Stealthy Measurement-Aided Pole-Dynamics
Attacks With Nominal Models
Dajun Du, Member, IEEE, Changda Zhang, Chen Peng, Senior Member, IEEE,
Minrui Fei, and Huiyu Zhou

Abstract—When traditional pole-dynamics attacks (TPDAs) are implemented with nominal models, model mismatch between exact and nominal models often affects their stealthiness, or even makes the stealthiness lost. To solve this problem, this article presents a novel stealthy measurement-aided pole-dynamics attacks (MAPDAs) method with model mismatch. First, the limitations of TPDAs using exact models are revealed. Second, to handle the limitations, the proposed MAPDAs method is designed by using an adaptive control strategy, which can keep the stealthiness. Moreover, it is easier to implement as only the measurements are needed in comparison with the existing methods requiring both measurements and control inputs. Third, the performance of the proposed MAPDAs method is explored using convergence of multivariate measurements, and MAPDAs with model mismatch have the same stealthiness and similar destructiveness as TPDAs. Finally, experimental results from a networked inverted pendulum system confirm the feasibility and effectiveness of the proposed method.

Index Terms—Adaptive control, convergence, model mismatch, pole-dynamics attacks (PDAs), stealthiness.

I. INTRODUCTION

NETWORKED control systems (NCSs) [1], [2], [3] deploy communication networks to exchange information between physical entities, such as plants, sensors, and controllers. Compared with traditional control systems, NCSs eliminate unnecessary wiring, reduce system complexity and cost, and improve system performance. However, the usage of networks makes NCSs open to the outer space and cost, and improve system performance. However, the NCSs eliminate unnecessary wiring, reduce system complexity and controllers. Compared with traditional control systems, NCSs eliminate unnecessary wiring, reduce system complexity and cost, and improve system performance. However, the usage of networks makes NCSs open to the outer space and the stealthiness of attacks.

The stealthiness of attacks has been known, model-based attacks can be of the stealthiness if they are designed to engage the measurements, control inputs and both. However, it is unrealistic to retrieve exact models used by the attacker or defender in many industrial control systems, leading to model mismatch between exact and nominal models. With nominal models, model-based attacks may lose their stealthiness. This brings a consequent question: Are model-based attacks helpless against model mismatch? The answer is no, and there actually are several improved model-based attacks methodologies, e.g., data-driven feedback-loop/two-loop covert attacks [16], [17], robust zero-dynamics attacks [18], and robust PDAs [19].

Although the existing improved model-based attacks methods have provided promising performance, control inputs are indispensable to the outcome of these methods, e.g., both control inputs and measurements are required to design attacks. This brings a new question: Can improved model-based attacks methods without using control inputs be working against model mismatch between exact and nominal models? In a practical sense, there exist several vulnerable-sensor-network-only NCSs (especially Internet of Things applications [20], [21], [22]) where a vulnerable wireless network may be used to link the sensors and the controller, and reliable cable networks may be used to connect the controller with the plant. Therefore, from the perspective of a defender, the above question is equivalent to this one: Are vulnerable-sensor-network-only NCSs safe enough from these stealthy

Manuscript received 20 February 2024; revised 24 June 2024 and 14 August 2024; accepted 31 August 2024. Date of publication 23 September 2024; date of current version 30 October 2024. This work was supported in part by the National Science Foundation of China under Grant 62373240, Grant 62273224, Grant 62173218, and Grant 62403182; and in part by the Project of Shanghai Science and Technology Commission of Shanghai Municipality under Grant 21190780300. This article was recommended by Associate Editor B. Ribeiro.

Dajun Du, *Changda Zhang* is with the Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China (e-mail: djd@shu.edu.cn; c.peng@shu.edu.cn; mfei@staff.shu.edu.cn).

Changda Zhang is with the School of Electrical Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China (e-mail: silenceupdown@hotmail.com).

Huiyu Zhou is with the School of Computing and Mathematical Sciences, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: hz143@leicester.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2024.3456084.

Digital Object Identifier 10.1109/TCYB.2024.3456084.
attacks thanks to model mismatch between exact and nominal models?

Motivated by the above observations, the following challenges and difficulties will be addressed.

1) Traditional PDAs (TPDAs) are implemented with exact models, which is impractical for some attackers. They have no choice but to use nominal models to design attacks, however model mismatch between exact and nominal models may lead to decline or even loss of stealthiness. Therefore, how to reveal the limitations of TPDAs with nominal models is the first challenge.

2) Some popular techniques (e.g., robust control and data driven) can be employed to improve the stealthiness of PDAs with nominal models requiring complete and accurate measurements and control inputs. It is difficult for the attacker to launch attacks by obtaining these signals especially in vulnerable-sensor-network-only NCSs. Therefore, how to propose a new attack method without control input is the second challenge.

3) The existing improved model-based attacks (e.g., robust zero-dynamics attacks and robust PDAs) have been mainly designed for single-input–single-output systems, which have rarely been implemented in multiple-input–multiple-output (MIMO) systems. When the above proposed attack method is employed in MIMO systems, identification of stealthiness and destructiveness is the third challenge.

To deal with the above challenges and difficulties, this article presents a stealth measurement-aided PDAs (MAPDAs) method with model mismatch for uncertain vulnerable-sensor-network-only NCSs. Comparative analysis between the existing methods and the proposed method is listed in Table I. The existing methods are mainly based on an exact model or a nominal model but require complete and accurate control inputs, but this article has revealed the limitations of TPDAs, proposed the new MAPDAs method with model mismatch, and provided the proof of stealthiness and destructiveness of MAPDAs. The main contributions of this article are summarized as follows.

1) The limitations of TPDAs using exact models are revealed, where exact models can ensure the stealthiness of TPDAs but model mismatch between the exact and nominal models may cause TPDAs to lose stealthiness.

2) To handle model mismatch, a new MAPDAs method is proposed using a model reference adaptive control strategy, which can keep the stealthiness. Moreover, it is easier to implement as only the measurements are needed in comparison with the existing methods requiring both the measurements and control inputs.

3) The stealthiness and destructiveness of the proposed MAPDAs in MIMO systems is explored by investigating the convergence of multivariate measurements, where MAPDAs with model mismatch have the same stealthiness and similar destructiveness as TPDAs. Specifically, MAPDAs with adaptive gains will remain stealthy at an acceptable detection threshold till destructiveness occurs.

The remainder of this article is organized as follows. Section II is problem formulation, where the limitations of TPDAs with model mismatch are discussed. Section III describes the proposed MAPDAs and analyzes the performance. Section IV provides the experiments where TPDAs and MAPDAs are embedded in the practical networked inverted pendulum visual servo system (NIPVSS), followed by the conclusions made in Section V.

Notation: Matrix $P > 0$ denotes that $P$ is positive definite symmetric. The one vector (all elements are 1) is denoted by $1_n \in \mathbb{R}^n$. Table II summarizes the notations most frequently used throughout the rest of this article.

II. PROBLEM FORMULATION

A. NCSs Under TPDAs With Exact Model

The framework of NCSs for TPDAs with exact auxiliary model (i.e., exact model) is shown in Fig. 1. First, the sensor obtains the measurement $x(t)$ from the plant. Then, $x(t)$ will be transmitted to the estimator and controller via networks, becoming $x_a(t)$ due to injection of attack signals $a(t)$ from possible TPDAs with exact auxiliary model. Using $x_a(t)$, the
controller calculates control input \( u(t) \) that is sent to the actuator to stabilize the plant and the estimator judges whether or not there exists an attack, and if there is an attack, the alarm will be triggered.

**Remark 1:** Fig. 1 shows the framework of NCSs for TPDAs with the exact auxiliary model, the construction of TPDAs adopts the exact auxiliary model that only needs the matrix \( A \) of the physical system, see [15]. However, when there exists model mismatch between the exact auxiliary model and the nominal model (i.e., the attackers cannot produce an exact auxiliary model), the complete and accurate measurement \( x(t) \) and control input \( u(t) \) have been applied to constructing model-based attacks for enhancing the stealthiness (see [16], [17], [18], [19]). However, could the stealthiness under model mismatch be enhanced by the only measurements \( x(t) \)? For the question, this article says yes and gives the corresponding method.

Consider continuous linear time-invariant (LTI) plant\(^1\)

\[
\dot{x}(t) = Ax(t) + Bu(t) \tag{1}
\]

\[
z(t) = Cx(t) \tag{2}
\]

where \( x(t) \in \mathbb{R}^n \) is system state and measurement, \( u(t) \in \mathbb{R}^m \) is control input, \( z(t) \in \mathbb{R}^r \) is controlled output, \( t \in [0, \infty) \), \( t_0 \) is the initial instant, \( x(t_0) \) is the initial state, and \( \|x(t_0)\| < \infty \). \( A, B, \) and \( C \) are constant matrices with appropriate dimensions. Without loss of generality, it is considered that (1) is controllable.

**Remark 2:** The general measured output is \( y(t) = Hx(t) \). For simplicity, \( H = I \) is considered as [24] and thus \( x(t) \) is the measured output. The condition \( H \neq I \) can be investigated in the future work.

System state \( x(t) \) will be transmitted to the controller of TPDAs via networks. TPDAs can maintain a continuous exact auxiliary model \( A_{c, t}(A) \) [15]

\[
\dot{x}_{eam}(t) = Ax_{eam}(t) \tag{3a}
\]

\[
a(t) = x_{eam}(t) \tag{3b}
\]

where \( x_{eam}(t) \) and \( a(t) \) are the state and the output of \( A_{c, t}(A) \). In a network, \( a(t) \) may be subtracted from \( x(t) \), and thus the network output becomes

\[
x_a(t) = x(t) - a(t). \tag{4}
\]

\(^1\)Strictly speaking, the poles set of an LTI system is a subset of the eigenvalues set of \( A \) [23, Lec. 19.2]. For this reason, \( \dot{x}(t) = Ax(t) \) is called pole-dynamics.

Using \( x_a(t) \), the controller calculates the control input

\[
u(t) = K x_a(t) \tag{5}
\]

where \( K \) is the controller gain and has been designed to make \( \Phi := A + BK \) stable (i.e., the eigenvalues of \( \Phi \) are located on the closed left half-plane). Then, \( u(t) \) will be sent to the actuator for stabilizing the plant.

**Remark 3:** The general controller is \( u(t) = K x_a(t) + N x(t) \) [25], where \( N \) is a gain matrix and \( x(t) \) is the reference input. Considering that the aim of the controller in the attack-free case is to bring \( x(t) \) to zero, it is set \( r(t) = 0 \) as [26]. Therefore, the controller form (5) is adopted.

To detect attacks, the common norm-based test is performed by the detector, i.e., if there is

\[
\|x_a(t)\| < \epsilon \tag{6}
\]

where \( \epsilon > 0 \) is a user-defined threshold, it means that there is no attack, otherwise, attacks emerge.

**Remark 4:** The threshold \( \epsilon \) of the detector (6) is important to check the validity of attack detection. The existing threshold selection methods generally include statistical analysis [27], theoretical derivation [28], machine learning [29], etc. To determine the proper threshold, the method of statistical analysis is used in Section IV.

To analyze the performance of attacks, according to the discussion on stealthiness [30] and destructiveness [18], the definitions of \( \epsilon \)-stealthiness and \( \xi \)-destructiveness in time period \( T := [t_0, t_f] \) of attacks are given in the following, where \( t_0, t_f \) are initial and finishing instants of attacks, respectively.

**Definition 1** (\( \epsilon \)-Stealthiness): An attack is said to be with \( \epsilon \)-stealthiness on the detector in \( T \) when \( (6) \) for \( t \in T \) always holds.

**Definition 2** (\( \xi \)-Destructiveness): An attack is said to be with \( \xi \)-destructiveness on the controlled output \( z(t_f) \) if

\[
\|z(t_f)\| \geq \xi \tag{7}
\]

where \( \xi \) is the admissible state limit. Specifically, \( \|z(t)\| < \xi \) for \( t \in T \) will run under control and \( \|z(t)\| \geq \xi \) for \( t \in T \) is actively out of control (e.g., takes active protection measures) to avoid possible severe accidents.

It is well believed that an ideal attack in \( T \) should be with both \( \xi \)-destructiveness and \( \epsilon \)-stealthiness. We may witness a more dangerous scenario where a quasi-ideal attack is with \( \epsilon \)-stealthiness and with no \( \xi \)-destructiveness in \( T \), but the controlled output is driven very close to \( \xi \). A quasi-ideal attack could be on the synchronous machines [31], where the attack will not make rotation rates of synchronous machines cross the admissible limit, but it pushes the rotation rate to be high. This will remarkably shorten the life of synchronous machines and even cause accidents. For simplicity, this article only focuses on ideal attacks.

**B. Performance of TPDAs With Exact Auxiliary Models**

Based on the above NCSs under TPDAs, the definitions and impact analysis of TPDAs with \( A_{c, t}(A) \) (3) in [15], and the stealthiness and destructiveness of TPDAs with \( A_{c, t}(A) \) (3) are presented in Theorem 1.
Before Theorem 1, Lemma 1 will be given. For exact auxiliary model $\mathcal{A}_{\varepsilon}(A)$ (3), considering the eigenvalues $\lambda_i$ ($i = 1, \cdots, p$) of the matrix $A$, $A$ can be expressed as $A = XJX^{-1}$, where $X$ is a constant matrix and $J = \text{diag}[J_1, \cdots, J_i, \cdots, J_r]$, $r_1 \leq p$ and when $\lambda_i$ is a single root, $J_i = \lambda_i$; when $\lambda_i$ is an $r_2$-fold root, i.e., $\lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+r_2-1}$, then

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & 1 & \\ \lambda_{i+r_2-1} & & & \end{bmatrix}.$$ 

To conveniently present Lemma 1, $\psi(t) := X^{-1}x_{\varepsilon}(t)$ is denoted.

**Lemma 1:** If and only if $\psi(t_0)$ with $\|\psi(t_0)\| < \infty$ satisfies:
1) when $\lambda_1$ locates on the open right half-plane, there are two cases: a) if $\lambda_1$ is a single root, $\psi(t_0) = 0$ and b) if $\lambda_1$ is an $r_2$-fold root, $\psi(t_0) = \cdots = \psi_{r_2-1}(t_0) = 0$;
2) when $\lambda_j$ locates on the closed left half-plane, if $\lambda_j$ is an $r_2$-fold root, $\psi_j(t_0) = \cdots = \psi_{j+r_2-1}(t_0) = 0$.
then the state $x_{\varepsilon}(t)$ of exact auxiliary model $\mathcal{A}_{\varepsilon}(A)$ (3) with unstable $A$ will converge to 0.

**Proof:** Considering $A = XJX^{-1}$, the solution of exact auxiliary model $\mathcal{A}_{\varepsilon}(A)$ (3) is

$$\psi(t) = e^{lt} \psi(t_0). \quad (8)$$

Considering the eigenvalues $\lambda_i$ ($i = 1, \cdots, p$) of matrix $A$, without loss of generality, we set that $\lambda_1$ is an $r_2$-fold root, i.e., $\lambda_1 = \lambda_2 = \cdots = \lambda_r$, and $\lambda_{r+1}, \cdots, \lambda_p$ are all single roots. Then, (8) can be rewritten as

$$\psi(t) = \begin{bmatrix} e^{\lambda_1 t} \psi_1(t_0) + e^{\lambda_2 t} \psi_2(t_0) + \cdots + e^{\lambda_{r_2-1} t} \psi_{r_2-1}(t_0) \\ \vdots \\ e^{\lambda_1 t} \psi_{r_2-1}(t_0) + e^{\lambda_1 t} \psi_2(t_0) \\ e^{\lambda_1 t} \psi_1(t_0) \\ \vdots \\ e^{\lambda_1 t} \psi_p(t_0) \end{bmatrix}.$$ 

$$\psi(t) = \begin{bmatrix} e^{\lambda_1 t} \psi_1(t_0) + e^{\lambda_2 t} \psi_2(t_0) + \cdots + e^{\lambda_{r_2-1} t} \psi_{r_2-1}(t_0) \\ \vdots \\ e^{\lambda_1 t} \psi_{r_2-1}(t_0) + e^{\lambda_1 t} \psi_2(t_0) \\ e^{\lambda_1 t} \psi_1(t_0) \\ \vdots \\ e^{\lambda_1 t} \psi_p(t_0) \end{bmatrix}.$$ 

**(9)**

**Sufficiency:** When the items 1) and 2) of Lemma 1 are satisfied, if the $r_2$-fold root $\lambda_1$ is located on the open right half-plane, $\psi_1(t_0) = \cdots = \psi_2(t_0) = 0$; if $\lambda_j$ is located on the closed left half-plane, $\psi_j(t_0) = 0$. It guarantees that $\psi(t_0)$, $\cdots$, $\psi_{r_2}(t_0)$ will converge to 0. When the items 1) and 2) of Lemma 1 are satisfied, if the single root $\lambda_j$, $j \in \{ r_2 + 1, \cdots, p \}$ is located on the open right half-plane, $\psi_j(t_0) = 0$. It guarantees that $\psi_{2r_j-1}(t_0), \cdots, \psi_p(t_0)$ will converge to 0. Therefore, the state $x_{\varepsilon}(t)$ of exact auxiliary model $\mathcal{A}_{\varepsilon}(A)$ (3) with unstable $A$ will converge to 0.

**Necessity:** When the state $x_{\varepsilon}(t)$ of exact auxiliary model $\mathcal{A}_{\varepsilon}(A)$ (3) with unstable $A$ converges to 0, it is necessary that $\psi_2(t_0) = \cdots = \psi_{r_2}(t_0) = 0$. If the $r_2$-fold root $\lambda_1$ is located on the open right half-plane, it is also necessary that $\psi_1(t_0) = 0$. Furthermore, if the single root $\lambda_j$, $j \in \{ r_2 + 1, \cdots, p \}$ is located on the open right half-plane, it is necessary that $\psi_j(t_0) = 0$. It completes the proof.

Now, based on Lemma 1, Theorem 1 will be given in the following.

**Theorem 1:** Considering system (1)–(5) under TPDAs with $\mathcal{A}_{\varepsilon}(A)$ (3), if $\Phi$ is stable, $A$ is unstable (i.e., at least one of eigenvalues of $A$ is located on the open right half-plane), and $x_{\varepsilon}(t_0)$ does not satisfy the item 1) or 2) of Lemma 1, then

$$\exists \epsilon, \|x_a(t)\| < \epsilon \quad \text{for} \quad t \in [t_0, \infty). \quad (10)$$

The norm of system state becomes unbounded, i.e.,

$$\lim_{t \to \infty} \|x(t)\| \to \infty. \quad (11)$$

**Proof:** Considering (1) and (3), the dynamics of $x_a(t)$ is

$$\dot{x}_a(t) = \Phi x_a(t). \quad (12)$$

If $\Phi$ is stable, it follows that

$$\|x_a(t)\| \leq \sup_{\varepsilon} \|x_a\|, \quad t \geq t_0 \quad (13)$$

where $\sup_{\varepsilon} \|x_a\| := \kappa \|x_a(t_0)\|$ and $\kappa$ is a constant. If $\epsilon > \sup_{\varepsilon} \|x_a\|$, then (10) is guaranteed.

It can be seen from (12) and (3) that $\lim_{t \to \infty} x_a(t) = 0$ if $\Phi$ is stable, and when $x_{\varepsilon}(t_0)$ does not satisfy the item 1) or 2) of Lemma 1, $\lim_{t \to \infty} \|x_{\varepsilon}(t)\| \to \infty$ holds if $A$ is unstable. Considering (4), to make $\lim_{t \to \infty} x_a(t_0) = 0$ hold, $\lim_{t \to \infty} \|x(t)\| = 0$ must hold when $\lim_{t \to \infty} \|x_{\varepsilon}(t)\| \to \infty$ holds. Therefore, (11) is guaranteed. It completes the proof.

**Remark 5:** For Theorem 1, there possibly exist two cases for TPDAs with $\mathcal{A}_{\varepsilon}(A)$ (3), i.e., $\epsilon > \sup_{\varepsilon} \|x_a\| \|x_a\|$ represents the upper bound of $\|x_a(t)\|$ under TPDAs with $\mathcal{A}_{\varepsilon}(A)$ (3) and $\epsilon \leq \sup_{\varepsilon} \|x_a\|$, which is shown in
Fig. 2(a). Therefore, for a given \( \epsilon \), when a small enough super Steve is selected, TPDAs with \( A_{c,1}(A) \) (3) can be with \( \xi \)-destructiveness and \( \epsilon \)-stealthiness in \( T \).

Remark 6: For Lemma 1, we examine whether or not the initial value of \( x_{e_{1}}(t_{0}) \) [i.e., \( x_{e_{1}}(t_{0}) \)] corresponding to the \( i \)th eigenvalue \( \lambda_{i} \) of \( A_{c,1} \) [i.e., \( A_{c,1} \) corresponding to the \( i \)th eigenvalue \( \lambda_{i} \)] equals to zero. There are two cases for \( x(t) \).

1) If \( x_{e_{1}}(t_{0}) \) satisfies the initial condition in Lemma 1, \( \lim_{t \to \infty} x_{e}(t) = 0 \) even if \( A \) is unstable. Furthermore, according to (3), (4), and (12), \( \lim_{t \to \infty} x(t) = 0 \) so that (11) will not hold.

2) If \( x_{e_{1}}(t_{0}) \) does not satisfy the initial condition in Lemma 1, \( \lim_{t \to \infty} x(t) \to \infty \) as \( A \) is unstable. Furthermore, according to (3), (4), and (12), \( \lim_{t \to \infty} \| x(t) \| \to \infty \), and (11) in Theorem 1 holds.

C. Limitation of TPDAs With Nominal Models

Although the above stealthiness and destructiveness of TPDAs with \( A_{c,1}(A) \) (3) look promising, it is unrealistic to obtain the exact model for the attacker (even for the defender). When the attacker only knows the nominal model of uncertain NCs (i.e., the nominal model \( A_{n} \) of \( A \)), they have to perform TPDAs with a continuous nominal auxiliary model \( A_{c,1}(A_{n}) \)

\[
\dot{x}_{a}(t) = A_{n}x_{a}(t) + \Phi(t)x_{a}(t) + \epsilon(t)
\]

\[
A(t) = A_{n}x_{a}(t) + \Phi(t)x_{a}(t) + \epsilon(t)
\]

According to impact analysis of TPDAs with \( A_{c,1}(A) \) (3) in [15], the limitation of TPDAs with \( A_{c,1}(A_{n}) \) (14) is presented in Theorem 2.

Before Theorem 2, Lemma 2 will be given in the following. For nominal auxiliary model \( A_{c,1}(A_{n}) \) (14), considering the eigenvalues \( \lambda_{n,i} \) (i.e., \( I, \ldots, p \)) of the matrix \( A_{n} \). \( A_{n} \) can be expressed as \( A_{n} = X_{n} \) where \( X_{n} \) is a constant matrix and \( J_{n} = \text{diag}(J_{n,1}, \ldots, J_{n,p}) \), \( J_{n} = \text{diag}(J_{n,1}, \ldots, J_{n,p}) \), \( n_{n,1} \leq n \) and when \( \lambda_{n,i} \) is a single root, \( J_{n,i} = \lambda_{n,i} \), when \( \lambda_{n,i} \) is an \( r_{n,i} \)-fold root, \( I, \ldots, \lambda_{n,i} \), \( \lambda_{n,i} = \lambda_{n,i+1} = \cdots = \lambda_{n,i+r_{n,2}-2,1} \), then

\[
J_{n,i} = \begin{bmatrix}
\lambda_{n,i} & 1 \\
& \ddots & 1 \\
& & \lambda_{n,i+r_{n,2}-2,1}
\end{bmatrix}.
\]

To conveniently present Lemma 2, \( \psi_{n}(t) := X_{n}^{-1}x_{n}(t) \) is denoted.

Lemma 2: If and only if \( \psi_{n}(t_{0}) \) or \( \| \psi_{n}(t_{0}) \| < \infty \) satisfies:

1) when \( \lambda_{n,i} \) locates on the open right half-plane, there are two cases: a) if \( \lambda_{n,i} \) is a single root, \( \psi_{n,i}(t_{0}) = 0 \) and b) if \( \lambda_{n,i} \) is an \( r_{n,2} \)-fold root, \( \psi_{n,i}(t_{0}) = \cdots = \psi_{n,i+r_{n,2}-2,1}(t_{0}) = 0 \);

2) when \( \lambda_{n,i} \) locates on the closed left half-plane, if \( \lambda_{n,i} \) is an \( r_{n,2} \)-fold root, \( \psi_{n,i}(t_{0}) = \cdots = \psi_{n,i+r_{n,2}-2,1}(t_{0}) = 0 \), then the state \( x_{n}(t) \) of nominal auxiliary model \( A_{c,1}(A_{n}) \) (14) with unstable \( A_{n} \) will converge to 0.

Proof: The proof is similar as that of Lemma 1, which is omitted.

Now, based on Lemma 2, Theorem 2 will be given in the following.

Theorem 2: Considering the system (1), (2), (4), (5) under TPDAs with \( A_{c,1}(A_{n}) \) (14), if \( \Phi \) is stable, \( x_{n}(t_{0}) \) does not satisfy the item 1) or 2) of Lemma 2, then

\[
\lim_{t \to \infty} \| x_{n}(t) \| \to \infty.
\]

The norm of system state becomes unbounded, i.e., (11).

Proof: Considering (1) and (14), the dynamics of \( x_{n}(t) \) are

\[
\dot{x}_{n}(t) = \Phi x_{n}(t) + (A - A_{n})x_{n}(t).
\]

If \( A_{n} \) is unstable, when \( x_{n}(t_{0}) \) does not satisfy the item 1) or 2) of Lemma 2, \( \lim_{t \to \infty} \| x_{n}(t) \| \to \infty \) (14) will hold. Therefore, although \( \Phi \) is stable, (15) is guaranteed as \( A \neq A_{n} \).

Considering (1), (4), (5), and (14), it follows that:

\[
\dot{x}(t) = \Phi x(t) - BK x_{n}(t).
\]

Since \( \lim_{t \to \infty} \| x_{n}(t) \| \to \infty \) in (14) holds when \( x_{n}(t_{0}) \) does not satisfy the item 1) or 2) of Lemma 2, (17) yields (11). It completes the proof.

Remark 7: For Theorem 2, there possibly exist two cases for TPDAs with \( A_{c,1}(A_{n}) \) (14), i.e., \( \epsilon > \| x_{n}(t) \| \) and \( \epsilon \leq \| x_{n}(t) \| \) (\( t_{f} \) is finishing instant of attacks), which is shown in Fig. 2(b). Therefore, when a small \( \epsilon \) is selected, TPDAs with \( A_{c,1}(A_{n}) \) (14) are with \( \xi \)-destructiveness but with no \( \epsilon \)-stealthiness in \( T \).

Remark 8: For Lemma 2, we have analyzed whether or not the initial value of \( x_{n}(t_{0}) \) (i.e., \( x_{n}(t_{0}) \)) corresponding to the \( i \)th eigenvalue \( \lambda_{n,i} \) of \( A_{n} \) equals to zero. There are two cases for \( x_{n}(t) \).

1) If \( x_{n}(t_{0}) \) satisfies the initial condition in Lemma 2, \( \lim_{t \to \infty} x_{n}(t) = 0 \) even if \( A_{n} \) is unstable. Furthermore, according to (4), (14), and (16), \( \lim_{t \to \infty} \| x_{n}(t) \| \to \infty \) so that (15) will not hold.

2) If \( x_{n}(t_{0}) \) does not satisfy the initial condition in Lemma 2, \( \lim_{t \to \infty} \| x_{n}(t) \| \to \infty \) because \( A_{n} \) is unstable. Furthermore, according to (4), (14), and (16), \( \lim_{t \to \infty} \| x_{n}(t) \| \to \infty \), and (15) in Theorem 2 holds.

Up to now, we understand the limitations of TPDAs with nominal models, i.e., TPDAs will be with \( \xi \)-destructiveness but with no \( \epsilon \)-stealthiness in \( T \). In the next section, to cope with this problem, we will present a measurements and adaptive control-based method into the attacks.

III. MEASUREMENT-AIDED POLE-DYNAMICS ATTACKS

We have analyzed NCs under TPDAs and the limitations of TPDAs with nominal models in the previous sections. To solve the problem, a stealthy MAPDAs method using measurements and an adaptive auxiliary model will be designed and discussed.

A. Design of MAPDAs

The framework of NCs under MAPDAs with an adaptive auxiliary model is shown in Fig. 3(a) and the framework of the adaptive auxiliary model is shown in Fig. 3(b). First, the sensor collects \( x(t) \) from the plant. Then, \( x(t) \) will be transmitted to the detector and controller via network, becoming \( x_{a}(t) \) due to possible attacks. The attacker obtains \( x_{a}(t) \) and uses it to construct adaptive auxiliary model with state \( x_{n}(t) \)
an attacker has nominal models (i.e., $A_n$ and output model. (b) Framework of an adaptive auxiliary model.

![Diagram](image)

and output $a(t)$, where $x_a(t)$ and $x_{s,am}(t)$ are used to produce adaptive gain $F_a(t)$ based on the designed adaptive laws, and $F_a(t)$ is used to update $x_{s,am}(t)$ and $a(t)$. Using $x_a(t)$, the controller calculates $u(t)$ that is sent to the actuator to stabilize the plant and the estimator scrutinizes whether or not there is an attack, and if there is an attack, the alarm will be triggered.

Consider continuous LTI plant (1) and (2) and that the attacker has nominal models (i.e., $A_n$, $B_n$, and $K_n$ of NCSs) and can obtain the data in the network. Motivated by the fact that the defender can adopt some control theories (e.g., model reference adaptive control [32], [33] and fuzzy control [34]) to make system with model mismatch convergent, the attacker can also cope with model mismatch by performing MAPDAs with a continuous adaptive auxiliary model $A_{a,1}(A_n, \Phi, x_a)$ and output $a(t)$, where $x_{s,am}(t)$ and $a(t)$ are, respectively, state and output of $A_{a,1}(A_n, \Phi, x_a)$, $F_a(t)$ is time-varying adaptive gain, $Z > 0$, $P > 0$, $Q > 0$ are constant matrices, and $x_{s,am}(t)$ is the network output in (4). The control input is $u(t)$ in (5) and the detector with the test (6) is used to detect attacks.

**Remark 9:** We will give a comprehensive explanation of adaptive control strategies from the perspective of the defender [as shown in Fig. 4(a)] and the attacker [as shown in Fig. 4(b)] in the following.

1) **Adaptive Control Strategy for the Defender:** Fig. 4(a) shows a common model reference adaptive control strategy [32], [33] for the defender (where the reference input $r(t) = 0$). The plant is expressed by (1) and the measurement is $x(t)$. To ensure that $x(t)$ is stable, a reference model is constructed as $\hat{x}_m(t) = A_m x_m(t)$ where $x_m(t)$ is the state and output of reference model, and $A_m$ is a constant matrix and needs to be Schur stable. The controller is $u(t) = K_m x(t)$ where $K_m(t)$ is a time-varying controller gain generated by adaptive law. To obtain adaptive law, letting $A_m = \hat{A} - B \hat{K}_m^\ast$, where $K_m^\ast$ is a constant matrix, the dynamic of $x_m(t) := x(t) - x_m(t)$ can be constructed as $\dot{x}_m(t) = A_m x_m(t) - B \Delta_m(t) x(t)$, where $\Delta_m(t) := K_m^\ast - K_m(t)$. Furthermore, a candidate Lyapunov function can be chosen as $V_m(t) = x_m^T(t) P_m x_m(t) + \text{tr}(\Delta_m(t) Z_m^\ast \Delta_m(t))$, where $P_m > 0$ and $Z_m > 0$ are constant matrices. The derivative $\dot{V}_m(t)$ of $V_m(t)$ is $\dot{V}_m(t) = x_m^T(t) (A_m^T P_m + P_m A_m) x_m(t) - 2 x_m^T(t) \Delta_m(t) B^T P_m x_m(t) + 2 \text{tr}(\Delta_m^T(t) Z_m^\ast \Delta_m(t))$. If

$$u(t) = K_m(t) x(t)$$

$$\dot{K}_m(t) = -Z_m B^T P_m x_m(t) x^T(t), \quad \text{[similar as] (18b)}$$

$$-Q_m = A_m^T P_m + P_m A_m, \quad \text{[similar as] (18d)}$$

where $Q_m > 0$ is a constant matrix, then $\dot{V}_m(t) < 0$, i.e., $\lim_{t \to \infty} x(t) = 0$. Due to $A_m$ is Schur stable (i.e., $\lim_{t \to \infty} x(t) = 0$), it can be yielded that $x(t)$ is stable (i.e., $\lim_{t \to \infty} x(t) = 0$).

2) **Adaptive Control Strategy for the Attacker:** Fig. 4(b) shows adaptive control strategy for the attacker proposed in this article. Compared with Fig. 4(a), main differences on Fig. 4(b) lie on that: a) the controller is constant as (5) (other than time-varying) and its input is $x_a(t)$ [other than $x(t)$]; b) the reference model becomes the adaptive auxiliary model that contains the nominal auxiliary model (14a), adaptive controller $F_a(t)$, and adaptive law on $F_a(t)$; and c) the error $x_m(t)$ (other than both the error and state) needs to converge to zero. To ensure that $x_m(t)$ is stable, the adaptive auxiliary model needs to construct (18a) with adaptive controller $F_a(t)$ generated.
by adaptive law. To obtain adaptive law, the dynamic of $x_a(t)$ can be constructed as

$$\dot{x}_a(t) = \Phi x_a(t) + (A - F_a(t) - A_n)x_{aam}(t). \quad (19)$$

Furthermore, a candidate Lyapunov function $V(t)$ can be chosen as (22) similar as $V_m(t)$, whose derivative is (23).

If there are (18), then $\dot{V}(t) < 0$, i.e., $\lim_{t \rightarrow \infty} x_a(t) = 0$.

Remark 10: The goal of MAPDAs with $A_{a,t}(A_n, \Phi, x_a)$ (18) is to drive $x_a(t)$ to converge to 0 as that of TPDAs with $A_{a,f}(A)$ (3). To achieve this goal, different from TPDAs with $A_{a,f}(A)$ (14), additional adaptive gain $F_a(t)$ and the measurement $x_a(t)$ are required for MAPDAs with $A_{a,t}(A_n, \Phi, x_a)$ (18). By using MAPDAs with $A_{a,t}(A_n, \Phi, x_a)$ (18) and considering the system (1), (2), (4), (5), the dynamics of $x_a(t)$ becomes (19). In (19), $x_a(t)$ will be driven to asymptotically converge to 0, which is proved by using Lyapunov stability theory in the next section.

B. Performance of MAPDAs

MAPDAs with $A_{a,t}(A_n, \Phi, x_a)$ (18) have been designed above and its performance will be presented in Theorem 3.

Theorem 3: Considering the systems (1), (2), (4), (5) under MAPDAs with $A_{a,t}(A_n, \Phi, x_a)$ (18), for $Q > 0$ and $Z > 0$, if $\Phi$ is stable, then

$$\lim_{t \rightarrow \infty} x_a(t) = 0. \quad (20)$$

Proof: Considering the systems (1), (2), (4), (5) under MAPDAs with $A_{a,t}(A_n, \Phi, x_a)$ (18), the dynamic of $x_a(t)$ is (19). Denoting $F_a^d(t) := F_a(t) + A_n - A$, the dynamic of $x_a(t)$ (19) can be re-expressed as

$$\dot{x}_a(t) = \Phi x_a(t) - F_a^d(t)x_{aam}(t). \quad (21)$$

To analyze the stability of $x_a(t)$, a candidate Lyapunov function is chosen as

$$V(t) = x^T_a(t)P x_a(t) + \operatorname{tr}\left((F_a^d(t))^T Z^{-1} F_a^d(t)\right). \quad (22)$$

Taking the derivative of $V(t)$ along $t$ leads to

$$\dot{V}(t) = x^T_a(t)(\Phi^T P + P \Phi)x_a(t) - 2x^T_a(t)P F_a^d(t)x_{aam}(t) + 2\operatorname{tr}\left((F_a^d(t))^T Z^{-1} F_a^d(t)\right). \quad (23)$$

If there are

$$F_a^d(t)^T Z^{-1} = x_{aam}(t)x_a(t)^T P \quad (24)$$

and (18d), i.e., (18b) and (18d) hold true, and we have

$$\dot{V}(t) = -\dot{x}_a^T(t)Q x_a(t) < 0. \quad (25)$$

Equation (25) yields (20). It completes the proof.

Remark 11: For Theorem 3, there possibly exist three types of MAPDAs, i.e., climbing type ($\epsilon < \sup_{aam}\|x_a\|$, $\sup_{aam}\|x_a\|$ represents the upper bound of $\|x_a(t)\|$ under the proposed MAPDAs), peak type ($\epsilon = \sup_{aam}\|x_a\|$), and descending type ($\epsilon > \sup_{aam}\|x_a\|$), which is shown in Fig. 5. It can provide the guideline for the attacker and defender. For the view of the attacker, they can select the proper parameters of MAPDAs for small $\|x_a\|$ (good stealthiness). However, for the view of the defender, it is suggested not to select too big threshold $\epsilon$. This article mainly focuses on the new stealthy MAPDAs method from the perspective of the attacker, so the discussion of these three types is valuable to help to select the proper parameters of MAPDAs.

Remark 12: The selection of parameters $Q$ and $Z$ of the proposed MAPDAs with $A_{a,t}(A_n, \Phi, x_a)$ (18) will affect the dynamics of $x_a(t)$ and $x(t)$, i.e., the upper bound of $\|x_a(t)\|$ and the limit-crossing speed of $\|x(t)\|$. When $Q$ and $Z$ are selected improperly, the upper bound of $\|x_a(t)\|$ could be close to the threshold and with a high limit-crossing speed of $\|x(t)\|$. On the contrary, when $Q$ and $Z$ are selected properly, the upper bound of $\|x_a(t)\|$ could be far less than the threshold and with a low limit-crossing speed of $\|x(t)\|$. These two cases are shown in Section IV. The parameters $Q$ and $Z$ can be selected by...
using some popular methods, such as trial-and-error method, optimization algorithm, and so on.

However, the attackers cannot obtain \( \Phi \) and thus they cannot calculate \( P \) by (18d) and \( F_a(t) \) in (18b), making MAPDAs with \( A_{\alpha,t}(A_n, \Phi, x_a) \) (18) unable to operate. To cope with this problem, the ideal \( A_{\alpha,t}(A_n, \Phi, x_a) \) (18) is slightly regulated into \( A_{\alpha,t}(A_n, \Phi_n, x_a) \)

\[
(18a)-(18c), \quad (26)
\]

\[ -Q = \Phi_n^T P + P \Phi_n \]

where \( \Phi_n := A_n + B_n K_n \) is the nominal part of \( \Phi \).

**Corollary 1:** Considering the systems (1), (2), (4), (5) under MAPDAs with \( A_{\alpha,t}(A_n, \Phi_n, x, x_a) \) (26), for \( Q > 0 \) and \( Z > 0 \), if \( \Phi \) is stable and (18d) holds, then (20) will hold.

**Proof:** The proof is similar as that of Theorem 3, which is thus omitted. \( \square \)

**Remark 13:** Corollary 1 indicates that after \( Q \) has been selected, \( P \) can be calculated by using (26). If the selected \( Q \) and the calculated \( P \) from (26) satisfy (18d), then (20) will hold, i.e., the stealthiness of MAPDAs is achieved. However, it is not easy to obtain exact auxiliary model for the attacker, so (18d) cannot be verified. Therefore, \( Q \) needs to be selected by using some methods as discussed in the above Remark 12.

The problem of PDAs without control inputs against model mismatch is completely solved along the following line (3) \(\rightarrow\) (14) \(\rightarrow\) (18) \(\rightarrow\) (26). First, in spite of the promising performance given in Theorem 1, TPDAs with (3) are denied due to model mismatch. Second, the limitation of TPDAs with (14) against model mismatch is revealed in Theorem 2 from perspective of \( \epsilon \)-stealthiness and \( \xi \)-destructiveness. Then, to deal with the limitation, MAPDAs with (18) are designed by introducing the measurements and an adaptive control method, whose performance is given in Theorem 3. Finally, to further cast off the dependence on exact models, the regulated MAPDAs with (26) are developed. The experimental demonstration will be given in the next section.

**Remark 14:** When the open-loop dynamic of the nominal system is stable (i.e., \( A_n \) is stable), (20) in Theorem 3 will still hold. When (20) holds, \( \|x_{\text{sam}}(t)\| \) could be unbounded or convergent, i.e., \( \|x(t)\| \) could be unbounded or convergent according to \( x_a(t) = x(t) - x_{\text{sam}}(t) \). The analysis is below using the contradiction method.

1) Suppose that both \( x_{\text{sam}}(t) \) and \( x_a(t) \) are convergent. Then, \( F_a(t) \) in (18b) is convergent, and thus \( A - F_a(t) - A_n \) in (19) is bounded. Using (19), it can be obtained that \( x_a(t) \) is convergent. Therefore, there is no contradiction with (20).

2) Suppose that \( \|x_{\text{sam}}(t)\| \) is unbounded and \( x_a(t) \) is convergent. Then, \( F_{\text{ad}}(t) \) in (18b) could be convergent because \( x_a(t)x_{\text{sam}}^T(t) \) is undetermined type, and thus \( A - F_{\text{ad}}(t) - A_n \) in (19) could be convergent. Using (19), we understand that \( x_a(t) \) are convergent. Therefore, there is no contradiction with (20).

Therefore, when the open-loop dynamic of the nominal system is stable, the proposed MAPDAs cannot ensure that the system state is divergent.

**Remark 15:** We will explain why the proposed results can be applied to the practical engineering applications with the disturbance. Specifically, when there exists no disturbance, the attack-free \( x_a(t) \) will arrive at 0. Thus, the attacker aims at driving \( x_a(t) \) to converge to 0. When there exists bounded disturbance, the attack-free \( x_a(t) \) will be bounded with respect to the disturbance. Thus, the attacker aims at driving \( x_a(t) \) to be bounded with respect to the disturbance. Therefore, the proposed MAPDAs can be applied to the practical engineering applications with the disturbance. It is worth noting that the detection threshold \( \epsilon \) in (6) need to be reselected with regard to the disturbance. According to the idea of the proposed MAPDAs method, the stealthiness [i.e., \( \|x_a(t)\| < \epsilon, t \in [0, \infty) \)] will still be guaranteed.

**C. MAPDAs for Digital Systems**

When the considered NCSs is a digital system (e.g., both the sensors and controller are digitized with the sampling periods), the attacker can adopt discrete-time TPDAs and MAPDAs transformed from continuous TPDAs (3) or (14) and MAPDAs (18) as follows.

1) TPDAs can use a discrete-time exact auxiliary model

\[
\begin{align*}
\varepsilon_{\text{sam}}((k+1)h) &= A_h^k \varepsilon_{\text{sam}}(kh) \quad (27a) \\
A(kh) &= x_{\text{sam}}(kh) \quad (27b)
\end{align*}
\]

where \( k = 0, 1, 2, \cdots, A_h := e^{Ah} \) and \( h \) is the sampling period.

2) TPDAs can use a discrete-time nominal auxiliary model

\[
\begin{align*}
\varepsilon_{\text{nm}}((k+1)h) &= A_h^k \varepsilon_{\text{nm}}(kh) \quad (28a) \\
A(kh) &= x_{\text{nm}}(kh) \quad (28b)
\end{align*}
\]

where \( A_h^k := e^{Ah} \).

3) MAPDAs can use a discrete-time adaptive auxiliary model

\[
\begin{align*}
\varepsilon_{\text{sam}}((k+1)h) &= J_n(kh)x_{\text{sam}}(kh) \quad (29a) \\
F_{\text{ad}}((k+1)h) &= F_{\text{ad}}(kh) \\
+ hZP x_{\text{ad}}(kh) x_{\text{sam}}^T(kh) \quad (29b) \\
A(kh) &= x_{\text{sam}}(kh) \quad (29c) \\
\phi &= F_{\text{n}}^T P + P \Phi_n \quad (29d)
\end{align*}
\]

where \( J_n(kh) := e^{(A_n + F_{\text{ad}}(kh))h} \).

The next aim is to analyze the effectiveness of discrete-time MAPDAs (29) (taken as an example) for the digital system. Considering that when the sensors and controller are digitized, the digital system becomes a sampled-data-based hybrid system. For this kind of hybrid system, referring to [35] and [36], it is commonly expressed as a time-delay system and can be expressed as the following time-delay model:

\[
(1), \ (4), \ \text{and} \ u(t) = K x_a(t - t_d) \quad (30)
\]

where \( t \in [kh, (k+1)h) \), \( k = 0, 1, 2, \cdots \), and \( t_d \in [0, h) \) is time delay and \( t_d = 1. \) Based on the above time-delay model (30), the stability criterion on delay-induced continuous MAPDAs to guarantee (20) will be presented in Theorem 4.
Theorem 4: Considering the constructed time-delay model (1), (4), and (30), for a scalar \( h \), the matrices \( P_1 > 0 \), \( P_2 > 0 \), \( P_3 > 0 \), \( P_4 > 0 \), and \( Z_1 > 0 \), if there are delay-induced continuous MAPDAs

\[
\dot{x}_{aam}(t) = (A_n + F_d(t))x_{aam}(t) \\
\dot{F}_d(t) = Z_1P_1x_{aam}(t)x_{aam}^T(t) \\
+ h^2Z_1P_4(Ax_a(t) + BKx_a(t - \tau))x_{aam}(t) \\
- \frac{1}{2}h^2Z_1P_4F_d(t)x_{aam}(t)x_{aam}^T(t) \\
a(t) = x_{aam}(t) \\
- Q = \Phi^TP + P\Phi.
\]

(31a)
(31b)
(31c)
(31d)

\( F_d(t) \) have been given in (21), and

\[
\Omega = \begin{bmatrix}
\Omega_{11} & \Omega_{12} & \Omega_{13} \\
* & \Omega_{22} & \Omega_{23} \\
* & * & \Omega_{33}
\end{bmatrix} < 0
\]

(32)

where \( \Omega_{11} = A^TP_1 + P_1A + P_2 + P_3 + h^2A^TP_4A + P_4 \), \( \Omega_{22} = -P_2 + h^2K^TB^TP_4BK \), \( \Omega_{33} = -P_3 - P_4 \), \( \Omega_{12} = P_1BK + h^2A^TP_4BK \), \( \Omega_{13} = P_4 \), and \( \Omega_{23} = 0 \), then (20) will hold.

Proof: From (30), the dynamics of \( x_a(t) \) becomes

\[
\dot{x}_a(t) = Ax_a(t) + BKx_a(t - \tau) - F_d(t)x_{aam}(t).
\]

(33)

To analyze the stability of \( x_a(t) \), we choose the candidate Lyapunov function

\[
V_{lds}(t) = x_a^T(t)P_1x_a(t) + \int_{t-r}^{t} x_a^T(s)P_2x_a(s)ds \\
+ \int_{t-r}^{t} x_a^T(s)P_3x_a(s)ds \\
+ \int_{t-r}^{t} h^2x_a^T(s)P_4x_a(s)dsdv \\
+ tr\left(\left(F_d(t)\right)^TZ^{-1}_1F_d(t)\right).
\]

(34)

Taking the derivative of \( V_{lds}(t) \) along with \( t \), denoting \( \chi_a(t) = \{x_a(t); x_a(t - \tau); x_a(t - h)\} \) and using Jensen’s inequality [38], [39], [40] leads to

\[
\dot{V}_{lds}(t) \leq x_a^T(t)\Omega x_a(t) \\
+ tr\left(\left(F_d(t)\right)^TZ^{-1}_1F_d(t)\right) \\
- \chi_a^T(t)P_1\chi_a(t) \\
- h^2\chi_a^T(t)x_{aam}(t)P_4(Ax_a(t) + BKx_a(t - \tau)) \\
+ \frac{1}{2}h^2\chi_a^T(t)x_{aam}(t)P_4F_d(t)x_{aam}(t) \\
+ tr\left(\left(F_d(t)\right)^TZ^{-1}_1F_d(t)\right) \\
- x_a^T(t)P_1F_d(t)x_{aam}(t) \\
- h^2(Ax_a(t) + BKx_a(t - \tau))P_4F_d(t)x_{aam}(t) \\
+ \frac{1}{2}h^2F_d(t)x_{aam}(t)P_4F_d(t)x_{aam}(t).
\]

(35)

If there are (31) and (32), then

\[
\dot{V}_{lds}(t) \leq x_a^T(t)\Omega x_a(t) < 0.
\]

(36)

Equation (36) yields (20). It completes the proof.

Remark 16: According to (31) in Theorem 4, a delay-induced discrete-time MAPDAs (37) can be obtained as

\[
x_{aam}(k+1) = \Xi_a(kh)x_{aam}(kh)
\]

(37a)

\[
F_a(k+1) = F_a(kh)
\]

(37b)

\[
+ hZ_1P_1x_a(kh)x_{aam}(kh) \\
+ h^2Z_1P_4x_a(kh)x_{aam}(kh) \\
- \frac{1}{2}h^2Z_1P_4F_a(kh)x_{aam}(kh)
\]

(37c)

\[
- Q = \Phi^TP + P\Phi_a.
\]

(37d)

Note that there exists only one difference between discrete-time MAPDAs (29) and delay-induced discrete-time MPADAs (37), i.e., the last three items related to the square of the sampling period are additional in (37b). When the sampling period is small, discrete-time MAPDAs (29) are a proper approximation of delay-induced discrete-time MAPDAs (37). Since discrete-time MPADAs (37) can guarantee the stealthiness (20), discrete-time MAPDAs (29) will be effective on guaranteeing the stealthiness (20) for digital system.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To validate the proposed MAPDAs, we consider the scenario when TPDAs [15], MAPDAs, DFLCAs [16], and TLCAs [17] are embedded in an NIPVSS [37] in Fig. 6.

A. Experimental Setup of NIPVSS

In the real NIPVSS, a single Ac640-120cm monochrome camera at 640 × 480 pixels with light sources acting as the sensor captures the images of GLIP2001C linear motion inverted pendulum. Then, the images are sent to the controller through the 1-Gb/s wired network (It means that transmission delay can be omitted), which runs Microsoft Visual Studio 2010 in Windows XP with an Intel Core i5 processor (3.2 GHz) and 4-GB RAM. After having received images, the controller processes the images to obtain state information of cart position and pendulum angle. Furthermore, state information is used to formulate the control signal in the controller with GT-400-SV motion control card. Finally, the control signal is...
TABLE III

|          | $\sup_{t}^\epsilon \|x_a\|$ | $\sup_{t}^\epsilon \|x_a\|$ |
|----------|-----------------------------|-----------------------------|
| Experiment 1 | 2.2544                     | Experiment 4                |
| Experiment 2 | 3.0560                     | Experiment 5                |
| Experiment 3 | 2.2544                     | Experiment 6                |

applied to the inverted pendulum by use of MSMA023A1A AC servo motor driver serving as the actuator.

B. Parameter Settings of NIPVSS

The state of NIPVSS is set as $x(t) = [\alpha(t), \theta(t), \dot{\alpha}(t), \dot{\theta}(t)]$, where $\alpha(t)$ is cart position, $\theta(t)$ is pendulum angle, and $\dot{\alpha}(t)$ and $\dot{\theta}(t)$ are cart and angular velocity, respectively. The acceleration-as-control-input nonlinear differential equation of inverted pendulum is

$$lmu \cos \theta + lmg \sin \theta = Jm \ddot{\theta}$$

where $l$ is the length from the pivot to the center of the pendulum, $m$ is the mass of the pendulum, $g$ is the acceleration of the gravity, $Jm$ is the moment of inertia about the pivot of the pendulum, and the values of $l$, $m$, $g$, and $Jm$ can be found in [37]. By linearizing (38) in $|\theta| \leq 0.2$ rad (i.e., $\cos \theta \approx 1$ and $\sin \theta \approx \theta$ in $|\theta| \leq 0.2$ rad), the nominal models $A_n$ and $B_n$ of (38) are given by

$$A_n = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{lmg}{Jm} & 0 & 0 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \frac{lm}{Jm} \end{bmatrix}$$

where $(lmg/Jm) = 29.4311$ and $(lm/Jm) = 3.0001$. $u = \ddot{\alpha}$ is control input. According to $A_n$ and $B_n$, the controller is designed as (5), where the controller gain has been calculated in [37], i.e.,

$$K = K_n = [3.7569, -29.6225, 4.0648, -5.4563]$$

The controlled outputs are $\alpha(t)$ and $\theta(t)$, and the aim of controller is to drive the pendulum being stable. The admissible limits are $|\alpha(t)| < 0.3$ m and $|\theta(t)| < 0.8$ rad.

C. Threshold of the Detector

To properly determine the threshold $\epsilon$ of the detector based on statistical analysis method [27], 500 experiments of attack-free NIPVSS are operated (see 6 experiments of 500 experiments in Fig. 7 and Table III). The frequency of different $\sup_{t}^\epsilon \|x_a\|$ (i.e., the upper bound of attack-free $\|x_a(t)\|$) is shown in Fig. 8. It can be seen from Fig. 7, Table III, and Fig. 8 that after the state of NIPVSS is stable, the threshold can be set as $\epsilon = 3.1$ based on the $3\sigma$ principle.

D. Performance of TPDAs and MAPDAs

For comparison, we construct two types of attacks with the values of $A_n$, $B_n$, and $K_n$. One is TPDAs [15] (28) when $t \geq 0$ s. Another is the proposed MAPDAs (29) when $t \geq 0$ s, and their parameters are set as $Q = I$ and $Z = 10000I$ in (26), and $P$ is calculated by using (26) as

$$P = \begin{bmatrix} 1.7760 & -2.0855 & 0.8362 & -0.3231 \\ 10.6948 & -2.9413 & 1.4742 \\ * & 1.6562 & -0.4646 \\ * & * & 0.2755 \end{bmatrix}$$
DU et al.: STEALTHY MAPDAs WITH NOMINAL MODELS

**Fig. 9.** Controlled output and detection results of NIPVSS under TPDAs\([15]\) (28) or the proposed MAPDAs (29) of \(Q = I\) and \(Z = 10000I\). Blue line: Under TPDAs (28). Red line: Under MAPDAs (29). Black line in (c): The detection threshold \(\epsilon = 3.1\). (a) \(\alpha(t)\). (b) \(\theta(t)\). (c) \(\|x_a(t)\|\).

The initial condition is set to \(x_{nam}(0) = 0.00011\) in (14), and \(x_{am}(0) = 0.00011\), \(F_a(0) = I\) in (28). Considering \(A_n = X_nJ_nX_n^{-1}\) where

\[
X_n = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 0.5 & 0.5 \\
0 & 1 & 0 & 0 \\
0 & 0 & -2.7125 & 2.7125
\end{bmatrix}
\]

\[
J_n = \begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & -5.425 & 0 \\
0 & 0 & 0 & 5.425
\end{bmatrix}
\]

the eigenvalues of \(J_n\) are \(\lambda_{n,1} = \lambda_{n,2} = 0\) (2-fold root), \(\lambda_{n,3} = -5.425\) (single root), and \(\lambda_{n,4} = 5.425\) (single root). Considering \(\psi_n(0) = X_n^{-1}x_{nam}(0) = [0.1; 0.1; 0.0816; 0.1184] \times 10^{-3}\), all the elements of \(\psi_n(0)\) are not zero, which does not satisfy the item 1) or 2) of Lemma 2. Therefore, it satisfies the initial condition of Theorem 2.

The experiments are operated using the above set parameters, and the experimental results of NIPVSS under TPDAs (28) and the proposed MAPDAs (29) are shown in Fig. 9. When there exists model mismatch, we observe that: 1) TPDAs (28) drive the pendulum angle of NIPVSS to cross the maximum allowable angle 0.8 rad shown as the blue line in Fig. 9(b), while the detector succeeds to detect them shown as the blue line in Fig. 9(c) and 2) MAPDAs (29) cannot be detected shown as the red line in Fig. 9(c) before driving the cart position to cross the allowable limit -0.3 m shown as red line in Fig. 9(a). It confirms that the proposed MAPDAs are not detectable before achieving successful destructiveness.

The above has presented rather good results for the proper parameters \(Q\) and \(Z\), however if the improper parameters \(Q\) and \(Z\) are chosen, it will produce less satisfactory experimental results. For example, when \(Q = I\) and \(Z = 0.5I\), Figs. 10 and 11 show the experiments of NIPVSS under TPDAs (28) when \(t \geq 6\) s or the proposed MAPDAs (29) when \(t \geq 6\) s. It can be seen from Figs. 10 and 11 that the upper bound of \(\|x_a(t)\|\) will almost touch the threshold \((\epsilon = 1.6\) from five attack-free experiments) and the limit-crossing speed of \(\|x(t)\|\) is high (as shown by the result that the blue lines touch the threshold later than red lines). It indicates that the parameters \(Q\) and \(Z\) are improperly chosen.

---

**Fig. 10.** Controlled output and detection results of NIPVSS under TPDAs (28) or the proposed MAPDAs (29) of \(Q = I\) and \(Z = 0.5I\). Blue line: Under TPDAs (28). Red line: Under MAPDAs (29). Green line: Instant when attacks start. Black line in (c): The detection threshold \(\epsilon = 1.6\). (a) \(\alpha(t)\). (b) \(\theta(t)\). (c) \(\|x_a(t)\|\).

**Fig. 11.** Detection results when \(t \geq 6\) s of Fig. 10(c). Red line: Under MAPDAs (29). Black line: The set detection threshold \(\epsilon = 1.6\).
TABLE IV
QUANTITATIVE COMPARISON AMONG MAPDAS, TPDAS WITH MODEL MISMATCH, DFLCAS [16], AND TLCAS [17]

|                  | MAPDAS | TPDAs with model mismatch | DFLCAS [16] | TLCAS [17] | TLCAs without using control signal |
|------------------|--------|---------------------------|-------------|------------|----------------------------------|
| ATI              | [0s, 1.82s] | [0s, 0.88s]                     | [0s, 0.72s] | [0s, 0.44s] | [0s, 0.44s]                        |
| I                | 1.82s  | 0.88s                     | 0.72s       | 0.44s      | 0.44s                             |
| N                | 0s     | 0.01s                     | 0.08s       | 0s         | 0.14s                             |
| Stealthiness: β | 0%     | 1.14%                     | 11.11%      | 0%         | 31.82%                            |
| Destructiveness  | | | | | |
| | |                          |                          | [∥xa(t)∥ > ϵ] > 0.3m, | [∥xa(t)∥ > ϵ] > 0.3m, | |
| | |                          |                          | [∥θ(t)∥ > 0.8 rad] > 0.8 rad, | [∥θ(t)∥ > 0.8 rad] > 0.8 rad, | |
| | |                          |                          | [∥xa(t)∥ > ϵ] > 0.3m, | [∥xa(t)∥ > ϵ] > 0.3m, | |
| | |                          |                          | [∥θ(t)∥ > 0.8 rad] > 0.8 rad, | [∥θ(t)∥ > 0.8 rad] > 0.8 rad, | |

Fig. 12. Controlled output and detection results of NIPVSS under DFLCAS [16], TLCAS [17], and the proposed MAPDAs (29). Purple line: Under DFLCAS. Green line: Under TLCAS. Blue line in (c): Detection result of TLCAs without using control input. Red line: Under MAPDAs (29). Black line in (c): The detection threshold $\epsilon = 3.1$. (a) $\alpha(t)$. (b) $\theta(t)$. (c) $\|x_a(t)\|$. (d) $\|\theta(t)\|$. (e) $\|x_a(t)\|$. (f) $\|\theta(t)\|$. (g) $\|x_a(t)\|$.

E. Performance of MAPDAs, DFLCAS, and TLCAS

For further comparison with the existing methods, three types of attacks are constructed: 1) DFLCAS [16] using both measurements and control input; 2) TLCAS [17] using both measurements and control input (or using only the measurements); and 3) the proposed MAPDAs using only measurements.

The experimental results of NIPVSS under DFLCAS, TLCAS and MAPDAs (29) are shown in Fig. 12. When there exists model mismatch, it is observed that: 1) DFLCAS drive the pendulum angle and cart position of NIPVSS to cross the allowable limit $-0.8$ rad and $-0.3$ m shown as the purple lines in Fig. 12(a) and (b), respectively, while the detector succeeds to detect them shown by the purple line in Fig. 12(c); 2) TLCAS drive the pendulum angle and cart position of NIPVSS to cross the allowable limit $-0.8$ rad and $-0.3$ m shown as the green lines in Fig. 12(a) and (b), respectively, while the detector fails to detect them shown as the green line in Fig. 12(c). It can be also seen from the blue line in Fig. 12(c) that once TLCAs do not use control input (i.e., cannot construct the covert agent), they will be detected by the detector; and 3) MAPDAs (29) drive the cart position of NIPVSS to cross the allowable limit $-0.3$ m shown as the red line in Fig. 12(a), while the detector fails to detect them shown as the red line in Fig. 12(c). Therefore, compared with DFLCAS and TLCAS, the proposed MAPDAs method using only the measurements can bypass the detector, i.e., achieve successfully stealthy attack.

A Thorough Quantitative Analysis of Experimental Results: From Figs. 9 and 12, a detailed comparison among MAPDAs, TPDAs with model mismatch and the existing attack methods (e.g., DFLCAS [16] and TLCAS [17]) supported by quantitative metrics is given in Table IV. To further analyze stealthiness and destructiveness, the detection rate $\beta$ is defined as $\beta = \frac{N}{I} \times 100\%$, where $I$ is the length of attack time interval (ATI), $N \in \mathbb{Z}$ is total time when $\alpha(t) > 0.3m$, $\theta(t) > 0.8 rad$. Recall the admissible limits $\alpha(t), \theta(t)$. It can be clearly seen from Table IV that: 1) maintaining the destructiveness, MAPDAs and TLCAs have the best stealthiness (i.e., $\beta = 0$) and 2) however, when without using control signal, TLCAs have the worst stealthiness ($\beta = 31.82\%$). Therefore, the proposed MAPDAs’ advantages over traditional methods have been clearly demonstrated.

V. CONCLUSION

We have shown in this article that stealthy attacks on vulnerable-sensor-network-only NCSs are possible, particularly when the measurements and adaptive control methods are employed. As a prototype, a stealthy MAPDAs method has been designed by using the measurements and model reference adaptive control methods, which is easier to implement in comparison with the existing methods requiring both the measurements and control inputs. Its promising performance on stealthiness and destructiveness are analyzed by using convergence of measurements.

In the future, our work can be further extended to control system with partial measurement (i.e., $y(t) = Hx(t) \in \mathbb{R}^n$, $H \neq I, n_y < p$). In such control system, the controller common employs more complex control structure, e.g., dynamic output feedback or static state feedback with Luenberger observer, etc. Thus, it is also an interesting work to apply adaptive control strategy to TPDAs with model mismatch for guarantee attack stealthiness and destructiveness.

On the other hand, MAPDAs adopt plant state in adaptive auxiliary model so that MAPDAs can achieve its stealthiness, hence system security is destroyed. To improve system...
security, it is necessary to use the probing signals (e.g., water-marking [37]) into plant state for hindering the stealthiness of MAPDAs.

REFERENCES

[1] J. Tao et al., “Event-triggered control for Markov jump systems subject to mismatched modes and strict dissipativity,” IEEE Trans. Cybern., vol. 53, no. 3, pp. 1537–1546, Mar. 2023.
[2] L. Zou, Z. Wang, H. Dong, X. Yi, and Q.-L. Han, “Recurrent filtering under probabilistic encoding-decoding schemes: Handling randomly occurring measurement outliers,” IEEE Trans. Cybern., vol. 54, no. 6, pp. 3378–3391, Jun. 2024.
[3] W. Chen, Z. Wang, D. Ding, X. Yi, and Q.-L. Han, “Distributed state estimation over wireless sensor networks with energy harvesting sensors,” IEEE Trans. Cybern., vol. 53, no. 5, pp. 3311–3324, May 2023.
[4] H. Ahn, J. Choi, and Y. H. Kim, “A mathematical modeling of Stuxnet-style autonomous vehicle malware,” IEEE Trans. Intell. Transp. Syst., vol. 24, no. 1, pp. 673–683, Jan. 2023.
[5] M. Cook, A. Marnerides, C. Johnson, and D. Pezaros, “A survey on industrial control system digital forensics: Challenges, advances and future directions,” IEEE Commun. Surveys Tuts., vol. 25, no. 3, pp. 1705–1747, 3rd Quart., 2023.
[6] M. Liu, C. Zhao, J. Xia, R. Deng, P. Cheng, and J. Chen, “PDLPD: Proactive distributed detection and localization against stealthy deception attacks in DC microgrids,” IEEE Trans. Smart Grid, vol. 14, no. 1, pp. 714–731, Jan. 2023.
[7] R. Kumar, R. Kela, S. Singh, and R. Trujillo-Rasua, “APT attacks on industrial control systems: A tale of three incidents,” Int. J. Crit. Infrastruct. Prot., vol. 37, Jul. 2022, Art. no. 100521.
[8] M.D. Firoozjaei, N. Mahmoudyar, Y. Baseri, and A. A. Ghorbani, “An evaluation framework for industrial control system cyber incidents,” Int. J. Crit. Infrastruct. Prot., vol. 36, Mar. 2022, Art. no. 100487.
[9] A. Naha, A. Teixeira, A. Ahlén, and S. Dey, “Sequential detection of replay attacks,” IEEE Trans. Autom. Control, vol. 68, no. 3, pp. 1941–1948, Mar. 2023.
[10] K. Jin and D. Ye, “Optimal innovation-based stealth attacks in networked LQG systems with attack cost,” IEEE Trans. Cybern., vol. 54, no. 2, pp. 787–796, Feb. 2024.
[11] X. Lei, G. Wen, W. X. Zheng, and J. Fu, “Security strategy against location-varying sparse attack on distributed state monitoring,” IEEE Trans. Autom. Control, vol. 69, no. 4, pp. 2514–2521, Apr. 2024.
[12] H. Song, P. Shi, C.-C. Lim, W.-A. Zhang, and L. Yu, “Attack and estimation design for multi-sensor systems with undetectable adversary,” Automatica, vol. 109, Nov. 2019, Art. no. 108545.
[13] D. Mikhaileyenko and P. Zhang, “Stealthy local covert attacks on cyber-physical systems,” IEEE Trans. Autom. Control, vol. 67, no. 12, pp. 6778–6785, Dec. 2022.
[14] K. Kimura and H. Ishii, “Quantized zero dynamics attacks against sampled-data control systems,” IEEE Trans. Autom. Control, vol. 69, no. 5, pp. 3418–3425, May 2024.
[15] S. Kim, K.-J. Park, and C. Lu, “A survey on network security for cyber-physical systems: From threats to resilient design,” IEEE Commun. Surveys Tuts., vol. 24, no. 3, pp. 1534–1573, 3rd Quart., 2022.
[16] Z. Li and G.-H. Yang, “A data-driven covert attack strategy in the closed-loop cyber-physical systems,” J. Frankl. Inst., vol. 355, no. 14, pp. 6645–6668, 2018.
[17] W. Li, L. Xie, and Z. Wang, “Two-loop covert attacks against constant value control of industrial control systems,” IEEE Trans. Ind. Informat., vol. 15, no. 2, pp. 663–676, Feb. 2019.
[18] G. Park, C. Lee, H. Shim, Y. Eun, and K.H. Johansson, “Stealthy adversaries against uncertain cyber-physical systems: Threat of robust zero-dynamics attack,” IEEE Trans. Autom. Control, vol. 64, no. 12, pp. 4907–4919, Dec. 2019.
[19] H. Jeon and Y. Eun, “A stealthy sensor attack for uncertain cyber-physical systems,” IEEE Internet Things J., vol. 6, no. 4, pp. 6345–6352, Aug. 2019.
[20] X. Wen et al., “Age-of-task-aware co-design of sampling, scheduling, and control for industrial IoT systems,” IEEE Internet Things J., vol. 11, no. 10, pp. 4227–4232, Feb. 2024.
[21] B. Babayigit and M. Abubaker, “Industrial Internet of Things: A review of improvements Over traditional SCADA systems for industrial automation,” IEEE Syst. J., vol. 18, no. 1, pp. 120–133, Mar. 2024.
Changda Zhang received the B.Sc. degree in electrical engineering and automation and the M.Sc. and Ph.D. degrees in control theory and control engineering from Shanghai University, Shanghai, China, in 2016, 2019, and 2023, respectively. He is currently a Lecturer with the Hebei University of Science and Technology, Shijiazhuang, China. His main research interests include secure control for networked control systems.

Chen Peng (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Chinese University of Mining Technology, Xuzhou, China, in 1999 and 2002, respectively. From September 2002 to August 2004, he was a Postdoctoral Research Fellow of Applied Mathematics with Nanjing Normal University, Nanjing, China. From November 2004 to January 2005, he was a Research Associate with Hong Kong University, Hong Kong, China. From July 2006 to August 2007, he was a Visiting Scholar with the Queensland University of Technology, Brisbane, QLD, Australia. From September 2010 to August 2012, he was a Postdoctoral Research Fellow with Central Queensland University, Rockhampton, QLD, Australia. He is currently a Professor with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China. His current research interests include networked control systems, intelligent control, complex system modeling, hybrid network systems, and field control systems.

Minrui Fei received the B.S. and M.S. degrees in industrial automation from the Shanghai University of Technology, Shanghai, China, in 1984 and 1992, respectively, and the Ph.D. degree in control theory and control engineering from Shanghai University, Shanghai, in 1997. Since 1998, he has been a Full Professor with Shanghai University. His research interests include networked control systems, intelligent control, complex system modeling, hybrid network systems, and field control systems.

Dr. Fei is the Chairman of the Embedded Instrument and System Subsociety, the Standing Director of the China Instrument and Control Society, the Chairman of the Life System Modeling and Simulation Subsociety, the Vice-Chairman of the Intelligent Control and Intelligent Management Subsociety, and the Director of the Chinese Artificial Intelligence Association.

Huiyu Zhou received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.Sc. degree in biomedical engineering from the University of Dundee, Dundee, U.K., in 2002, and the D.Phi. degree in computer vision from Heriot–Watt University, Edinburgh, U.K., in 2006. He is currently a Full Professor with the School of Computing and Mathematical Sciences, University of Leicester, Leicester, U.K. He has authored or co-authored over 350 peer-reviewed papers in the field. His research work has been or is being supported by U.K. EPSRC, MRC, AHRC, ESRC, EU ICT, Royal Society, Innovate U.K., Leverhulme Trust, Invest NI, Puffin Trust, Alzheimer’s Research U.K., and industry.