2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Dynamic Watermarking for Cyber-Security of Nonlinear Stochastic Systems

Tzu-Hsiang Lin and P. R. Kumar

Abstract— **As cyber-physical systems form the core of many critical infrastructures, ensuring their safety is essential. The sensor measurements of networked cyber-physical systems can potentially be compromised, resulting in the misbehavior of the overall system. Indeed there have been a number of such attacks. Dynamic Watermarking is a proactive method whose goal is to detect such cyber attacks. It superimposes a small secret stochastic excitation onto signals in the system, such as control inputs of actuators or sensor measurements. Based on an examination of the signals purportedly returned, for example, by the sensors, it determines if the measurements have been tampered with. Previous theory for detection guarantees provided by Dynamic Watermarking has been restricted to linear stochastic systems.**

This paper examines the Dynamic Watermarking method for nonlinear stochastic systems. We show that Dynamic Watermarking for detecting attacks can be extended to certain systems in backstepping form. We present the analytical proofs, as well as simulation results.

I. Introduction

Cyber-Physical Systems (CPS) integrate sensing, computation, control, and networking with physical components. They have become an essential part of critical infrastructures such as transportation, energy, and industrial control systems. However, being networked, the systems are susceptible to cyber attacks on network nodes or links. Sensors in particular are vulnerable nodes in the system. Both sensor measurements and information flows from the sensors can be compromised, leading to malfunctioning of the system. There are several kinds of attacks, such as false data injection [1], replay attack [2], etc. Examples can be found in [3]–[5]. Being at the core of critical infrastructures also means that any malfunction of the CPS can result in great damage. Therefore, the safety of CPS becomes crucial.

The watermarking method [6]–[8] is one of the methods used for cyber security of CPS. It is a method that injects a private signal into the system. In this work, the private signal is superimposed onto a nominal control input. The private signal is typically a random process whose statistics can be disclosed to others, but whose waveform is kept secret. If the system is not under attack, the private signal should appear in the returned sensor measurement, appropriately transformed according to the system's parameters. Dynamic Watermarking [8] should thereby be able to detect attacks. However, the mathematical proofs have been restricted to linear stochastic systems.

Nonlinear stochastic systems can be linearized if they operate around a fixed setpoint. In that case, the Dynamic Watermarking method for linear stochastic systems may be sufficient. However, if the system has a time-varying setpoint, or it has a large excursion, then it is necessary to design a Dynamic Watermarking method specifically for nonlinear stochastic systems.

This paper addresses the challenge of detecting cyber attacks in *nonlinear* stochastic control systems. We describe the watermarking method for discrete-time stochastic systems in backstepping form. We prove that it can be used to detect any nonzero power distortions of sensor measurements. We also present simulation results on false data injection attacks and replay attacks.

II. Related Work

Aside from the watermarking method, other methods have been proposed to detect cyber attacks on cyber-physical systems. In [9], a study of delay insertion attacks on the feedback path has been conducted. A method using recursive prediction error has been proposed to detect such attacks. Additionally, in [10], a reachable set-based detection method targeting false data injection attacks has been proposed. A more comprehensive survey of the literature is provided in [11].

Regarding the watermarking method, various aspects need additional consideration. For example, since the watermarking method injects a private signal into the system, the magnitude of the signal can impact both normal system behavior and attack detection performance. If the signal size is significant, it may cause significant fluctuations in output measurements, thereby affecting system performance. Conversely, if the signal is too small, it may not be detectable in the output measurement, or there could be a large detection delay. In [12], a method is proposed to minimize the impact of the watermark on the system while maximizing attack detection performance.

In [13], Dynamic Watermarking has been applied to a particular nonlinear system - a "bicycle model" for vehicles. The authors demonstrate that as long as the variance of the noise and watermark remains bounded, dynamic watermarking is effective. In [14], a Dynamic Watermarking method that can be used for linear time-varying systems is proposed.

This material is based upon work partially supported by the National Science Foundation under Contract Numbers CNS-2328395 and CMMI-2038625, US Army Contracting Command under W911NF-22-1-0151, US ARO under W911NF-21-2-0064, US Office of Naval Research under N00014-21-1-2385 and N00014-24-1-2615, and U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number DE-EE0009031.

The authors are with Texas A&M University, Department of Electrical and Computer Engineering, College Station, TX 77843 USA {thl1246, prk}@tamu.edu.

The majority of the works mentioned above focus on linear systems. Applying these methods to nonlinear systems with time-varying setpoints is challenging. This work studies a general method that can be applied to nonlinear stochastic systems.

III. Dynamic Watermarking

In this section, a mathematical proof of how Dynamic Watermarking can be applied to two important classes of nonlinear stochastic dynamical systems is presented. Consider a multiple-input–multiple-output (MIMO) nonlinear system with additive Gaussian noise, described by

$$\mathbf{x}[t+1] = f(\mathbf{x}[t]) + B\mathbf{u}[t] + \mathbf{w}[t+1], \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^m$, $B \in \mathbb{R}^{n\times m}$and $\{\mathbf{w}\}$ is i.i.d. Gaussian distributed random vector $\sim N(0, \sigma_w^2 I)$. The case where $f(\mathbf{x}[t])$ is a linear function has been covered in [8]. Let the history dependent control policy intending to be applied on the ith input be $g^i = (g_1^i, g_2^i, ..., g_t^i, ...)$, with the input at time $u_i[t] = g_t^i(\mathbf{x}^t)$ where $\mathbf{x}^t \coloneqq (\mathbf{x}[0], \mathbf{x}[1], ..., \mathbf{x}[t])$. However, the actuator does not have access to $\mathbf{x}^t$; it only has access to the *reported* output measurements, denoted as $\mathbf{z}^t$. If $\mathbf{x}[t] \equiv \mathbf{z}[t]$, i.e., $\mathbf{x}[t] = \mathbf{z}[t]$ for all t, the reported output measurements are honest, but we interested in the case where $\mathbf{z}^t \not\equiv \mathbf{x}^t$, i.e., a malicious agent is reporting false measurements.

To implement Dynamic Watermarking, a random signal $e_i[t]$ called "watermark" is superimposed onto the ith input $g_t^i(z^t)$ of the actuator. The watermark is a white Gaussian sequence with variance σ_e^2, i.e., i.i.d. $\sim N(0, \sigma_e^2)$. Also $e_i[t]$ is independent of $e_j[k]$ for $(i,t) \neq (j,k)$, $\mathbf{x}$[m], $\mathbf{z}$[m] for $m \leq t$, and for all $\mathbf{w}$. The values of $e_i[t]$ are only known to the actuator. By adding $e_i[t]$, the input of the system becomes

$$u_i[t] = g_t^i(\mathbf{z}^t) + e_i[t]. \tag{2}$$

So the non-linear system with watermark injected is

$$\mathbf{x}[t+1] = f(\mathbf{x}[t]) + Bg_t(\mathbf{z}^t) + B\mathbf{e}[t] + \mathbf{w}[t+1]. \tag{3}$$

The actuator does not directly measure the state $\mathbf{x}[t]$. Rather it receives a measurement $\mathbf{z}[t]$ purported to be that of the state. Using (3) as a touchstone, the honest actuator subjects the reported measurements $\mathbf{z}[t]$ to the following two tests:

Test 1: Is

$$\begin{aligned}&\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\mathbf{z}[k+1] - f(\mathbf{z}[k]) - Bg_k(\mathbf{z}^k))\\ &(\mathbf{z}[k+1] - f(\mathbf{z}[k]) - Bg_k(\mathbf{z}^k))^T \stackrel{?}{=} \sigma_e^2 BB^T + \sigma_w^2 I.\end{aligned} \tag{4}$$

Test 2: Is

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1} e_i[k](\mathbf{z}[k+1] - f(\mathbf{z}[k]) - Bg_k(\mathbf{z}^k))\\ \stackrel{?}{=} B_{\cdot,i}\sigma_e^2.\end{aligned} \tag{5}$$

The above asymptotic tests can be converted in standard ways to finite-time statistical χ^2-tests.

Define

$$\mathbf{v}[t+1] \coloneqq \mathbf{z}[t+1] - f(\mathbf{z}[t]) - Bg_t(\mathbf{z}^t) - B\mathbf{e}[t] - \mathbf{w}[t+1].$$

If the reported measurements are honest, which means $z[t] \equiv x[t]$, then $v[t] \equiv 0$. The following proof of Theorem 1 is a straight forward extension of the proof in [8] which replaces $A\mathbf{z}[t]$ with $f(\mathbf{z}[t])$.

Theorem 1: If $\{\mathbf{z}[t]\}$ passes tests (4) and (5), then

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}||\mathbf{v}[k]||^2 = 0. \tag{6}$$

Proof: Since $\{\mathbf{z}[t]\}$ satisfies (5). $\forall i \in \{1, 2, ..., m\}$, it can be written as

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1} e_i[k](\mathbf{v}[k+1] + B\mathbf{e}[k] + \mathbf{w}[k+1])\\ = B_{\cdot,i}\sigma_e^2.\end{aligned}$$

It follows that for all $i \in \{1, 2, ..., m\}$

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1} e_i[k]\mathbf{v}[k+1] = 0.$$

Therefore,

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbf{e}[k]\mathbf{v}^T[k+1] = 0. \tag{7}$$

Since $\{\mathbf{z}[t]\}$ also satisfies (4), it can be written as

$$\begin{aligned}&\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\mathbf{v}[k+1] + B\mathbf{e}[k] + \mathbf{w}[k+1])\\ &(\mathbf{v}[k+1] + B\mathbf{e}[k] + \mathbf{w}[k+1])^T = \sigma_e^2 BB^T + \sigma_w^2 I_n.\end{aligned} \tag{8}$$

Using (7), and the fact that $\mathbf{w}[k+1]$ is independent of $\mathbf{e}[k]$, the above can be written as

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\mathbf{v}[k+1]\mathbf{w}[k+1]) + (\mathbf{v}[k+1]\mathbf{w}[k+1])^T\\ + (\mathbf{v}[k+1]\mathbf{v}^T[k+1]) = 0.\end{aligned} \tag{9}$$

Let S_{k+1} be the $\sigma-$algebra generated by $(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{e}^{k-1})$. Since

$$B\mathbf{e}[k] + \mathbf{w}[k+1] = \mathbf{x}[k+1] - f(\mathbf{x}[k]) - Bg_k(\mathbf{z}^k)), \tag{10}$$

it is S_{k+1} measurable. Therefore, $\hat{\mathbf{w}}[k+1] \coloneqq E[\mathbf{w}[k+1]|S_{k+1}] = E[\mathbf{w}[k+1]|B\mathbf{e}[k] + \mathbf{w}[k+1]]$.

The conditional mean estimate given the Gaussian distributions of $\mathbf{w}$ and $\mathbf{e}$ is

$$\begin{aligned}\hat{\mathbf{w}}[k+1] &= \sigma_w^2(\sigma_e^2 BB^T + \sigma_w^2 I)^{-1}(B\mathbf{e}[k] + \mathbf{w}[k+1])\\ &= K_w(B\mathbf{e}[k] + \mathbf{w}[k+1]).\end{aligned} \tag{11}$$

Since $\mathbf{e}[k]$ is S_{k+2} measurable, it follows that $\mathbf{w}[k+1]$ and $\mathbf{v}[k+1]$ are also S_{k+2} measurable. Let $\tilde{\mathbf{w}}[k] \coloneqq \mathbf{w}[k] - \hat{\mathbf{w}}[k]$. Then $(\tilde{\mathbf{w}}[k+1], S_{k+2})$ is a Martingale difference sequence. From the Martingale Stability Theorem [15], we have

$$\sum_{k=0}^{T-1} \tilde{\mathbf{w}}[k+1]\mathbf{v}^T[k+1] = \begin{bmatrix} o(\sum_{k=0}^{T-1} v_1^2[k+1]) & \dots & o(\sum_{k=0}^{T-1} v_p^2[k+1]) \\ o(\sum_{k=0}^{T-1} v_1^2[k+1]) & \dots & o(\sum_{k=0}^{T-1} v_p^2[k+1]) \\ \vdots & \dots & \vdots \\ o(\sum_{k=0}^{T-1} v_1^2[k+1]) & \dots & o(\sum_{k=0}^{T-1} v_p^2[k+1]) \end{bmatrix} + O(1) \tag{12}$$

where $v_i[k+1]$ denotes the ith element of $\mathbf{v}[k+1]$, and o and O denote little-o and big O order notation, respectively. Using the above, we have

$$\begin{aligned} \mathbf{w}[k+1] &= \hat{\mathbf{w}}[k+1] + \tilde{\mathbf{w}}[k+1] \\ &= K_w(B\mathbf{e}[k] + \mathbf{w}[k+1]) + \tilde{\mathbf{w}}[k+1]. \end{aligned} \tag{13}$$

Based on the assumption that the rank of B holds, it follows that K_w has all eigenvalues strictly less than unity. Therefore, we have:

$$\mathbf{w}[k+1] = (I - K_w)^{-1}K_wB\mathbf{e}[k+1] + (I - K_w)^{-1}\tilde{\mathbf{w}}[k+1]. \tag{14}$$

Substituting this into (9), the first two terms can be simplified using (7) and (12), and the third term can be expanded as

$$\sum_{k=0}^{T-1} \mathbf{v}[k+1]\mathbf{v}^T[k+1] = \begin{bmatrix} \sum_{k=0}^{T-1} v_1^2[k+1] & \dots & \sum_{k=0}^{T-1} v_1[k+1]v_n[k+1] \\ \sum_{k=0}^{T-1} v_2[k+1]v_1[k+1] & \dots & \sum_{k=0}^{T-1} v_2[k+1]v_n[k+1] \\ \vdots & \dots & \vdots \\ \sum_{k=0}^{T-1} v_n[k+1]v_1[k+1] & \dots & \sum_{k=0}^{T-1} v_n^2[k+1] \end{bmatrix}. \tag{15}$$

From the Cauchy-Schwarz inequality,

$$\left(\sum_{k=0}^{T-1} v_1[k+1]v_2[k+1]\right)^2 \leq (\sum_{k=0}^{T-1} v_1^2[k+1])(\sum_{k=0}^{T-1} v_2^2[k+1]). \tag{16}$$

Therefore, all the non-diagonal terms can be ignored. Equating the qth entry along the diagonal, we have

$$\sum_{k=0}^{T-1} v_q^2[k+1] + o(\sum_{k=0}^{T-1}(v_q^2[k+1]) = o(T). \tag{17}$$

Since this is true for all $q \in \{1, 2, ..., n\}$, dividing the above by T, and limiting $T \to \infty$ completes the proof. ■

Noting that $\mathbf{z}[t+1] = f(\mathbf{z}[t]) + B(g_t(\mathbf{z}^t) + \mathbf{e}[t]) + (\mathbf{w}[t+1]+]\mathbf{v}[t+1])$, the implication is that the adversary can at best additively corrupt the ambient noise $\{\mathbf{w}[t]\}$ by a signal $\{\mathbf{v}[t]\}$ of zero power if it is to stay undetected.

IV. Nonlinear Stochastic Systems

In the previous section, it was proved that Dynamic Watermarking can detect any attacks for which $\{\mathbf{v}[t]\}$ is of non-zero power provided the nonlinear stochastic system is representable in the form depicted in (1). In this section, we show that there are two kinds of systems, discrete-time versions of backstepping type, that can be so represented.

A. Backstepping

Motivated by the strict-feedback form for backstepping [16], [17], we consider the Euler version of such a system with step size h. We also allow for each state to have a white measurement noise w_i, a Gaussian distributed random vector $\sim N(0, \sigma_w^2)$. First we examine the case below where each g_i is a constant:

$$\begin{aligned} x_i[t+1] = x_i[t] + h(f_i(x_1[t], ..., x_i[t]) + g_i x_{i+1}[t]) \\ + w_i[t+1], \end{aligned} \tag{18}$$

for $1 \leq i \leq n-1$, and

$$\begin{aligned} x_n[t+1] = x_n[t] + h(f_n(x_1[t], ..., x_n[t]) + g_n u[t]) \\ + w_n[t+1]. \end{aligned} \tag{19}$$

By rearranging the (18,19) it can be written in the following form, which is of the same form of (1).

$$\begin{bmatrix} x_1[t+1] \\ x_2[t+1] \\ \vdots \\ x_{n-1}[t+1] \\ x_n[t+1] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ hg_nu[t] \end{bmatrix} + \begin{bmatrix} w_1[t+1] \\ w_2[t+1] \\ \vdots \\ w_{n-1}[t+1] \\ w_n[t+1] \end{bmatrix} + \begin{bmatrix} x_1[t] + h(f_1(x_1[t]) + g_1x_2[t]) \\ x_2[t] + h(f_2(x_1[t], x_2[t]) + g_2x_3[t]) \\ \vdots \\ x_{n-1}[t] + h(f_{n-1}(x_1[t], ..., x_{n-1}[t]) + g_{n-1}x_n[t]) \\ x_n[t] + hf_n(x_1[t], ..., x_n[t]) \end{bmatrix}. \tag{20}$$

This system in "strict-feedback" form can be stabilized by applying the discrete version of backstepping recursively. Therefore, for those discrete-time systems that can be stabilized by backstepping and have constant coefficients $g_1, g_2, ..., g_n$, Dynamic Watermarking can be applied to detect attacks.

V. Simulation Results

In this section, we consider a specific example of a nonlinear stochastic system on which we evaluate Dynamic Watermarking. The Tunnel-Diode Circuit depicted in Figure (1) is a system that consists of a resistor, an inductor, a capacitor, and a diode. It is a system with two outputs, v_c and i_L, and one input, E. The goal of the system's controller is to regulate the voltage of the capacitor, v_c, to the desired setpoint.

Let $x_1 = v_c$, $x_2 = i_L$, and $u = E$. The current i_c is given by $C\frac{dx_1}{dt}$, the voltage v_L by $L\frac{dx_2}{dt}$, and the current i_R

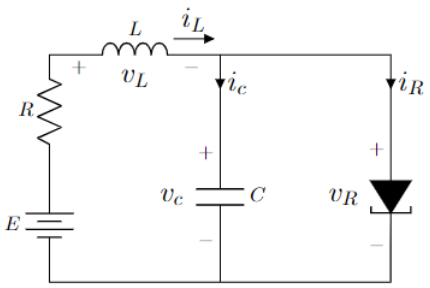

Fig. 1: Tunnel-Diode Circuit

by $f_1(x_1)$, where $f_1(x_1)$ is a nonlinear function represented as $(17.76x_1 - 103.79x_1^2 + 229.62x_1^3 - 226.31x_1^4 + 83.72x_1^5)$. According to Kirchhoff's law, the system dynamics can be presented in strict-feedback form as

$$\begin{aligned} \dot{x_1} &= \frac{-1}{C} f_1(x_1) + \frac{1}{C} x_2, \\ \dot{x_2} &= \frac{1}{L}[-x_1 - Rx_2] + \frac{1}{L} u. \end{aligned} \tag{21}$$

Therefore, the Tunnel-Circuit is suitable for Dynamic Watermarking.

The above motivates the discrete-time analog:

$$\begin{aligned} \begin{bmatrix} x_1[t+1] \\ x_2[t+1] \end{bmatrix} &= \begin{bmatrix} x_1[t] - \frac{h}{C} f_1(x_1[t]) + \frac{h}{C} x_2[t] \\ \frac{-h}{L} x_1[t] + (1 - \frac{hR}{L}) x_2[t] \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ \frac{h}{L} u[t] \end{bmatrix} + \begin{bmatrix} w_1[t+1] \\ w_2[t+1] \end{bmatrix}. \end{aligned} \tag{22}$$

In the simulation experiment, two types of attacks have been tested: false data injection attacks and replay attacks. For false data injection attacks, both bias injection and noise injection attacks have been studied. The setpoint of v_c changes periodically, as shown in Figure 2. All attacks begin at the 8 second mark. For the tests in (4) and (5), instead of taking $T \to \infty$, the sliding window method was used to calculate the values. This results in a 2×2 matrix for Test 1 and a 2×1 matrix for Test 2. Let $\mathbf{z}[k+1] - f(\mathbf{z}[k]) - Bg_k(\mathbf{z}^k) := [r_1, r_2]^T$, where r_1 and r_2 denote the watermark and the noise in v_c and i_L, respectively. The tests reduce to calculating the variance of the following matrices.

$$\begin{bmatrix} r_1 r_1 & r_1 r_2 \\ r_2 r_1 & r_2 r_2 \end{bmatrix}, \begin{bmatrix} e r_1 \\ e r_2 \end{bmatrix}. \tag{23}$$

Since, the measurement of v_c is under attack, one can expect that the test values that are related to v_c should change after the attack. The attack also affects the i_L; however, in this specific example, the changes are too small to be noticeable. From (23), except for the $r_2 r_2$ in Test 1 and $e r_2$ in Test 2, the rest should respond to the attack.

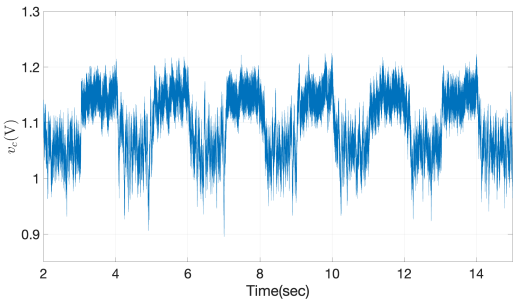

Fig. 2: v_c measurement during nominal operation.

A. Bias Injection Attack

In the bias injection attack, a constant is added to the output measurement. This alteration does not render the system unstable, but it does shift the steady state of the system, potentially impacting its overall performance. Figure 3 provides an example illustrating how a bias injection attack alters the system's steady state. Dynamic watermarking Test 1 and Test 2 are shown in Figure 4 and Figure 5, respectively. The tests detect the attacks within 0.1 seconds.

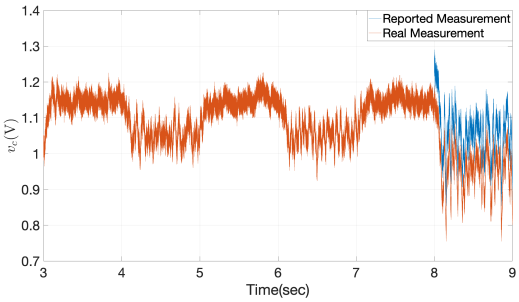

Fig. 3: v_c measurement under bias injection attack.

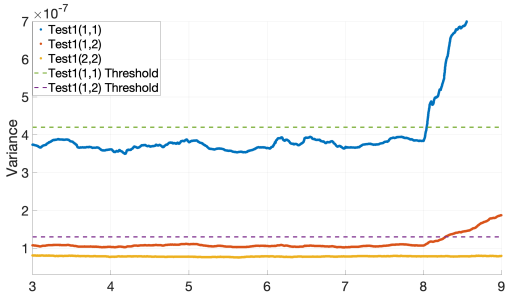

Fig. 4: Dynamic Watermarking Test 1 for v_c under bias injection attack.

B. Noise Injection Attack

In a noise injection attack, the malicious agent generates a random value at every time step and adds it to the output measurement. Typically, this attack does not cause instability in the system. It may be difficult to discern whether the system is under attack simply by observing the output value.

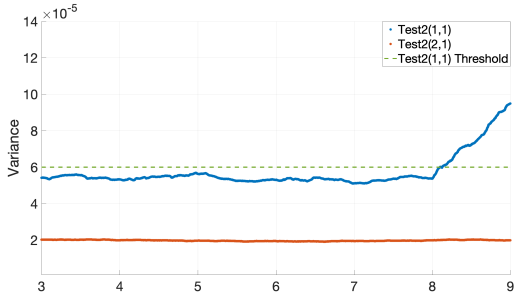

Fig. 5: Dynamic Watermarking Test 2 for v_c under bias injection attack.

An example of a noise injection attack is illustrated in Figure 6. However, the additional noise introduced can degrade the performance of the system. Dynamic watermarking Test 1 and Test 2 are shown in the Figure 7 and Figure 8, respectively. The tests detect such attacks within 0.3 seconds.

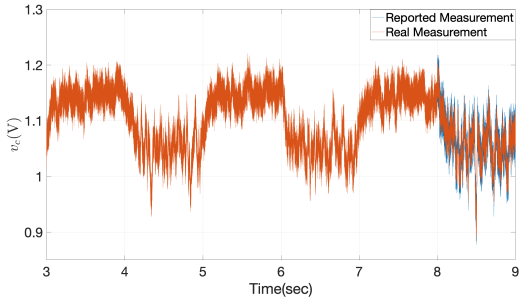

Fig. 6: v_c measurement under noise injection attack.

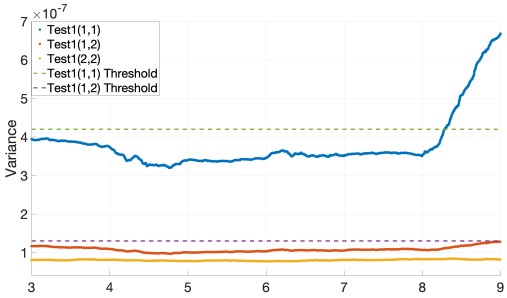

Fig. 7: Dynamic Watermarking Test 1 for v_c under noise injection attack.

C. Replay Attack

In a replay attack, a series of output measurements under nominal operation is recorded. When the attack begins, the prerecorded data is used to replace the real output measurement and is sent to the feedback loop. Figure 9 illustrates a replay attack on the v_c measurement. Dynamic watermarking Test 1 and Test 2 are shown in Figure 10 and Figure 11, respectively. The tests detect such attacks within 0.2 seconds.

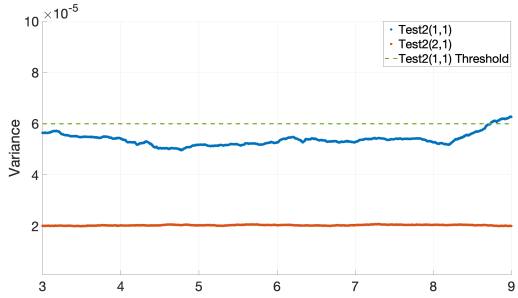

Fig. 8: Dynamic Watermarking Test 2 for v_c under noise injection attack.

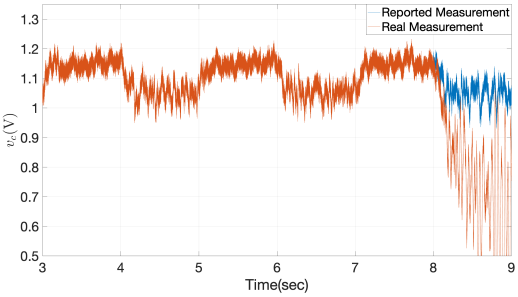

Fig. 9: v_c measurement under replay attack.

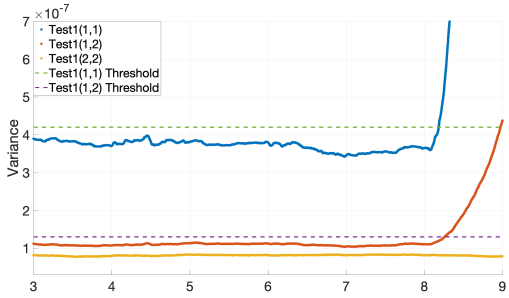

Fig. 10: Dynamic Watermarking Test 1 for v_c under replay attack.

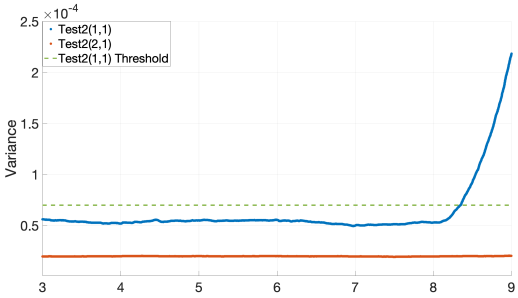

Fig. 11: Dynamic Watermarking Test 2 for v_c under replay attack.

VI. Conclusion

In this paper, the applicability of Dynamic Watermarking for attack detection in certain nonlinear stochastic cyber-physical systems is addressed. It has been shown how certain

backstepping type systems are suitable for Dynamic Watermarking. Thus Dynamic Watermarking can be potentially applied to a wider range of systems beyond linear systems. The simulation results show how Dynamic Watermarking detects both false data injection attacks and replay attacks in such nonlinear stochastic systems.

REFERENCES

[1] A. Chattopadhyay and U. Mitra, "Security against false data-injection attack in cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 2, pp. 1015–1027, 2020.

[2] H. Liu, Y. Mo, and K. H. Johansson, *Active Detection Against Replay Attack: A Survey on Watermark Design for Cyber-Physical Systems*, pp. 145–171. Cham: Springer International Publishing, 2021.

[3] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, S. Sastry, *et al.*, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, vol. 5, Citeseer, 2009.

[4] 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, 2012.

[5] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[6] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *47th Annual Allerton Conference on Communication, Control, and Computing*, Sept 2009.

[7] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems*, vol. 35, pp. 93–109, Feb 2015.

[8] B. Satchidanandan and P. R. Kumar, "Dynamic watermarking: Active defense of networked cyber–physical systems," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 219–240, 2017.

[9] T. Wigren and A. Teixeira, "Feedback path delay attacks and detection," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 3864–3871, 2023.

[10] S. Narasimhan, N. H. El-Farra, and M. J. Ellis, "A reachable set-based cyberattack detection scheme for dynamic processes," in *2023 American Control Conference (ACC)*, pp. 3777–3782, 2023.

[11] W. Duo, M. Zhou, and A. Abusorrah, "A survey of cyber attacks on cyber physical systems: Recent advances and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 5, pp. 784–800, 2022.

[12] R. Goyal, C. Somarakis, E. Noorani, and S. Rane, "Co-design of watermarking and robust control for security in cyber-physical systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 4821–4826, 2022.

[13] W.-H. Ko, B. Satchidanandan, and P. R. Kumar, "Theory and implementation of dynamic watermarking for cybersecurity of advanced transportation systems," in *2016 IEEE Conference on Communications and Network Security (CNS)*, pp. 416–420, 2016.

[14] M. Porter, P. Hespanhol, A. Aswani, M. Johnson-Roberson, and R. Vasudevan, "Detecting generalized replay attacks via time-varying dynamic watermarking," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3502–3517, 2021.

[15] T. L. Lai and C. Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *The Annals of Statistics*, vol. 10, no. 1, pp. 154–166, 1982.

[16] P. Kokotovic, "The joy of feedback: nonlinear and adaptive," *IEEE Control Systems Magazine*, vol. 12, no. 3, pp. 7–17, 1992.

[17] I. Kanellakopoulos, P. Kokotovic, and A. Morse, "Systematic design of adaptive controllers for feedback linearizable systems," *IEEE Transactions on Automatic Control*, vol. 36, no. 11, pp. 1241–1253, 1991.