \( C_{con} \) is the controllability matrix of \( (A_d, B_d) \) and \( C_{con} \) is that of \( (A_{con}, B_{con}) \), namely,
\[
A_{con} = \begin{bmatrix} B_d & A_dB_d & \cdots & A_d^{n-1}B_d \end{bmatrix}, 
C_{con} = \begin{bmatrix} B_{con} & A_{con}B_{con} & \cdots & A_{con}^{n-1}B_{con} \end{bmatrix}.
\]

Since \( (A_d, B_d) \) is controllable, we have from (17) that \( \tilde{C}_d = C_{con}C_{con}C_d^{-1} \). In addition, it is trivial to see that \( \tilde{D}_d = k_d \). Thus, the existence of \( \tilde{C}_d \) and \( \tilde{D}_d \) is proved.

One can easily show the relation (15) by rewriting (12) as
\[
\hat{C}_d = w [\tilde{C}_d e^{A_{con}^T}] \quad (18)
\]
\[
\hat{D}_d = w [\tilde{D}_d e^{A_{con}^T}]
\]

which completes the proof. \( \square \)

From Lemma 1, we propose a design procedure for GS as follows.

**Design Procedure 1:**
1) Choose the number of subintervals \( N \).
2) Choose \( n \) desired zeros \( z_{d,1}, z_{d,2}, \ldots, z_{d,n} \) and the gain \( k_d \).

3) Compute \( \tilde{C}_d = C_{\text{con}}C_{\text{con}}^{-1} \) and \( \tilde{D}_d = k_d \).

4) The weights \( w_1, \ldots, w_N \) are given by \( w = [\tilde{C}_d, \tilde{D}_d] M^\top \).

Sometimes, it is desirable that \( \sum w_i = 1 \) to ensure that \( \dot{y}_k = y(kT) \) when \( y(t) \) is constant on \((k - 1)T, kT\). In this case, solve (15) with \( k_d = 1 \) and \( w \) replaced by \( w^\top \), and then set \( k_d = 1/(\sum w_i^\top) \) and \( w = k_d w^\top \).

Example 3: In this example, we would like to apply the developed zero assignment method to the two-mass system shown in Fig. 3. Let the desired zero be chosen as \( z_{d,1} = e^{-T_1} \), \( z_{d,2} = z_{d,3} = z_{d,4} = 0 \). Following Design Procedure 1, we have \( w = [-6.75, 17.49, -1.41, -27.64, 19.30] \). It is noted that the zeros of sampled-data system can also be assigned if we employ GH instead of ZOH [13]. Following the design of GH proposed in [13] with the desired zeros \( z_{d,1} = e^{-T_1} \), \( z_{d,2} = z_{d,3} = 0 \), we have a piecewise constant GH with the gain \( h = [20.89, -21.97, 3.14, 1.94]^\top \).

Fig. 7 shows step responses of system: one with GH and SS, and the other with ZOH and GS. The response with GH shows a severe fluctuation in the signal \( \tilde{y}(t) \) (shown in red) between sampling instants. This phenomenon is typically observed when GH is used instead of ZOH because the continuous-time input \( u(t) \) during one sampling interval depends on the time-varying pattern of GH. On the contrary, for the case with GS, the inter-sample behavior is significantly improved (shown in blue) because \( u(t) \) is constant due to ZOH. If the intrinsic zero is shifted by 0.002%, the signal \( \tilde{y}_{\text{GH}}(t) \) fluctuates more severely (shown in Fig. 7a (right)). For the case with GS, \( \tilde{y}_{\text{GS}}(t) \) remains near the system output \( y(t) \) even if the intrinsic zero is shifted by 0.002% and 5% (shown in Fig. 7b).

IV. NEUTRALIZATION OF ZERO-DYNAMICS ATTACK VIA WEIGHTED AVERAGING GENERALIZED SAMPLER

As discussed in Section II-A, ZDA exploits the property that the unstable attack signal, that is constructed using unstable mode of the zero-dynamics, cannot be detected by monitoring the system output, i.e., ZDA is effective for non-minimum phase systems. Thus, if we can place all the zeros inside the unit circle, then zero-dynamics of new sampled-data system has stable zero-dynamics so that ZDA is not effective anymore. As described in Section III, GS can indeed do this, and thus it can be used as a promising security tool against ZDA.

The following result establishes that if we can design a GS so that the system with new output \( \tilde{y}_k \) becomes of minimum phase, then any diverging ZDA is detected by monitoring the signal \( \tilde{y}_k \).

Lemma 2: Suppose that the discrete-time system (2) with ZOH and SS is of non-minimum phase and asymptotically stable. Let \( a_k \) be a ZDA designed for this system. If there exists a GS of the form (9) such that the transfer function (13) with the GS is of minimum phase, then \( \tilde{y}_k \) can detect the ZDA. That is to say, the output \( \tilde{y}_k \) of the system

\[
\begin{align*}
x_{k+1} &= A_d x_k + B_d a_k \\
\tilde{y}_k &= \tilde{C}_d x_{k-1} + \tilde{D}_d a_{k-1}
\end{align*}
\]

becomes unbounded as \( k \) goes to infinity.

Proof: Let \( \tilde{y}(z) \) and \( a(z) \) be the \( z \)-Transform of \( \tilde{y}_k \) and \( a_k \), respectively. From the dynamics attack signal (6), one has

\[
a(z) = -\frac{1}{g_d} \Psi_d^\top (zI - S_d)^{-1} z_0 =: \frac{p_a(z)}{q_a(z)}
\]

where \( z_0 \) is the initial condition of attack dynamics. Then, from (14), \( \tilde{y}(z) \) becomes

\[
\tilde{y}(z) = G_d^\ast(a(z)) = \frac{p^\ast(z)}{z q^\ast(z)} q_a(z)
\]

where \( p^\ast(z) = (z-z_{d,1}) \cdots (z-z_{d,n}) \) with \( z_{d,1}, \ldots, z_{d,n} \) being the new zeros determined by the weights \( w_1, \ldots, w_N \) of GS. By assumption, all the zeros \( z_{d,1}, \ldots, z_{d,n} \) are located inside the unit circle, while at least one root of \( q_a(z) \) is unstable (\( S_d \) is unstable). If this root is denoted by \( p_a \), then, \( |p_a| > 1 \) and \( \tilde{y}_k \) contains a term \( e^{\rho_k} \), \( \rho_k \neq 0 \), which diverges. This completes the proof. \( \square \)

As a security tool, GS has several advantages over existing strategies. First, we do not need to hide key information such as the weights and the number sub-intervals. Indeed, suppose the hacker has full information on the system and GS. If a ZDA is constructed based on this information, the attack will
converge to zero since the target sampled-data system has stable zero-dynamics. In the case that the attacker assumes that ZOH and SS are used and a ZDA is constructed using this information, then the zero-dynamics will be quite different from the real one, hence it will be detected by the new signals from GS. Second, there is no theoretical limit on the choice of zeros under mild conditions (see Lemma 1) and this is because the degree of freedom for the design of GS is quite large. It is noted that the approach by [15] guarantees that the lifted system has no unstable zero except 1, but they did not provide a zero assignment strategy. Third, as discussed in Example 3, the proposed approach does not induce violent inter-sample behavior that is a drawback of the GH based approach. It is also observed that the proposed approach is remarkably insensitive to the shift of intrinsic zeros compared to [13].

Remark 2: The proposed solution does not require any additional sensor to measure state variables. This means that the system structure remains unchanged and this is in sharp contrast to the solutions such as [9], [12] that involve modification of system structure. Meanwhile, the approaches that modify system structure using additional sensors (or actuators) can change controllability and observability of a system, but this does not mean that the zeros can be assigned as desired. In this paper, we restrict ourselves that the system structure is fixed and would like to focus on the zero assignment. Clearly, combining the idea of modification of system structure and zero assignment is a very interesting future research topic.

In the next example, we illustrate that ZDA can be detected by using GS.

Example 4: Consider the system discussed in Example 3. The sampled-data system under ZOH and SS has unstable zero-dynamics. Using this information, a ZDA is constructed as shown in Fig. 8a. As can be seen from Fig. 8b, the sampled output $y_{k,SS}$ generated by SS still remains zero and the ZDA is not detected, while the continuous-time output diverges. Now we consider the case with GS that has been designed with the desired zeros chosen in Example 3. With this GS, the attack is clearly detected (Fig. 8b) because the sampled output $\tilde{y}_{k,Zd}$ generated by GS also diverges.

V. OPTIMAL DESIGN OF WEIGHTED AVERAGING GENERALIZED SAMPLER

As shown in the previous section, we can assign zeros by employing GS and the weights are design parameters determined from the desired zeros. Although the location of zeros can be arbitrary, we consider the case where the desired zeros lie inside the unit disk so that the approach can be applied to develop a countermeasure against ZDA.

In this section, we develop an optimal design for the desired zeros that minimize the discrepancy between the output from SS ($y_k$) and that of GS ($\tilde{y}_k$). In addition, optimal designs considering the size of the weights are also covered. It is noted that the proposed designs are in fact solved by convex optimization tools [21].
\[ k_d = \tilde{D}_d = wD_{d,N}. \]

Applying the sufficient condition for Schur stability derived by [22], i.e., a polynomial \( \alpha_n \varepsilon^0 + \alpha_{n-1} \varepsilon^{n-1} + \cdots + \alpha_0 \) is Schur stable if \( \sum_{i=0}^{n-1} |\alpha_i| \leq |\alpha_0| \), we arrive at a constraint for stable zero-dynamics given by

\[ \sum_{i=0}^{n-1} |w(\Omega_{i+1} + D_{d,N}d_i)| < |wD_{d,N}| \tag{20} \]

where \( \Omega_i \) is the \( i \)th column of \( C_{d,N}C_\text{con}^{-1} \).

The resulting optimization problem for the zero-assignment problem is then described by

\[
\begin{align*}
\min_w f(w) \\
\text{s.t.} \quad \sum_{i=0}^{n-1} |w(\Omega_{i+1} + D_{d,N}d_i)| < |wD_{d,N}|. 
\end{align*}
\tag{P.1}
\]

Note that the problem is not convex due to the inequality constraint. However, as described in the work by [13], we can obtain a practical solution by solving two convex optimization problems and take the one with smaller \( f(w) \); one with constraints \( \sum_{i=0}^{n-1} |w(\Omega_{i+1} + D_{d,N}d_i)| \leq wD_{d,N} - \epsilon \) and \( wD_{d,N} \geq \epsilon \), and the other with \( \sum_{i=0}^{n-1} |w(\Omega_{i+1} + D_{d,N}d_i)| \leq -wD_{d,N} - \epsilon \) and \( wD_{d,N} \leq -\epsilon \), where \( \epsilon \) is a small positive number. See [13] for more details.

**Example 5:** In this example, we continue Example 3 to illustrate that the GS obtained by solving the optimization problem (P.1) results in smaller output deviation compared to the one from the GS with pre-determined desired zeros. With \( \epsilon = 10^{-10} \) and \( N = 5 \), the convex optimization solver CVX [21] gives the optimal weight \( w^{*}_{P1} \approx [-8.57, 30.81, -36.94, 15.69, 0.00] \). The corresponding zeros of the sampled-data system are \( e^{-T_5}, -0.01, \) and \( -0.52 \pm j0.82 \), which make the zero-dynamics Schur stable. Fig. 9 shows the system output \( y(t) \) and the signals obtained by GSs when the unit step input is applied to the sampled-data system. In the figure, \( y_k, z_k \) denotes the output of GS, which is obtained by following **Design Procedure 1** with the set of desired zeros \( Z_4 = \{e^{-T_l}, 0, 0, 0\} \), and \( y^*_k, P1 \) denotes that of the optimal GS for the problem (P.1).

It is clearly seen that \( y^*_k, P1 \) is closer to \( y_k \) than \( y_k, z_k \). The values of the objective function for the cases are \( f(w_{Z_4}) = 9.7 \times 10^{-3} \) and \( f(w^{*}_{P1}) = 2.36 \times 10^{-6} \), where \( w_{Z_4} \) denotes the sampling weight associated to \( Z_4 \).

**B. OPTIMAL DESIGN WITH REGULARIZATION**

In practice, it is inevitable that the measured output of the system is contaminated by sensor noise. Since the new output \( \tilde{y}_k \) is a weighted average of multiple measurements, large weights will amplify the measurement noise.

Motivated by this, we introduce a regularization term [23] to have new objective function \( f(w) = f(w) + \gamma \|w\|_2^2 \) where \( \gamma > 0 \) is a scaling parameter. In addition, another constraint \( w1_N = 1, \) \( I_N := [1 \cdots 1]^T \in \mathbb{R}^N \), is introduced to ensure that two signals generated by GS and SS are identical for the case of constant output.

Taking both the regularization term and the constraint for constant measurement into account, we have the second optimization problem as

\[
\begin{align*}
\min_w \tilde{f}(w) &= f(w) + \gamma \|w\|_2^2 \\
\text{s.t.} \quad \sum_{i=0}^{n-1} |w(\Omega_{i+1} + D_{d,N}d_i)| < |wD_{d,N}|, \\
& \quad w1_N = 1. 
\end{align*}
\tag{P.2}
\]

One may explicitly add a constraint concerning the size of \( w \) instead of including the regularization term in the cost function. Denoting the maximum allowable size of \( \|w\|_2^2 \) by \( \delta > 0 \), we have

\[
\begin{align*}
\min_w f(w) \\
\text{s.t.} \quad \sum_{i=0}^{n-1} |w(\Omega_{i+1} + D_{d,N}d_i)| < |wD_{d,N}|, \\
& \quad w1_N = 1, \\
& \quad \|w\|_2^2 \leq \delta. 
\end{align*}
\tag{P.3}
\]

**Example 6:** In this example, we investigate the effect of measurement noise on GS in terms of false alarm. For the system considered in previous examples, we assume that the measurement of GS is contaminated by noise \( v_{k,l} \), applied at \( t = kT_s + lT_s/N, k = 0, 1, \ldots, N-1, \) and that \( v_{k,l} \) is zero-mean white Gaussian with variance \( R \). We consider an attack detector employing GS, which decides whether the system is under attack or not. When the output of GS exceeds a threshold, it is determined that the system is under attack and an alarm is raised. Under this setting, we define a false alarm as a case that the output signal exceeds the threshold even though there is no attack, or that an attack is present but the output signal is within the threshold [24], [25].

Fig. 10a shows the effect of noise with \( R = 10^{-5} \) for the system considered in Example 4 under a GS designed in that example. It is seen that several samples of \( \tilde{y}_k, z_k \) exceed the threshold although there is no attack, leading to a false alarm.
This is because large sampling weights of GS amplify the effect of noise. This undesirable situation can be avoided by the optimized design with regularization.

Fig. 10b shows the sampled output from GS that is designed by solving (P.2). The parameter $\gamma$ is chosen as $5.0 \times 10^{-4}$ and the optimal weights of GS are determined as $w_{P2}^* = [2.11, -0.59, -2.05, -1.25, 2.78]$. It is seen that the effect of noise is reduced so that $\hat{y}_{k,P2}$ lies within the threshold before an attack is injected, and the attack is successfully detected.

Table 1 shows the false alarm rate of each GS, which is examined through repeated simulations. In each simulation, we examine the situation for 5 seconds each, before and after the attack, and the number of false alarm is counted. This simulation repeats three times and the false alarm rate is obtained by putting collected over the entire simulation. Four GSs, the one obtained by Design Procedure 1 and the others from (P.1) to (P.3), are compared. The weights of GSs are denoted by $w_{Z,d}$, $w_{P1}^*$, $w_{P2}^*$, and $w_{P3}^*$, where the first two weights are those obtained in Examples 4 and 5, the third one obtained above in this example, and the last one, obtained by solving (P.3), is $w_{P3}^* = [0.10, 2.59, -1.95, -4.31, 4.56]$ ($\delta$ is chosen as 50). As can be seen from Table 1, the false alarm rate is substantially decreased when the size of weight is included in the cost function or a constraint on the size is imposed.

VI. CONCLUSION

A new countermeasure against the zero-dynamics attack has been proposed. It employs the generalized sampler, which takes a weighted average of inter-samples, instead of simple sampler that is frequently used in practice. Although this approach shares the same idea of zero assignment with the generalized hold based approach, the proposed strategy seems to be more effective since the unfavorable inter-sample behavior can be avoided. Compared to other approaches, the proposed idea does not need to hide information on the system and the generalized sampler, which is an additional benefit. Optimal designs for GS considering the output deviation and the effect of noise have been provided, and it is illustrated that the designs can be directly applied to neutralize the zero-dynamics attack.

We are currently working on a robust zero assignment problem that can allow system uncertainties and plan to conduct real experiments. Extension to MIMO systems and application to output feedback stabilization of non-minimum phase systems are also interesting future research topics.

Example 7: In this example, we would like to illustrate that the regularization term added to the objective function can be used to reduce the effect of noise. We consider the case mentioned in Example 5 and assume that process noise and measurement noise, denoted by $w_{k,l}$ and $v_{k,l}$, are added to the state and output at $t = kT_s + lT_s/N$, $k = 0, 1, \ldots, l = 0, 1, \ldots, N - 1$, where $w_{k,l}$ and $v_{k,l}$ are zero-mean white Gaussian with variance $0.1 I_4$ and $10^{-3}$, respectively.

Fig. 11 shows the output signals generated by four GSs used in the previous example where $\hat{y}_{k,P2}^*$ and $\hat{y}_{k,P3}^*$ denote the system output at $t = kT_s + lT_s/N$ corrupted with noise $v_{k,l}$. It can be observed from Fig. 11 that the impulse response of the system ($y_{k,l}$) is slightly different from the case without process noise (See Fig. 9). It can be observed that the effect of noise is substantially reduced in the signals $\hat{y}_{k,P2}^*$ and $\hat{y}_{k,P3}^*$ compared to the signals $\hat{y}_{k,Zd}$ and $\hat{y}_{k,P1}^*$. It is mainly because the weights $w_{Zd}$ and $w_{P1}^*$ are relatively very larger than $w_{P2}^*$ and $w_{P3}^*$.
ACKNOWLEDGMENT

A preliminary version of this article was presented at the 21st World Congress of the International Federation of Automatic Control (IFAC), July 13–17, 2020, Germany [26].

REFERENCES

[1] J. P. Farwell and R. Rohozinski, “Stuxnet and the future of cyber war,” Survival, vol. 53, no. 1, pp. 23–40, 2011.
[2] Analysis of the Cyber Attack on the Ukrainian Power Grid, Defense Use Case, Electr. Inf. Sharing Anal. Center (E-ISAC), Washington, DC, USA, 2016.
[3] Y. Mo and B. Sinopoli, “Secure control against replay attacks,” in Proc. 47th Ann. Allerton Conf. Commun., Control, Comput. (Allerton), Sep. 2009, pp. 911–918.
[4] A. Gupta, C. Langbort, and T. Basar, “Optimal control in the presence of an intelligent jammer with limited actions,” in Proc. 49th IEEE Conf. Decision Control, Dec. 2010, pp. 1096–1101.
[5] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, “Attacks against process control systems: Risk assessment, detection, and response,” in Proc. 6th ACM Symp. Inf. Comput. Commun. Secur. (ASIACCS), New York, NY, USA, 2011, pp. 355–366.
[6] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “A secure control framework for resource-limited adversaries,” Automatica, vol. 51, pp. 135–148, Jan. 2015.
[7] M. Jin, J. Lavaei, and K. H. Johansson, “Power grid AC-based state estimation: Vulnerability analysis against cyber attacks,” IEEE Trans. Autom. Control, vol. 64, no. 5, pp. 1784–1799, May 2019.
[8] L. Cao, X. Jiang, Y. Zhao, S. Wang, D. You, and X. Xu, “A survey of network attacks on cyber-physical systems,” IEEE Access, vol. 8, pp. 44219–44227, 2020.
[9] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “Revealing stealthy attacks in control systems,” in Proc. 50th Ann. Allerton Conf. Commun., Control, Comput. (Allerton), Oct. 2012, pp. 1806–1813.
[10] F. Pasqualetti, F. Dörfler, and F. Bullo, “Attack detection and identification in cyber-physical systems,” IEEE Trans. Autom. Control, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.
[11] Y. Chen, S. Kar, and J. M. Moura, “Dynamic attack detection in cyber-physical systems with side initial state information,” IEEE Trans. Autom. Control, vol. 62, no. 9, pp. 4618–4624, Sep. 2017.
[12] A. Hoehn and P. Zhang, “Detection of covert attacks and zero dynamics attacks in cyber-physical systems,” in Proc. Amer. Control Conf. (ACC), Jul. 2016, pp. 302–307.
[13] J. Kim, J. Back, G. Park, C. Lee, H. Shim, and P. G. Voulgaris, “Neutralizing zero dynamics attack on sampled-data systems via generalized holds,” Automatica, vol. 113, Mar. 2020, Art. no. 108778.
[14] J. I. Yuz and G. C. Goodwin, Sampled-Data Models for Linear and Nonlinear Systems. London, U.K.: Springer-Verlag, 2014.
[15] M. Nghihaeian, N. Hirzallah, and P. G. Voulgaris, “Dual rate control for security in cyber-physical systems,” in Proc. 54th IEEE Conf. Decis. Control (CDC), Dec. 2015, pp. 1415–1420.
[16] H. Shim, J. Back, Y. Eun, G. Park, and J. Kim, “Zero-dynamics attack, variations, and countermeasures,” 2021, arXiv:2101.00556. [Online]. Available: http://arxiv.org/abs/2101.00556
[17] C. T. Chen, Linear System Theory and Design, 4th ed. London, U.K.: Oxford Univ. Press, 2013.
[18] H. K. Khalil, Nonlinear Systems, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[19] P. T. Kabamba, “Control of linear systems using generalized sampled-data hold functions,” IEEE Trans. Autom. Control, vol. AC-32, no. 9, pp. 772–783, Sep. 1987.
[20] C. T. Chen, Linear System Theory and Design. London, U.K.: Oxford Univ. Press, 1998.
[21] M. Grant and S. Boyd, CVX: MATLAB Software for Disciplined Convex Programming, Version 1.21. [Online]. Available: http://cvxr.com/cvx
[22] T. Morí, “Note on the absolute value of the roots of a polynomial,” IEEE Trans. Autom. Control, vol. AC-29, no. 1, pp. 54–55, Jan. 1984.
[23] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[24] L. H. Chiang, E. L. Russell, and R. D. Braatz, Fault Detection and Diagnosis in Industrial Systems. London, U.K.: Springer-Verlag, 2000.
[25] I. Izadi, S. L. Shah, D. S. Shook, and T. Chen, “An introduction to alarm analysis and design,” IFAC Proc. Volumes, vol. 42, no. 8, pp. 645–650, 2009.
[26] D. Kim, K. Ryu, and J. Back, “Security enhancement of sampled-data systems: Zero assignment via generalized sampler,” in Proc. 21st IFAC World Congr., 2020, pp. 1–6.

DAEHAN KIM received the B.S. and M.S. degrees from the School of Robotics, Kwangwoon University, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. His research interests include control system theory and design, power electronics, motor control, and renewable energy systems.

KUNHEE RYU (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in control and instruments engineering from Kwangwoon University, in 2012 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the School of Robotics. His research interests include robotics, distributed systems, and security.

JUNG HOON KIM (Member, IEEE) was born in Incheon, Republic of Korea, in September 1988. He received the B.E. degree in electrical and electronic engineering, the M.E. and Ph.D. degrees in electrical engineering from Kyoto University, Kyoto, Japan, in 2012, 2014, and 2015, respectively. He worked as a Young Scientist Research Fellow with the Japan Society for the Promotion of Science (JSPS), DC1, from 2014 to 2015, and a JSPS Postdoctoral Research Fellow with Kyoto University, in 2015, for two months. He was a Postdoctoral Researcher and a Senior Research Scientist with the Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology (KIST), from June 2015 to March 2019. Since April 2019, he has been with the Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), where he is currently an Assistant Professor. His current research interests include dynamical system theory and control applications, such as $L_1$ optimal control, sampled-data systems, time-delay systems, and robotics.

JUHOON BACK (Member, IEEE) received the B.S. and the M.S. degrees in mechanical design and production engineering from Seoul National University, in 1997 and 1999, respectively, Seoul, South Korea, and the Ph.D. degree from the School of Electrical Engineering and Computer Science, Seoul National University, Seoul, in 2004. From 2005 to 2006, he worked as a Research Associate with the Control and Power Group, Electrical and Electronic Engineering, Imperial College London, U.K. He is currently a Professor with the School of Robotics, Kwangwoon University, Seoul. His research interests include nonlinear control, estimation, multi-agent systems, and their applications renewable energy systems, vehicle control systems, and robotics. He is currently serving as an Associate Editor for the Systems and Control Letters and IEEE Control Systems Letters.