2015 American Control Conference
Palmer House Hilton
July 1-3, 2015. Chicago, IL, USA

Secure Networked Control Systems Against Replay Attacks Without Injecting Authentication Noise

Bixiang Tang, Luis D. Alvergue, and Guoxiang Gu

Abstract— **This paper studies detection of replay attacks on networked control systems, assuming that the actuation and sensing signals are transmitted over an additive white Gaussian noise channel. It explores the use of the spectral estimation technique to detect the presence of a replay attack without injecting authentication noise to the control signal at the plant input. Our proposed detection method is applicable to networked feedback systems equipped with stable controllers and designed with classical and modern control system techniques. A numerical example from the literature is used to illustrate the detection results for replay attacks on networked control systems.**

I. Introduction

Networked control systems (NCSs) have received great attention in the control system community. Such a feedback system employs information technology (IT) to connect the plant and controller, and allow communication through a shared network. Hence, this configuration has a wide range of applications, including mobile sensor networks [16], multi-agent systems [17], and automated highway systems [19], among others. For a more thorough overview of the state of the art see the special issues [1], [2], and the references therein. However, NCSs suffer from a greater vulnerability due to the presence of both cyber and physical attacks [15]. For this reason security issues in NCSs have attracted considerable interest [4], [6], [7], [22].

This paper addresses a particular attack, termed as replay attack studied first in [14], one year before the Stuxnet worm was exposed in the news media. Allegedly Stuxnet was initially used to counter Iran's nuclear program [5], [11], but it can be used by any malicious attacker who wishes to disrupt any system that relies on feedback. So long as attackers have remote access to sensing and actuation devices and they are able to modify the software or reprogram the devices, it is possible to launch coordinated attacks against the system infrastructure (that is full of feedback control systems) without being detected by the underlying NCS until it is too late.

Replay attacks assume that the sensing data are secretly recorded by the attacker, which are then replayed back to the monitor center while conducting the attack on the physical system. The deception created by replay is often seen in movies and spy fiction. A solution proposed in [14] injects a known independently identically distributed (i.i.d.) zero-mean Gaussian noise into the control signal at the plant input that serves as the authentication signal. Assuming an LQG control system, a χ^2 detector can then be used to detect the presence of the replay attack. It is shown in [14] that when the replay attack is present, the normalized error covariance of the innovation signal of the Kalman filter deviates from identity with a higher variance dependent on the variance of the injected noise. As its variance increases, the detection rate improves but the control performance suffers. There exists a trade-off between the detection rate and loss of the control performance in terms of the variance of the authentication signal. A method is proposed in [8] for designing the covariance of the authentication signal to minimize the performance loss while guaranteeing a certain probability of the detection rate. A different method is proposed in [13] by switching the feedback controller between the LQG (with no added noise) and the secure (with added noise) controllers. Results from non-cooperative stochastic games are used to minimize the worst-case control and detection cost. Another method injects i.i.d. Gaussian noise to the control signal on and off periodically [21], which can also provide similar trade-off between the control performance loss and the false detection rate. A natural question arises: Is it possible to detect the presence of the replay attack without injecting Gaussian noise? This problem becomes meaningful for NCSs due to the information distortion induced by the network channels, which may play a similar role to the injected noise at the control input. This paper considers a special type of NCSs involving additive white Gaussian noise (AWGN) channels. Such channels are well-studied in wireless communications [18], and in NCSs over fading channels [3], [9]. However, the χ^2-detector does not seem to work well due to the small bandwidth of the high loop gain. A spectral estimation approach is thus proposed to estimate the frequency response of the plant measurements at some specific frequency at which the plant gain is high while the controller gain is not small. It will be shown that the spectral estimator at this frequency provides a viable detector for replay attacks.

II. Preliminary Analysis

An NCS has the feedback controller situated in a different physical location from that of the plant, and it communicates with the plant via a (often wireless) network. The use of networks in feedback control systems thus creates a vulnerability for malicious attacks which seek to destabilize and damage the physical system. Specifically, consider the discrete-time feedback control system shown in Figure 1 in which $d(t)$ represents the disturbance input, $\eta(t)$ consists

The authors are with the School of Electrical Engineering and Computer Science, Louisiana State University, Baton Rouge, LA 70803-5901, USA lalver1@tigers.lsu.edu

This research is supported in part by NASA/LEQSF(2013-15)-Phase3-06 through grant NNX13AD29A and by the 111 project from NEU of China.

of both measurement noise, $\eta_0(t)$, and communication error $\eta_c(t)$, where t is integer-valued. Mathematically $\eta(t) = \eta_0(t) + \eta_c(t)$.

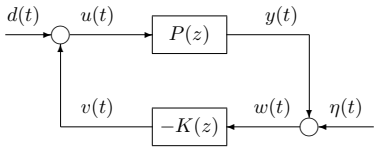

Fig. 1. Feedback control system

In Fig. 1, the reference or command input is assumed to be zero or removed together with its steady-state response in order to understand better the real issue in secure feedback control. The closed-loop transfer matrix from the exogenous inputs $(\{d(t)\}, \{\eta(t)\})$ to the controller output and input $(\{v(t)\}, \{w(t)\})$ signals is given by

$$T_K(z) = \begin{bmatrix} -K(z) \\ I \end{bmatrix} [I + P(z)K(z)]^{-1} \begin{bmatrix} P(z) & I \end{bmatrix}.$$

Without loss of generality the multi-input/multi-output (MIMO) plant model $P(z)$ with m-input/p-output is assumed to admit a stabilizable and detectable state-space realization. Its transfer matrix is described by

$$P(z) = D + C(zI - A)^{-1}B := \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \quad (1)$$

As a result, a stabilizing state feedback gain F and a stabilizing state estimation gain L exist such that $(A+BF)$ and $(A+LC)$ are both a Schur stability matrix. It is well known that $P(z)$ admits left/right coprime factorizations (LCF/RCF) [10]

$$P(z) = \tilde{M}(z)^{-1}\tilde{N}(z) = N(z)M(z)^{-1}$$

with $\{\tilde{M}(z), \tilde{N}(z), M(z), N(z)\}$ all stable transfer matrices. Assume that the feedback system in Fig. 1 is internally stable. Then the controller $K(z)$ admits LCF/RCF given by

$$K(z) = \tilde{V}(z)^{-1}\tilde{U}(z) = U(z)V(z)^{-1} \quad (2)$$

with $\{\tilde{V}(z), \tilde{U}(z), V(z), U(z)\}$ all stable transfer matrices satisfying the Bezout identity

$$\begin{bmatrix} \tilde{V}(z) & \tilde{U}(z) \\ -\tilde{N}(z) & \tilde{M}(z) \end{bmatrix} \begin{bmatrix} M(z) & -U(z) \\ N(z) & V(z) \end{bmatrix} = I_{m+p} \ \ \forall\, |z| \geq 1. \quad (3)$$

It is emphasized that the LCF/RCF for the plant model and for the stabilizing controller always exist and satisfy the Bezout identity in (3). The computation of such LCF/RCF for the plant and controller can be simplified for the observer-based controller

$$K_{\rm o}(z) = F(zI - A - BF - LC - LDF)^{-1}L.$$

Recall that $(A + BF)$ and $(A + LC)$ are both stability matrices. In this case, $K_{\rm o}(z)$ admits LCF/RCF

$$K_{\rm o}(z) = \tilde{V}_{\rm o}(z)^{-1}\tilde{U}_{\rm o}(z) = U_{\rm o}(z)V_{\rm o}(z)^{-1}$$

with realizations of its coprime factors together with those of coprime factors of $P(z)$ specified as

$$\begin{bmatrix} \tilde{V}_{\rm o}(z) & \tilde{U}_{\rm o}(z) \\ -\tilde{N}(z) & \tilde{M}(z) \end{bmatrix} = \left[\begin{array}{c|cc} A+LC & -(B+LD) & L \\ \hline F & I_m & 0 \\ \Omega^{-1}C & -\Omega^{-1}D & \Omega^{-1} \end{array}\right], \quad (4)$$

$$\begin{bmatrix} M(z) & -U_{\rm o}(z) \\ N(z) & V_{\rm o}(z) \end{bmatrix} = \left[\begin{array}{c|cc} A+BF & B & -L\Omega \\ \hline F & I_m & 0 \\ C+DF & D & \Omega \end{array}\right],$$

for each square and nonsingular Ω. The above is a slight modification from the existing literature [10]. Then any stabilizing controller $K(z)$ for the feedback system in Fig. 1 has the form

$$\begin{aligned} K(z) &= \left(\tilde{V}_{\rm o} + J\tilde{N}\right)^{-1}\left(\tilde{U}_{\rm o} - J\tilde{M}\right) \\ &= (U_{\rm o} - MJ)\,(V_{\rm o} + NJ)^{-1} \end{aligned} \quad (5)$$

for some stable $J(z)$. It follows that the coprime factors of $K(z)$ in (2) given by

$$V(z) = V_{\rm o}(z) + N(z)J(z),\ U(z) = U_{\rm o}(z) - M(z)J(z) \quad (6)$$

$$\tilde{V}(z) = \tilde{V}_{\rm o}(z) + J(z)\tilde{N}(z),\ \tilde{U}(z) = \tilde{U}_{\rm o}(z) - J(z)\tilde{M}(z), \quad (7)$$

satisfy the Bezout identity (3). Conversely, if the LCF/RCF of $K(z)$ shown above are available, then

$$J(z) = \tilde{U}_0(z)V(z) - \tilde{V}_0(z)U(z) = \tilde{V}(z)U_0(z) - \tilde{U}(z)V_0(z).$$

Normally the exogenous inputs $\{d(t)\}$ and $\{\eta(t)\}$ are wide-sense stationary (WSS) random processes, and are often white processes. However when attacks are present, $\{d(t)\}$ and $\{\eta(t)\}$ are replaced by $\{\alpha_u(t) + d(t)\}$ and $\{\alpha_y(t) + \eta(t)\}$, respectively, as shown in Figure 2.

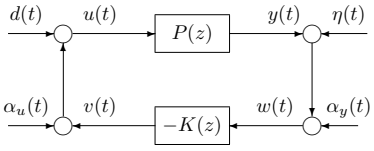

Fig. 2. Feedback control system under attack

It is assumed that the signals available for monitoring are at the controller site, and hence $\{v(t), w(t)\}$ can be logged by the controller, while the most valuable data $\{u(t), y(t)\}$ from the physical system are unavailable. In order to conceal the attack and induce damages, it is very likely that $\{\alpha_u(t)\}$ are unbounded, but $\{\alpha_y(t)\}$ are bounded. Even if a bounded $\alpha_u(t)$ is used for the malicious attack, it can be very irregular and disruptive in order to cause hardware damage to the physical system. As a result, $\{\alpha_u(t)\}$ and $\{\alpha_y(t)\}$ injected by the malicious attacker have different objectives: the former is aimed at replacing $u(t)$ so to damage the physical system while the latter is aimed at concealing the true output $y(t)$.

This paper will be focused on the replay attack studied initially in [14], [15]. The notorious Stuxnet worm is a prime example of such an attack. It fits into the framework in Figure 2 by taking $\alpha_u(t) = 0$ and

$$\begin{aligned} &\alpha_y(t) = (q^{-\tau_\alpha} - 1)[y(t) + \eta_{\rm o}(t)] \\ \Longrightarrow\quad &w(t) = y(t - \tau_\alpha) + \eta_{\rm o}(t - \tau_\alpha) + \eta_{\rm c}(t) \end{aligned} \tag{8}$$

for $t \geq t_\alpha$, where q^{-1} is the unit delay operator. Recall

$$\eta(t) = \eta_0(t) + \eta_c(t). \tag{9}$$

where both terms are i.i.d. random noise. From (8) it is clear that the replay attack has the effect of substituting the output (and measurement noise) at time t with the τ_α-samples delayed output. This way the attack conceals the real-time output of the plant and it is probably the easiest way to fake the output of a normally operating plant. With $\tau_\alpha >> 1$, the catastrophic result of the true $y(t)$ under attack $\alpha_u(t)$ is not observed at the controller site until a very long time later.

The replay attack results in

$$w(t) = P(q)[v(t - \tau_\alpha) + d(t - \tau_\alpha)] + \eta_{\rm o}(t - \tau_\alpha) + \eta_{\rm c}(t)$$

for $t \in [t_\alpha,\ t_\alpha + \tau_\alpha)$. Contrast the above to the case in absence of attacks:

$$w(t) = P(q)[v(t) + d(t)] + \eta_{\rm o}(t) + \eta_{\rm c}(t).$$

The replay attack is very effective so long as the controller $K(z)$ is stable and the feedback system is in steady-state, in light of the fact that $w(t)$ in absence of attacks is statistically no different from that in presence of attacks. A significant challenge posed by the replay attack lies in proposing a method to detect the attack without injecting noise and that is also applicable to commonly used control systems other than LQG. This problem will be studied in the next section.

III. Detection of Replay Attacks

Injection of authentication noise is effective to detect if a replay attack is present. However the noise injected at the plant input has to be large enough in order to achieve good detection performance, which deteriorates the control system performance. So there is a tradeoff between detection performance and control performance as demonstrated in [8], [13], [21]. In this paper the underlying NCS is assumed to employ network communications between the plant output and controller input over an AWGN channel, and thus $\eta(t)$ present at the plant output has the form of (9). It will be shown that the AWGN channel, while introducing information distortion, can help detection of the replay attack without injecting i.i.d. Gaussian noise at the plant input, provided that the noise power due to the AWGN channel is not too small. The following result provides two different LCFs of the plant model which will be useful later. The symbols $'$ and * stand for transpose and conjugate transpose, respectively.

Lemma 1 *Assume that the plant model $P(z)$ in (1) admits a stabilizable and detectable realization, and $\{d(t), \eta(t)\}$ are both temporal white processes with covariance Q_d and Q_η, respectively.*

(i) *Let $Y_{\rm n} \geq 0$ be the stabilizing solution to the discrete-time algebraic Riccati equation* (DARE)*:*

$$\begin{aligned} Y_{\rm n} = {} & AY_{\rm n}A' - (AY_{\rm n}C' + BQ_dD')Z_{\rm n}^{-1}(AY_{\rm n}C' + BQ_dD')' \\ & + BQ_dB' \end{aligned}$$

where $Z_{\rm n} = Q_\eta + DQ_dD' + CY_{\rm n}C'$. Then with $L = L_{\rm n} := -(AY_{\rm n}C' + BQ_dD')Z_{\rm n}^{-1}$, the left coprime factors of $P(z)$ in (4) are now given by $P(z) = \tilde{M}_{\rm n}(z)^{-1}\tilde{N}_{\rm n}(z)$ with realization

$$\left[\begin{array}{cc} \tilde{M}_{\rm n}(z) & \tilde{N}_{\rm n}(z) \end{array}\right] = \left[\begin{array}{c|cc} A + L_{\rm n}C & L_{\rm n} & (B + L_{\rm n}D) \\ \hline Z_{\rm n}^{-1/2}C & Z_{\rm n}^{-1/2} & Z_{\rm n}^{-1/2}D \end{array}\right]. \tag{10}$$

Moreover $\{\tilde{M}_{\rm n}(z), \tilde{N}_{\rm n}(z)\}$ satisfy the following normalization condition

$$\tilde{N}_{\rm n}(z)Q_d\tilde{N}_{\rm n}(z)^* + \tilde{M}_{\rm n}(z)Q_\eta\tilde{M}_{\rm n}(z)^* = I \quad \forall\ |z| = 1. \tag{11}$$

(ii) *Let $Y_0 \geq 0$ be the stabilizing solution to* DARE

$$Y_0 = AY_0A' - (AY_0C' + BD')Z_0^{-1}(CY_0A' + DB') + BB'$$

where $Z_0 = I + DD' + CY_0C'$. Then with $L = L_0 := -(AY_0C' + BD')Z_0^{-1}$, the left coprime factors of $P(z)$ in (4) are now given by $P(z) = \tilde{M}_0(z)^{-1}\tilde{N}_0(z)$ with realization

$$\left[\begin{array}{cc} \tilde{M}_0(z) & \tilde{N}_0(z) \end{array}\right] = \left[\begin{array}{c|cc} A + L_0C & L_0 & (B + L_0D) \\ \hline Z_0^{-1/2}C & Z_0^{-1/2} & Z_0^{-1/2}D \end{array}\right].$$

Moreover $\{\tilde{M}_0(z), \tilde{N}_0(z)\}$ satisfy the following normalization condition

$$\tilde{N}_0(z)\tilde{N}_0(z)^* + \tilde{M}_0\tilde{M}_0(z)^* = I \quad \forall\ |z| = 1. \tag{12}$$

Due to the space limit, the proof is omitted □

Consider first the LCF in Lemma 1 (i). It indicates that under (i),

$$\mathcal{N}(t) := \tilde{N}_{\rm n}(q)d(t) + \tilde{M}_{\rm n}(q)\eta(t) \tag{13}$$

is a white process with mean zero and covariance identity. That is, the power spectral density (PSD) of $\mathcal{N}(t)$ is identity at all frequencies. Let $K(z) = U_{\rm n}(z)V_{\rm n}(z)^{-1}$ be an RCF for some stable and proper $V_{\rm n}(z)$ and $U_{\rm n}(z)$ by taking $V(z) = V_{\rm n}(z)$ and $U(z) = U_{\rm n}(z)$ in (2). Note that $K(z)$ may not be an observer-based controller. In fact it can be PID or lead/lag compensator, provided that it stabilizes the feedback system in Figure 1 or 2. Now choosing $L = L_{\rm n}$ implies $\tilde{N}(z) = \tilde{N}_{\rm n}(z)$, and $\tilde{M}(z) = \tilde{M}_{\rm n}(z)$ for the LCF of $P(z)$ in (4). Since $K(z)$ is stabilizing, $V_{\rm n}(z)$ and $U_{\rm n}(z)$ can be chosen such that

$$V_{\rm n}(z) = V_{\rm o}(z) + N(z)J(z), \quad U_{\rm n}(z) = U_{\rm o}(z) - M(z)J(z),$$

for some stable $J(z)$ in light of (6) by simply setting $V(z) = V_{\rm n}(z)$ and $U(z) = U_{\rm n}(z)$. In addition there holds

$$\tilde{M}_{\rm n}(z)V_{\rm n}(z) + \tilde{N}_{\rm n}(z)U_{\rm n}(z) = I \ \ \forall\ |z| \geq 1. \tag{14}$$

The right coprime factorization of $P(z) = N(z)M(z)^{-1}$ can be obtained by taking some stabilizing state feedback gain

F, and thus $V_{\rm o}(z), U_{\rm o}(z)$, and $\tilde{V}_{\rm o}(z), \tilde{U}_{\rm o}(z)$ are also available by using $L = L_{\rm n}$ and the chosen stabilizing state feedback gain F.

Consider first the case of no attack. Using the coprime factorization description of the plant and controller, and (14), it can be shown that

$$T_K(z) = \left[\begin{array}{c} -U_{\rm n}(z) \\ V_{\rm n}(z) \end{array} \right] \left[\begin{array}{cc} \tilde{N}_{\rm n}(z) & \tilde{M}_{\rm n}(z) \end{array} \right]. \qquad (15)$$

Stability of $K(z)$ implies that $V_{\rm n}(z)^{-1}$ is also a stable and causal transfer matrix. It follows that $w(t) = V_{\rm n}(q)\mathcal{N}(t)$ by (15) and the definition of $\mathcal{N}(t)$ in (13). Thus $V_{\rm n}(z)^{-1}$ represents a whitening filter in the sense that the filtered signal

$$s(t) = V_{\rm n}(q)^{-1} w(t) = \mathcal{N}(t) \qquad (16)$$

is a white process with covariance identity, in light of the normalized left coprime factorization in (11) and the discussion after Lemma 1. Consequently the PSD of $s(t)$ is given by $\Phi_s(\omega) = I$ for all ω. Suppose that the replay attack takes place at $t = t_\alpha$ for the duration of $\tau_\alpha >> 1$. The PSD, $\Phi_w(\omega)$, for $w(t)$ is given in the following result.

Theorem 1 *Suppose that $\eta_{\rm o}(t)$ and $\eta_{\rm c}(t)$ are independent white processes for all t. Let $\{\tilde{M}_{\rm n}(z), \tilde{N}_{\rm n}(z)\}$ in* (10) *be* LCF *of $P(z)$ satisfying* (11)*, and $\{V_{\rm n}(z), U_{\rm n}(z)\}$ be* RCF *of the stabilizing controller satisfying* (14)*. Under the replay attack, the* PSD *of $w(t)$ is given by*

$$\begin{aligned} \Phi_w(\omega) = V_{\rm n}(e^{j\omega}) \left[I - \tilde{M}_{\rm n}(e^{j\omega}) Q_{\eta_{\rm c}} \tilde{M}_{\rm n}(e^{j\omega})^* \right] V_{\rm n}(e^{j\omega})^* \\ + N(e^{j\omega}) \tilde{U}(e^{j\omega}) Q_{\eta_{\rm c}} \tilde{U}(e^{j\omega})^* N(e^{j\omega})^* + Q_{\eta_{\rm c}} \end{aligned}$$

where $Q_{\eta_{\rm c}}$ is the covariance of $\eta_{\rm c}(t)$. In this case the PSD *of $s(t)$ in* (16) *is given by*

$$\begin{aligned} \Phi_s(\omega) = I - \tilde{M}_{\rm n}(e^{j\omega}) Q_{\eta_{\rm c}} \tilde{M}_{\rm n}(e^{j\omega})^* \\ + V_{\rm n}(e^{j\omega})^{-1} \left[N(e^{j\omega}) \tilde{U}(e^{j\omega}) Q_{\eta_{\rm c}} \tilde{U}(e^{j\omega})^* N(e^{j\omega})^* \right. \\ \left. + Q_{\eta_{\rm c}} \right] V_{\rm n}(e^{j\omega})^{*-1}. \end{aligned} \qquad (17)$$

Proof: Suppose that $t > t_\alpha$. Denote

$$\mathcal{N}_\alpha(t) = \mathcal{N}(t - \tau_\alpha) = \tilde{N}_{\rm n}(q) d(t - \tau_\alpha) + \tilde{M}_{\rm n}(q) \eta(t - \tau_\alpha).$$

Over the time interval of $[t_\alpha,\ t_\alpha + \tau_\alpha)$,

$$\begin{aligned} w(t) = w_\alpha(t) &= y(t - \tau_\alpha) + \eta_{\rm o}(t - \tau_\alpha) + \eta_{\rm c}(t) \\ &= w(t - \tau_\alpha) - \eta_{\rm c}(t - \tau_\alpha) + \eta_{\rm c}(t). \end{aligned}$$

By the relation $w(t) = V_{\rm n}(q)\mathcal{N}(t)$, there holds

$$\begin{aligned} w_\alpha(t) &= V_{\rm n}(q) \mathcal{N}_\alpha(t) - \eta_{\rm c}(t - \tau_\alpha) + \eta_{\rm c}(t) \\ &= V_{\rm n}(q) \left[\tilde{N}_{\rm n}(q) d(t - \tau_\alpha) + \tilde{M}_{\rm n}(q) \eta_{\rm o}(t - \tau_\alpha) \right] \\ &\quad - \left[I - V_{\rm n}(q) \tilde{M}_{\rm n}(q) \right] \eta_{\rm c}(t - \tau_\alpha) + \eta_{\rm c}(t). \end{aligned}$$

The Bezout identity in (3) can now be written as

$$\left[\begin{array}{cc} M(z) & -U_{\rm n}(z) \\ N(z) & V_{\rm n}(z) \end{array} \right] \left[\begin{array}{cc} \tilde{V}(z) & \tilde{U}(z) \\ -\tilde{N}_{\rm n}(z) & \tilde{M}_{\rm n}(z) \end{array} \right] = I_{m+p} \quad \forall\ |z| \geq 1. \qquad (18)$$

The above implies $I - V_{\rm n}(z)\tilde{M}_{\rm n}(z) = N(z)\tilde{U}(z)$, leading to

$$\begin{aligned} w_\alpha(t) = V_{\rm n}(q) \left[\tilde{N}_{\rm n}(q) d(t - \tau_\alpha) + \tilde{M}_{\rm n}(q) \eta_{\rm o}(t - \tau_\alpha) \right] \\ - N(q) \tilde{U}(q) \eta_{\rm c}(t - \tau_\alpha) + \eta_{\rm c}(t). \end{aligned} \qquad (19)$$

The expression of $w_\alpha(t)$ is different from $w(t)$ prior to t_α. As a result $s(t) = V_{\rm n}(q)^{-1} w_\alpha(t)$ is not a white process in general. Since the four terms in (19) are all uncorrelated, the PSD of $w(t)$ over $[t_\alpha,\ t_\alpha + \tau_\alpha)$ can be easily obtained as in the proposition in which the normalization property (11) is used in obtaining the PSD expression. The PSD of $s(t)$ in (17) follows by $s(t) = V_{\rm n}(q)^{-1} w(t)$. □

The above result shows that if $\|\Phi_s(\omega) - I\|$ is significantly greater than zero in most of the frequency range, then successful detection of the presence of replay attacks has a high probability. As a result, detection of the replay attack is equivalent to detecting whether or not $s(t)$ is white. However the hypothesis on large deviation for $\|\Phi_s(\omega) - I\|$ in most of the frequency range does not hold in engineering practice. For this reason, it is more meaningful to consider the PSD of $w(t)$ or $s(t)$ at some specific frequency for which the LCF in (ii) of Lemma 1 plays an important role. It is also important to point out that the LCF in (ii) of Lemma 1 does not make use of the covariance of the input disturbance or/and output noise, and is thus more advantageous and admits robustness against the inaccuracy of the covariance matrices Q_d and Q_η.

Now consider the LCF in (ii) of Lemma 1, and RCF $P(z) = N(z)M(z)^{-1}$ based on some stabilizing state feedback gain F. Let $K(z) = U(z)V(z)^{-1}$ be a given stabilizing controller. Then in light of the parameterization of the stabilizing controllers in (5) and (6),

$$V(z) = V_{\rm o}(z) + N(z)J(z), \quad U(z) = U_{\rm o}(z) - M(z)J(z)$$

for some stable and proper $J(z)$ where $K_{\rm o}(z) = U_{\rm o}(z)V_{\rm o}(z)^{-1}$ is an observer-based controller using the state feedback gain F and state estimation gain L_0. It follows from $P(z) = \tilde{M}_0(z)^{-1}\tilde{N}_0(z)$ that $T_K(z)$ in (15) can be written as

$$T_K(z) = \left[\begin{array}{c} -U(z) \\ V(z) \end{array} \right] \left[\begin{array}{cc} \tilde{N}_0(z) & \tilde{M}_0(z) \end{array} \right].$$

The next result provides the expression of the PSD of $w(t)$ at high-gain frequency ω_h with $P(e^{j\omega}) \to \infty$ as $\omega \to \omega_h$.

Theorem 2 *Suppose that the input noise $d(t)$ is a white process with covariance $Q_d = \sigma_d^2 I_m$, and the* AWGN *noise have covariance $Q_{\eta_{\rm c}} = \sigma_{\eta_{\rm c}}^2 I_p$. Assume that $p \leq m$, and $\sigma_i[P(e^{j\omega})]$ has infinity gain at $\omega = \omega_h$ for $1 \leq i \leq p$ where $\sigma_i(\cdot)$ is the ith singular value arranged in descending order. Let $\Phi_w(\omega)$, and $\Phi_s(\omega)$ be the* PSD *for $w(t)$, and $s(t)$, respectively. In the absence of replay attacks,*

$$\Phi_w(\omega_h) = \sigma_d^2 V(e^{j\omega_h}) V(e^{j\omega_h})^*, \quad \Phi_s(\omega_h) = \sigma_d^2 I_m. \qquad (20)$$

In the presence of replay attacks,

$$\begin{aligned} \Phi_w(\omega_h) &= \sigma_d^2 V(e^{j\omega_h}) V(e^{j\omega_h})^* + 2\sigma_{\eta_{\rm c}}^2 I_m, \\ \Phi_s(\omega_h) &= \sigma_d^2 I_m + 2\sigma_{\eta_{\rm c}}^2 V(e^{j\omega_h})^{-1} V(e^{j\omega_h})^{*-1}. \end{aligned} \qquad (21)$$

Proof: Under the assumption that all singular values of $P(e^{j\omega})$ tend to infinity as $\omega \to \omega_h$,

$$\tilde{M}_0(e^{j\omega})\tilde{M}_0(e^{j\omega})^* \;\to\; 0, \quad \tilde{N}_0(e^{j\omega})\tilde{N}_0(e^{j\omega})^* \;\to\; I_p, \tag{22}$$

as $\omega \to \omega_h$ in light of (12). It follows that $w(t)$ now has an expression

$$w(t) = V(q)\left[\tilde{N}_0(q)d(t) + \tilde{M}_0(q)\eta(t)\right]$$

in absence of attacks. Thus there holds $V(e^{j\omega})\tilde{M}_0(e^{j\omega}) \to 0$ as $\omega \to \omega_h$. We can now conclude that the PSD of $w(t)$ at $\omega = \omega_h$ is given by

$$\begin{aligned}\Phi_w(\omega_h) &= V(e^{j\omega_h})\tilde{N}_0(e^{j\omega_h})Q_d\tilde{N}_0(e^{j\omega_h})^*V(e^{j\omega_h})^* \\ &= \sigma_d^2 V(e^{j\omega_h})V(e^{j\omega_h})^*, \end{aligned}\tag{23}$$

if $Q_d = \sigma_d^2 I$ that verifies (20). In addition there holds

$$\begin{bmatrix} M(z) & -U(z) \\ N(z) & V(z) \end{bmatrix}\begin{bmatrix} \tilde{V}(z) & \tilde{U}(z) \\ -\tilde{N}_0(z) & \tilde{M}_0(z) \end{bmatrix} = I_{m+p} \quad \forall\ |z| \ge 1.$$

The above is the same as (18) except that $\tilde{M}_{\rm n}(e^{j\omega})$ and $\tilde{N}_{\rm n}(e^{j\omega})$ are replaced by $\tilde{M}_0(e^{j\omega})$ and $\tilde{N}_0(e^{j\omega})$, respectively, and $V_{\rm n}(z), U_{\rm n}(z)$, are replaced by $V(z)$ and $U(z)$, respectively. As a result,

$$\begin{aligned} & I - V(z)\tilde{M}_0(z) = N(z)\tilde{U}(z) \ \ \forall\ |z| \ge 1 \\ \Longrightarrow\ & N(e^{j\omega_h})\tilde{U}(e^{j\omega_h}) = I. \end{aligned}$$

If the replay attack is present, then $\Phi_w(\omega_h)$ changes its value. Specifically (19) is modified into

$$\begin{aligned} w_\alpha(t) = V(q)\left[\tilde{N}_0(q)d(t-\tau_\alpha) + \tilde{M}_0(q)\eta_{\rm o}(t-\tau_\alpha)\right] \\ -N(q)\tilde{U}(q)\eta_{\rm c}(t-\tau_\alpha) + \eta_{\rm c}(t). \end{aligned}$$

The results in (22) and (23) then lead to

$$\Phi_w(\omega_h) = \sigma_d^2 V(e^{j\omega_h})V(e^{j\omega_h})^* + 2Q_{\eta_{\rm c}}$$

that verifies the expression of $\Phi_w(\omega_h)$ in (21), if in addition $Q_{\eta_c} = \sigma_{\eta_{\rm c}}^2 I$. The derivation for $\Phi_s(\omega_h)$ is similar which is omitted. □

Remark 1 Consider the single-input/single-output (SISO) case. The normalized coprime factors in (ii) of Lemma 1 implies that $|\tilde{N}_0(e^{j\omega_h})| = 1$ by $\tilde{M}_0(e^{j\omega_h}) = 0$ due to the assumption of $|P(e^{j\omega_h})| = \infty$. In addition $M(e^{j\omega_h}) = 0$. Thus $U(e^{j\omega_h})\tilde{N}_0(e^{j\omega_h}) = 1$. It follows that $|U(e^{j\omega_h})| = 1$ as well, and hence $|V(e^{j\omega_h})|$ has a small value, provided that $K(z) = U(z)V(z)^{-1}$ has a reasonably large gain at ω_h, which is ensured by several different design techniques, including the classical Bode design method and $\mathcal{H}_\infty$ loop shaping [12]. Often $\omega_h = 0$ and in this case if $|V(e^{j\omega_h})|$ is not very small, a lag type compensator can be added in by replacing $K(z)$ by

$$K(z) + \frac{\delta}{z - 1 + \epsilon}$$

with both $\delta > 0$ and $\epsilon > 0$ much smaller than 1 and $\frac{\delta}{\epsilon} >> 1$. Hence the gain of $V(z)^{-1}$ can be boosted to help increase the detection rate without sacrificing the false alarm rate. The addition of the lag type compensator $\frac{\delta}{z-1+\epsilon}$ helps to improve the steady-state performance while compromising little of the transient response. See our numerical example in Section IV. □

Theorem 2 and Remark 1 show that

$$\Phi_s(\omega_h) = \begin{cases} \sigma_d^2, & \text{if } \mathcal{H}_0, \\ \sigma_d^2 + 2\sigma_{\rm c}^2|V(e^{j\omega_h})|^{-2}, & \text{if } \mathcal{H}_1, \end{cases}$$

where $\mathcal{H}_0$ represents the hypothesis for the absence, and $\mathcal{H}_1$ for the presence of the replay attack. The two terms on the right hand sides can have large difference from each other. Therefore a threshold $\pi(\alpha\%)$, based on the given false alarm rate $\alpha\%$, can be setup for the following detector:

$$\Phi_s(\omega_h) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \pi(\alpha\%). \tag{24}$$

IV. An Illustrative Example

An example from [14] is employed to illustrate our proposed detection technique. This example considers a room temperature control system. Let $T(t)$ be the room temperature at time index t and T_* be the desired temperature. Denote $x(t) = T(t) - T_*$ as the state variable. Then the following describes the dynamic process of the temperature deviation:

$$x(t+1) = x(t) + u(t) + d(t), \quad y(t) = x(t) + \eta(t).$$

It is assumed in [14] that the disturbance $d(t)$ and measurement noise $\eta_{\rm o}(t)$ are i.i.d. Gaussian processes with variance 1 and 0.1 respectively. This paper assumes that $\eta_{\rm c}(t)$, the AWGN of the communication channel at the plant output, is also i.i.d. Gaussian. Clearly the plant model satisfies the hypothesis of Theorem 2 by taking $\omega_h = 0$ at which the plant has an infinity gain. In contrast to the LQG controller in [14], a simple deadbeat controller $K(z) \equiv 1$ is taken that places the closed-loop pole to the origin. Since this controller does not admit small gain for $|V(e^{j\omega_h})|$ at $\omega_h = 0$, the deadbeat controller is replaced by

$$K(z) = 1 + \frac{0.09}{z - 0.99} = \frac{z - 0.9}{z - 0.99},$$

that is a lag compensator, following Remark 1. Hence the value of $|V(e^{j\omega_h})|^{-1}$ is boosted 10 times. We employ the detector in (24) with false alarm rate $\alpha = 5\%$. The detection performance is shown in Figure 3 based on an average of 10,000 ensembles with the window size 5 for spectrum estimation. The solid (blue), dotted (black), and dashed line (red) curves show the detection rates when $\sigma_{\eta_{\rm c}}^2 = 0.1, 0.2, 0.6$, respectively. The performance curves are much better than those of [14] in which the detection rate is only 0.35 when the noise injected to the control signal at the plant input has variance 0.6. Intuitively speaking, the authentication noise injected at the plant input is not as effective as the communication noise present at the plant output, considering that the noise at the plant input has to

propagate through the plant dynamics before showing up at the plant output.

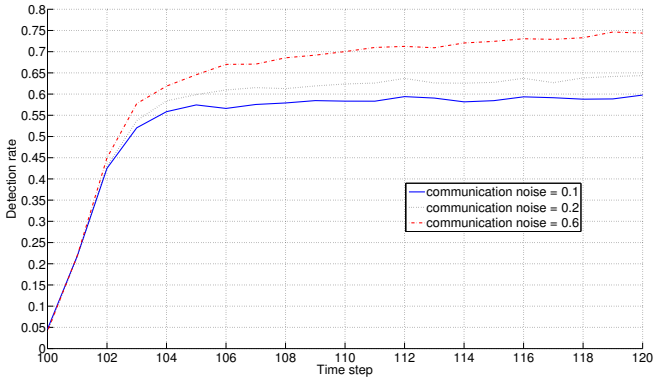

Fig. 3. Detection rate at each time step for different variances

Figure 4 shows the detection performance for different window sizes when the communication noise has $\sigma^2_{\eta_c} = 0.1$. The solid (red), dashed (blue), and dotted (black) lines correspond to window size of 5, 10, and 20, respectively. It is seen that as window size increases, the detection performance improves. This observation is important, because the variance of the communication noise cannot be chosen by the designer. So if the variance is too small, the window size can be increased in order to improve the detection performance. However as the window size increases, the timely detection of the replay attack can be a problem. It is clear then, that there is a tradeoff between the window size and timely detection. More details on the computational aspects involved in computing $V(z)$ in Theorem 2 and on spectrum estimation can be found in [20].

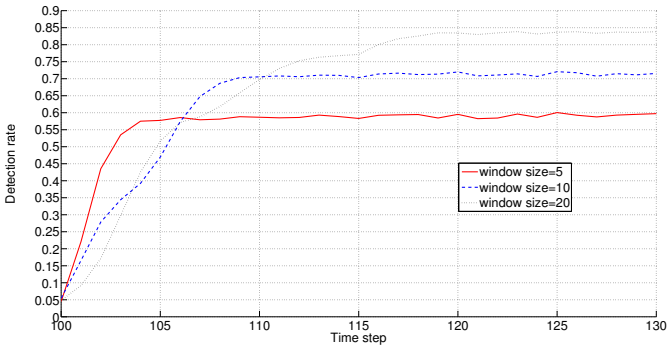

Fig. 4. Detection rate at each time step for different window sizes

V. Conclusion

This paper considers the problem of the replay attack on NCSs. While this problem has been studied in a number of papers, authentication noise has to be injected to the control signal which degrades the control system performance. In contrast, we propose to utilize the communication noise that already exists in NCSs for detecting the replay attack. Although the communication noise is unknown at each time sample, it can also serve as a time stamp, in a similar manner as the authentication noise. A spectral estimation method is developed to estimate the spectrum of the received signal at the controller site at a specific frequency. Its value or its filtered value differ between the presence and absence of the replay attack. A numerical example has been worked out to illustrate our proposed detection algorithm.

References

[1] P. Antsaklis and J. Baillieul (Guest Editors), "Special Issue on Networked Control Systems", *IEEE Trans. Automat. Contr.*, vol. 49, 2004.

[2] P. Antsaklis and J. Baillieul (Guest Editors), "Special Issue on Technology of Networked Control Systems", *Proc. of IEEE*, vol. 95, 2007.

[3] J.H. Braslavsky, R.H. Middleton, and J.S. Freudenberg, "Feedback stabilization over signal-to-noise ratio constrained channels," *IEEE Trans. Automat. Contr.*, vol. 52, pp. 1391-1403, Aug. 2007.

[4] E. Byres and J. Lowe, "The myths and facts behind cyber security risks for industrial control systems," *Proceedings of the VDE Kongress*, VDE, Berlin, Oct. 2004.

[5] W.J. Broad, J. Markoff, and D.E. Sanger, "Israeli test on worm called crucial in iran nuclear delay," *The New York Times*, January 2011.

[6] A.A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," *Proceedings of the 3rd Conference on Hot Topics in Security*, pp. 1-6, Berkeley, CA, USA: USENIX Association, 2008.

[7] A.A. Cárdenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Proceedings of 28th International Conference on Distributed Computing Systems Workshops*, ICDCS'08, pp. 495-500, 2008.

[8] R. Chabukswar, Y. Mo, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," in *Proceedings of the 18th IFAC World Congress*, pp. 11239-11244, Milano, Italy, 2011.

[9] J. Freudenberg, R. Middleton and J. Braslavsky, "Minimum variance control over a Gaussian communication channel", *IEEE Trans. Automat. Contr.*, vol. 56, pp. 1751-1765, 2011.

[10] G. Gu, *Discrete-Time Linear Systems: Theory and Design with Applications*, Springer, March 2012.

[11] David Kushner, "The real story of stuxnet," *IEEE Spectrum*, vol. 50, pp. 48-53, March 2013.

[12] D. McFarlane and K. Glover, *Robust Controller Design Procedure Using Normalized Coprime Factor Plant Descriptions*, vol. 138, Lecture Notes in Control and Information Sceinces, Springer-Verlag, 1990.

[13] F. Miao, M. Pajic, and G.J. Pappas, "Stochastic game approach for replay attack detection," *Proceedings of the 52nd IEEE Conference on Decision and Control*, pp. 1854?859, Firenze, Italy, Dec. 2013.

[14] Y. Mo and B. Sinopoli, "Secure control against replay attacks," *Proceedings of 47th Annual Allerton Conference* (UIUC, IL), pp. 911-918, Sept 30 - Oct. 2, 2009.

[15] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no.1, pp. 195-209, Jan. 2012.

[16] P. Ogren, E. Fiorelli, and N. E. Leonard, "Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment", *IEEE Trans. Automat. Contr.*, vol. 49, pp. 1292-1302, 2004.

[17] R. Olfati-Saber, and R. Murray , "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp. 1520-1533, 2004.

[18] T.S. Rappaport, *Wireless Communications: Principles and Practice*, second edition, Prentice Hall, 2002.

[19] P. Seiler and R. Sengupta, "Analysis of communication losses in vehicle control problems", *Proc. 2001 Amer. Contr. Conf.*, vol. 2, pp. 1491-1496, 2001.

[20] B. Tang, *New approaches to smart grid security with SCADA systems*, Ph.D. Dissertation, ECE, Louisiana State University, July 2014.

[21] T.-T. Tran, O.-S. Shin, and J.-H. Lee, "Detection of replay attacks in smart grid systems," *Proc. Int. Conf. Comput. Mangement and Telecommun.*, Ho Chi Minh City, Vietnam, pp. 298-302, 2013.

[22] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," *Proc. IEEE Int. Conf. Smart Grid Commun*, pp. 226-231, Oct. 2010.