A novel resilient control scheme for a class of Markovian jump systems with partially unknown information

Zhang, Kun; Su, Rong; Zhang, Huaguang

2021

Zhang, K., Su, R. & Zhang, H. (2021). A novel resilient control scheme for a class of Markovian jump systems with partially unknown information. IEEE Transactions On Cybernetics. <https://dx.doi.org/10.1109/TCYB.2021.3050619>

<https://hdl.handle.net/10356/152246>

<https://doi.org/10.1109/TCYB.2021.3050619>

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: <https://doi.org/10.1109/TCYB.2021.3050619>.

Downloaded on 26 Apr 2023 18:38:42 SGT

A Novel Resilient Control Scheme for a Class of Markovian Jump Systems with Partially Unknown Information

Kun Zhang, *Member, IEEE*, Rong Su, *Senior Member, IEEE*, and Huaguang Zhang, *Fellow, IEEE*

Abstract—In the complex practical engineering systems, many interferences and attacking signals are inevitable in industrial applications. This paper investigates the reinforcement learning (RL) based resilient control algorithm for a class of Markovian jump systems with completely unknown transition probability information. Based on the Takagi-Sugeno logical structure, the resilient control problem of nonlinear Markovian systems is converted into solving a set of local dynamic games, where the control policy and attacking signal are considered as two rival players. Combining the potential learning and forecasting abilities, the new integral RL (IRL) algorithm is designed via system data to compute the zero-sum games without using the information of stationary transition probability. Besides, the matrices of system dynamics can also be partially unknown, and the new architecture requires less transmission and computation during the learning process. The stochastic stability of the system dynamics under the developed overall resilient control is guaranteed based on Lyapunov theory. Finally, the designed IRL based resilient control is applied to a typical multi-mode robot arm system, and implementing results demonstrate the practicality and effectiveness.

Index Terms—Resilient control, integral reinforcement learning (IRL), zero-sum game, adaptive dynamic programming.

I. INTRODUCTION

IN control and many other engineering fields, there has been a large growing interest in computational intelligence technologies and implementations during these past few years. In particular, the artificial intelligent (AI) methods based on the so-called reinforcement learning (RL) technique have rapidly attracted the scholars' and engineers' attention, since the "AlphaGo" defeated champion of human beings in man-machine games [1]–[3]. Combining the abilities of learning and prediction, RL methods show the superiority in dealing with kinds of problems under incomplete information. Optimal control problem can be abstracted to be a partial differential

equation, algebraic Riccati equation (ARE) for some linear systems or complex Hamilton-Jacobi-Bellman (HJB) equation for a slew of nonlinear systems, which is hardly solved by mathematical methods directly [4]–[6]. In recent research, the integral RL (IRL) methods or adaptive/approximate dynamic programming (ADP) techniques are proposed in the literatures [7]–[9] to solve the optimal control policy by scholars. As we know, although the solutions above are approximately obtained for a fixed numerical system, operational plants in engineering applications unavoidably undergo some stochastic variations in system parameters or structures resulting from several different failures, sudden environmental change and noises, etc. Unfortunately, for the Markovian jump dynamics, let alone general stochastic systems, the overall, even local optimal solutions are difficult to obtain, if it is possible.

It should be mentioned that the Takagi-Sugeno (T-S) logical structure [10]–[12], which has been widely applied to analyze the complicated Markov jump models with stochastic nonlinear dynamics in industries [13]–[15]. For robust stabilization of nonlinear Markovian jump systems, a mode-independent controller was proposed by using the linear matrix inequalities in [16]. With the aim to be stochastic stability, the robust control was designed in [17] subject to Markovian jump parameters of nonlinear systems, where the H_∞ performance was considered. Based on the logical models, [18] proposed an adaptive output-feedback control approach to stochastic stabilize uncertain strict-feedback systems, which combined the dynamic surface technique with an observer design. Based on the RL technique, [19] attempted to solve the tracking control with partially unknown information in an interconnected system, which provides a potential superiority of RL in dealing with some complex systems. Then, to improve the efficiency at each transmission and sampling instance, the event-based control problem for complex stochastic systems is investigated in [20]. However, the information of transition probability and system dynamics are necessary in these researches, and almost constantly transmitting is required in the solving process.

Besides, against the cyber-environment and cyber-physical system attacks in industrial applications, some security and resilient control methods are provided to resist the negative effects [23]–[26]. In networked control systems (NCSs), denial-of-service (DoS) attacks is induced by radio-interference signals (also referred to as jamming signals), and causes the inaccurate transmission or information, even all legitimate requests from being fulfilled, for example, in the topology networks of multiple wheeled mobile robots and other multi-

This work was supported by the National Postdoctoral Program for Innovative Talents under Grant BX20200357, the China Postdoctoral Science Foundation under Grant 2020M680718, the Singapore National Research Foundation Delta-NTU Corporate Lab Program (DELTA-NTU CORP-SMA-RP2), and the Singapore Ministry of Education Tier 1 Academic Research Grant 2013-T1-002-177.

K. Zhang is with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: nukgnahz@163.com).

R. Su is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore (e-mail: rsu@ntu.edu.sg).

H. Zhang is with the State Key Laboratory of Synthetical Automation for Process Industries, and also with the School of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: hgzhzhang@ieee.org).

agent systems [27]–[33]. Some resilient optimal and guaranteed cost control methods were proposed to overcome the perturbations and attacks in [34]–[37], and the asymptotically stability was guaranteed in [38] by a new mode-dependent resilient control policy, where a weak sufficient condition was developed. A novel distributed filter was first constructed [39] to practically reflect the impact from both cyber-attacks and gain perturbations in power systems, then, [40] designed a recursive distributed set-membership scheme to fulfill the scalability requirement of the sensor networks. Moreover, resilient control matching H_∞ performance was discussed in detail in [41], where the control protocols guaranteed bounded L_2 gains of certain errors in terms of the L_2 norms of fault signals. By using the infinite horizon performance index, a zero-sum game with two players is equal to finding the saddle point solution from a saddle surface of policies, which has been explored by [42]–[45]. Thus, as pointed out in [23], solving the zero-sum game is a way for both cyber defenders and DoS attackers. However, the physical setting and engineering environment are complicated and changeable in practical applications, and the dynamics under attacks have random components, which requires lots of information unavailable in real time.

To solve the resilient control problem of a nonlinear Markovian jump system (MJS) with completely unknown transition probability, and motivated by the potentiality of RL techniques, a novel data based IRL control algorithm is designed in this paper. The main contributions can be summarized in the follows: **(1)** A resilient control scheme is designed in the Markov process by solving a set of zero-sum games, where we consider the unknown signal as a control policy from the hostile player, who aims to damage the system performances with its maximum destructive capability. **(2)** Compared with some existing methods in controlling the nonlinear MJSs [16]–[19], [38], the transition probability information of a homogeneous Markov process is not needed in the novel IRL algorithm. Besides, the dynamic information can also be partially unknown during the solving process. **(3)** The developed IRL control technique is data driven, and aims for achieving the stochastic stability in a resilient control problem, where the attacking signals from different components occur simultaneously.

The architecture of this work can be organized as follows. The main idea is inspired by how to use the RL technique and system data to solve the resilient control problem, as presented in the aforementioned introduction of Section I. The problem formulation of the resilient control in nonlinear MJSs is shown in Section II. The mode-dependent control policies for zero-sum games are developed in Section III, where the problem is transformed to solve the coupled stochastic algebraic Riccati equations (AREs). Section IV proposes the novel IRL based control algorithm for MJSs, and both the stationary transition probability and partial system dynamics can be avoided during the iterative learning process. A typical multi-mode robot arm system is implemented to verify the designed IRL based algorithm in Section V, where practicality and effectiveness of the new control scheme are demonstrated. Section VI concludes the main results of the work.

II. PRELIMINARIES

A. The system under DoS attacks

In this paper, let us consider a kind of partially unknown nonlinear Markovian jump systems in some probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where the sample space is Ω , the σ -algebra is $\mathcal{F}$, and $\mathcal{P}$ indicates the probability measure defined on $\mathcal{F}$ [42]. And the system dynamics suffer some uncertain attacks during cyber-networked control processes, which can be described by the differential equation:

$$\dot{x}(t) = f(x, \eta_t) + g(x, \eta_t)\mu(t) + \varsigma(t, \eta_t) \quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the state vector; $f(x, \eta_t)$ and $g(x, \eta_t)$ denote the smooth dynamic functions with $f(0, \eta_t) = 0$ and $g(0, \eta_t) = 0$; the system mode $\eta_t = l$ is generated from a continuous time Markov process with discrete state, which takes values in the finite mode set $\mathcal{S} \triangleq \{1, \dots, s\}$ and the set $\mathcal{S}$ consists all the operation modes from the system; $\tilde{f}(x, \eta_t)$ and $\tilde{g}(x, \eta_t)$ are the uncertain signals on the dynamic functions $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$, and $\varsigma(t, \eta_t) = \tilde{f}(x, \eta_t) + \tilde{g}(x, \eta_t)\mu(t)$ is the uncertain system attacking signal; $\mu(t)$ is the input vector, which also suffers the attack from external signals.

The transition probability of the homogeneous Markovian process from the mode $\eta_t = l$ to a mode $\eta_{t+\Delta t} = k$, with $l, k \in \mathcal{S}$ is described as

$$P_r\{\eta_{t+\Delta t} = k | \eta_t = l\} = \begin{cases} \pi_{kl}\Delta t + o(\Delta t), & l \neq k \\ 1 - \pi_{ll}\Delta t + o(\Delta t), & l = k \end{cases} \quad (2)$$

where $\Delta t > 0$ is a small time constant, $o(\cdot)$ is the higher order infinitesimal operator, i.e., $\lim_{\Delta t \rightarrow 0} (\frac{o(\Delta t)}{\Delta t}) = 0$, $\pi_{kl} \geq 0$ is the transition probability between modes l and k for $l \neq k$, and for all $l \in \mathcal{S}$, $\pi_{ll} := -\pi_{ll} = -\sum_{k=1, k \neq l}^s \pi_{lk}$. The matrix $\Pi = [\pi_{lk}]_{s \times s}$ is called the transition-probability-rate matrix.

To better process the multi-mode complex dynamic, let $z(t)$ be a mode indicator variable, η_0 be the initial mode and $\mathbb{M}_l$ be a crisp set, thus, the Markov jump nonlinear system can be represented by the logical model [17], [38]. This system is established by the crisp sets $\mathbb{M}_l$ related to the jumps in the continuous Markov process, which ranges in a lower and upper levels. The system mode is identified by $z(t)$ and dynamic is structured by the judging sets $M_{l,ij}$ resulting from system state $x(t)$, then, the l th mode of the considered system (1) is further described as follows:

Mode l :

IF $z(t)$ is $\mathbb{M}_l$

THEN

(3)

Plant Rule i :

IF $\xi_1(t)$ is $M_{l,i1}$, $\dots$, $\xi_{r_{li}}(t)$ is $M_{l,ir_{li}}$,

THEN $\dot{x}(t) = (A_{l,i} + \tilde{A}_{l,i}(t))x(t) + (B_{l,i} + \tilde{B}_{l,i}(t))\mu(t)$

where $i \in \mathcal{I} \triangleq \{1, \dots, r_l\}$ and the number with scalar of IF – THEN rules is r_l in l th mode, $M_{l,ij}$ is the judging set with $i \in \mathcal{I}, j \in \{1, \dots, r_{li}\}$; the premise variables $z(t)$ and $\xi_1(t), \dots, \xi_{r_{li}}(t)$ are given here or to be only a function

of system parameters and states; $A_{l,i}$ and $B_{l,i}$ represent the real dynamic matrices, where $A_{l,i}$ is unknown; $\tilde{A}_{l,i}(t)x(t)$ and $\tilde{B}_{l,i}(t)\mu(t)$ denote the time-varying uncertain sensor attacking signals on dynamic functions in the system. In the terminology of such a logical system, as proposed by Takagi and Sugeno, the IF-part indicates the premise part and the THEN-part means the consequent part, which describe the equation (1) more logical.

By using the logical model, the overall system dynamic is resulted as:

$$\begin{aligned}\dot{x}(t) &= \sum_{l=1}^s \zeta_l(z(t)) \frac{\sum_{i=1}^{r_l} \nu_{l,i}(\xi(t)) [A_{l,i}(t)x(t) + B_{l,i}(t)\mu(t)]}{\sum_{i=1}^{r_l} \nu_{l,i}(\xi(t))} \\ &= \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l(z(t)) h_{l,i}(\xi(t)) [A_{l,i}(t)x(t) + B_{l,i}(t)\mu(t)]\end{aligned}\quad (4)$$

where $\mathcal{A}_{l,i}(t) = A_{l,i} + \tilde{A}_{l,i}(t)$, $\mathcal{B}_{l,i}(t) = B_{l,i} + \tilde{B}_{l,i}(t)$; $\zeta_l(z(t)) = 1$ if $z(t)$ is $\mathbb{M}_l$, otherwise $\zeta_l(z(t)) = 0$; $\xi(t) = [\xi_1(t), \dots, \xi_{r_{li}}(t)]$, $\nu_{l,i}(\xi(t))$ and $h_{l,i}(\xi(t))$ are defined as

$$h_{l,i}(\xi(t)) = \frac{\nu_{l,i}(\xi(t))}{\sum_{i=1}^{r_l} \nu_{l,i}(\xi(t))}, \nu_{l,i}(\xi(t)) = \prod_{j=1}^{r_{li}} M_{l,ij}[\xi(t)], \quad (5)$$

$$\nu_{l,i}(\xi(t)) \geq 0, 0 \leq h_{l,i}(\xi(t)) \leq 1, i \in \mathcal{I},$$

and the grade of membership of $\xi_j(t)$ in $M_{l,ij}$ is $M_{l,ij}[\xi(t)]$.

To avoid triviality, we assume that $\sum_{i=1}^{r_l} \nu_{l,i}(\xi(t)) > 0$ and,

thus, $\sum_{i=1}^{r_l} h_{l,i}(\xi(t)) = 1$ for all t . The universe of discourse $\mathbb{X} : \mathbb{R}^n \times \mathcal{S} \rightarrow \mathbb{R}^n$ for the MJS is given as

$$\mathbb{X} = \bigcup_{l=1}^s \text{Mode } l = \text{Mode } 1 \bigcup \text{Mode } 2 \bigcup \dots \bigcup \text{Mode } s,$$

and $\text{Mode } l \cap \text{Mode } k = \emptyset, \forall l \neq k, l, k \in \mathcal{S}$. In addition, we denote $\zeta_l(z(t))$ and $h_{l,i}(\xi(t))$ by ζ_l and $h_{l,i}$ for notational simplicity in the subsequent development.

For thus a nonlinear dynamic, we use the following control input similar as that in [38], which is mode-dependent with the logical system (3):

Mode l :

IF $z(t)$ is $\mathbb{M}_l$

THEN

Control Rule i :

IF $\xi_1(t)$ is $M_{l,i1}, \dots, \xi_{r_{li}}(t)$ is $M_{l,ir_{li}}$,

THEN $\mu(t) = (\mathcal{K}_{l,i} + \Delta\mathcal{K}_{l,i}(t))x(t)$

where $\mathcal{K}_{l,i} \in \mathbb{R}^{n \times m}$ denotes the gain matrix in control $\mu(t)$, and $\Delta\mathcal{K}_{l,i}(t)x(t)$ captures the actuator attacking signal.

B. The resilient control problem

The control object here is to find a resilient control policy to achieve stochastic stability for partially unknown MJSs in the suffering of uncertain sensor and actuator attacks. Similar to [26], [28], the sensor and actuator attacks are considered to be state dependent. Thus, similar to [21], [22], [38], the $\tilde{A}_{l,i}(t)$ and $\tilde{B}_{l,i}(t)$ are assumed with the following form:

$$[\tilde{A}_{l,i}(t), \tilde{B}_{l,i}(t)] = H_{l,i}F(t) [D_{1l,i}, D_{2l,i}] \quad (7)$$

where $F(t)$ is an *unknown* time-varying matrix with Lebesgue measurable elements and satisfies $F^T(t)F(t) \leq I$, for all $t \geq 0$, matrices $H_{l,i}$, $D_{1l,i}$ and $D_{2l,i}$ for $i \in \mathcal{I}$, $l \in \mathcal{S}$ are known and characterize the structures of these uncertainties. Without loss of generality, here we assume that the actuator attacking signal $\Delta\mathcal{K}(t) = \hat{H}F(t)\hat{D}$, where $\hat{H}$ and $\hat{D}$ are constant matrices for all $t \geq 0$, since an actuator attack in the worst case will lead to the amplitude saturation for actuator in practice.

Let the overall control input be defined as $\mu(t) = \bar{\mathcal{K}}x(t) = (\mathcal{K} + \Delta\mathcal{K})x(t)$. By substituting $\mu(t)$ into system (4) and using equation (7), it will yield the following system equation:

$$\begin{aligned}\dot{x}(t) &= \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} [A_{l,i}x(t) + B_{l,i}\mu(t) + \tilde{A}_{l,i}(t)x(t) \\ &\quad + \tilde{B}_{l,i}(t)\mu(t)] \\ &= \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} [A_{l,i}x(t) + B_{l,i}u(t) + B_{l,i}\hat{H}F(t)\hat{D}x(t) \\ &\quad + H_{l,i}F(t)(D_{1l,i} + D_{2l,i}\bar{\mathcal{K}})x(t)] \\ &= \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} [A_{l,i}x(t) + B_{l,i}u(t) + E_{l,i}w(t)]\end{aligned}\quad (8)$$

where the control policy is $u(t) = \mathcal{K}x(t)$, $E_{l,i} \in \mathbb{R}^{n \times n}$ is the identity matrix, and $w(t) = (B_{l,i}\hat{H}F(t)\hat{D} + H_{l,i}F(t)(D_{1l,i} + D_{2l,i}\bar{\mathcal{K}}))x(t)$ denotes the overall attack signal from physical components, consisting of the actuator and sensor attacks, which are injected into the system.

Remark 1. Note that a control method is developed to prevent not only attacks on sensors, i.e., the uncertain signals $\tilde{A}_{l,i}(t)$ and $\tilde{B}_{l,i}(t)$; but also attacks on actuator as $\Delta\mathcal{K}(t)$. All these unknown attacking signals are reorganized into the term $w(t)$ in system (8), which will worsen the system performances with its maximum destructive capability.

In particular, when $F(t) \equiv 0$, there is $w(t) \equiv \mathbf{0}$, $\mathbf{0}$ is the zero vector with an appropriate dimension, the MJS (8) degenerates into a nominal MJS, where no attacks occur. As in general, the matrices $B_{l,i}$, $i \in \mathcal{I}$ for any $l \in \mathcal{S}$ are chosen to be the same in the logical model.

Assumption 1. We assume that the system in this paper is stochastically controllable, and a mode-dependent control policy exists and can stabilize the system (8). The values of k is available at time t , and mode jumps with a short dwell time, which means that the minimum dwell time $\bar{\Delta} > 0$ in the homogeneous continuous Markovian process.

It is reasonable that the system mode can not be switched ceaselessly in an instant by nature, and the mode dwell time must exist. Let the system state from the initial mode η_0 and initial state $x(0) = x_0$ be $x(t, \eta_0, x_0)$, and we give the following definitions.

Definition 1. *The nominal continuous-time Markovian jump system (8) with $w(t) \equiv \mathbf{0}$ is defined as stochastically stabilizable, if for any initial mode $\eta_0 \in \mathcal{S}$ and state $x_0 \in \mathbb{R}^n$, there exists a control policy $u(t)$ satisfying*

$$\lim_{T \rightarrow \infty} \mathbf{E} \left\{ \int_0^T x^T(t, u) x(t, u) dt | x_0, \eta_0 \right\} \leq x_0^T M x_0 \quad (9)$$

where $\mathbf{E}\{\cdot\}$ means the mathematical expectation and $M = M^T$ is a symmetric positive definite matrix.

A positive definite matrix $\mathcal{X} > 0$ indicates that for all nonzero vector z , $z^T \mathcal{X} z > 0$, which is the so called first kind of generalized positive definite matrix, and it is asymmetric if it is not pointed out in the paper. The stochastic stability of the nominal MJS means that from the mean square sense, a feedback control policy (6) exists and can asymptotically drive system states to zeros from any initial condition.

Note that the problem we address is designing the feedback gain matrix $\mathcal{K}$ for the overall resilient control $\mu(t) = (\mathcal{K} + \Delta \mathcal{K})x(t)$ that stochastically stabilizes MJS (8) under the attacking signal $w(t)$. Considering the closed-loop system (8), there is

$$\begin{aligned} & w^T(t)w(t) \\ &= x^T(t)(B_{l,i}\hat{H}F(t)\hat{D} + H_{l,i}F(t)(D_{1l,i} + D_{2l,i}\bar{\mathcal{K}}))^T \\ & \quad \times (B_{l,i}\hat{H}F(t)\hat{D} + H_{l,i}F(t)(D_{1l,i} + D_{2l,i}\bar{\mathcal{K}}))x(t) \\ & \leq x^T(t)C^T C x(t) \end{aligned} \quad (10)$$

where matrix C is a constant matrix. According to the definition of L_2 -space, the signal satisfies $w(t) \in \mathcal{L}_2[0, \infty)$. Let the mode-dependent H_∞ performance index associated with system (8) be defined as:

$$\begin{aligned} \bar{J}(x(t), \eta_t) &= \mathbf{E} \left\{ \int_t^\infty [x^T(\tau)Q_1 x(\tau) + u^T(\tau)u(\tau) \right. \\ & \quad \left. - \gamma^2 w^T(\tau)w(\tau)] d\tau | x(t), \eta_t \right\} \end{aligned} \quad (11)$$

where matrix $Q_1 > 0$ is symmetric, and $\gamma > 0$. Hence, the stochastically stability is defined as follows.

Definition 2. *The MJS (8) with uncertain attacking signal $w(t)$ is stochastically stable along with an attenuation level $\gamma > 0$, if for all initial state x_0 and η_0 , nonzero $w(t) \in \mathcal{L}_2[0, \infty]$, there is an $\mathcal{L}_2$ -gain not more than γ satisfying*

$$\begin{aligned} & \mathbf{E} \left\{ \int_0^\infty [x^T(t)Q_1 x(t) + u^T(t)u(t)] dt | x_0, \eta_0 \right\} \\ & \leq \gamma^2 \int_0^\infty w^T(t)w(t) dt. \end{aligned} \quad (12)$$

This is also the called H_∞ performance, and the stochastic stability will be guaranteed by the control gain matrix $\mathcal{K}$ when the overall resilient control is solved. Let us regard $u(t)$ and $w(t)$ as the control policies from two rival players, then, a pair of control policies will satisfy the H_∞ performance (12), and the resilient control will be obtained by the distributed parallel compensation scheme in the following sections.

III. THE CONTROL DESIGN WITH ZERO-SUM GAMES FOR THE MARKOVIAN JUMP SYSTEMS

In this part, the resilient control problem of nonlinear MJS (3) can be transformed into solving a set of zero-sum games with two players, where the control policy is designed against the attacking signal, which worsens the system performance with its maximum destructive capability.

Considering an overall performance index $J(t)$, we denote that the attacking signal $w(t)$ as the opponent's policy, then, the objective of this paper is finding a control policy pair (u, w) to make thus a system (8) stable such that satisfying the H_∞ performance (12).

To guarantee the existence of saddle points, it should be guaranteed that $\gamma \geq \gamma^*$, where γ^* is the smallest L_2 -gain value of system (8). Assume that $\gamma \geq \gamma^*$ is satisfied, we have that

$$\begin{aligned} J(x(t), \eta_t) &= \min_{u(t)} \max_{w(t)} \mathbf{E} \left\{ \int_t^\infty \mathcal{U}(x, u, w) d\tau | x(t), \eta_t \right\} \\ &= \max_{w(t)} \min_{u(t)} \mathbf{E} \left\{ \int_t^\infty \mathcal{U}(x, u, w) d\tau | x(t), \eta_t \right\} \end{aligned} \quad (13)$$

where $\mathcal{U}(x, u, w) = x^T(\tau)Qx(\tau) + u^T(\tau)u(\tau) - \gamma^2 w^T(\tau)w(\tau)$, matrix $Q > 0$ is symmetric, and $J(x(t), \eta_t)$ is the overall performance index.

By using the parallel design with distributed compensation, the mode-dependent feedback control policies with a corresponding performance index in logical rules can be developed as follows:

Mode l :

IF $z(t)$ is $\mathbb{M}_l$

THEN

(14)

Control Rule i :

IF $\xi_1(t)$ is $M_{l,i1}, \dots, \xi_{r_{li}}(t)$ is $M_{l,ir_{li}}$,

THEN $\begin{cases} u_{l,i}(t) = K_{l,i}x(t); w_{l,i}(t) = L_{l,i}x(t) \\ J_i(x(t), \eta_t) = \mathbf{E} \left\{ \int_t^\infty \mathcal{U}(x, u_{l,i}, w_{l,i}) d\tau | x(t), \eta_t \right\} \end{cases}$

where $\mathcal{U}(x, u_{l,i}, w_{l,i}) = x^T(\tau)Qx(\tau) + u_{l,i}^T(\tau)u_{l,i}(\tau) - \gamma^2 w_{l,i}^T(\tau)w_{l,i}(\tau)$.

Thus the overall mode-dependent control policies are obtained as:

$$\begin{cases} u(t) = \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} K_{l,i} x(t) = \mathcal{K}x(t) \\ w(t) = \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} L_{l,i} x(t) = \mathcal{L}x(t) \end{cases} \quad (15)$$

Note that $J(x, \eta_t) \neq \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} J_i(x, \eta_t)$, $J(x, \eta_t)$ is the defined overall performance index with two overall control policies $u(t)$ and $w(t)$, and is not equal to the blending of indexes $J_i(x, \eta_t)$, $i \in \mathcal{I}$, consisting of $u_{l,i}(t)$ and $w_{l,i}(t)$.

Lemma 1. Let $S, S_i > 0$ be symmetrical positive definite matrices satisfying $SS_i = S_iS$, then, both the matrices SS_i and S_iS are positive definite.

Proof. Considering the symmetrical positive definite matrices S and S_i , based on the matrix analysis theory of linear algebra, $(SS_i)^T = S_iS = SS_i$, and the invertible matrices $\bar{X}$ and $\underline{X}$ can be founded satisfying $S = \bar{X}^T \bar{X}$ and $S_i = \underline{X}^T \underline{X}$.

Thus it yields

$$\begin{aligned} (\bar{X}^T)^{-1}(SS_i)\bar{X}^T &= (\bar{X}^T)^{-1}(\bar{X}^T \bar{X} \underline{X}^T \underline{X})\bar{X}^T \\ &= \bar{X} \underline{X}^T \underline{X} \bar{X}^T = \bar{X} \underline{X}^T (\bar{X} \underline{X}^T)^T > 0 \end{aligned} \quad (16)$$

which denotes that matrix SS_i is positive definite by similar matrices property. By this way, it can be also obtained that matrix S_iS is positive definite. The proof is completed. $\square$

Lemma 2. Given two quadratic functions as $V(t) = z^T(t)SX Sz(t)$ and $V_i(t) = z^T(t)SX S_i z(t)$, where $S, S_i > 0$ are symmetrical, $SS_i = S_iS$, and $X = X^T$ is a symmetric matrix, thus, one has $V(t) = aV_i(t)$, $a > 0$ is a constant.

Proof. According to Lemma 1, it can be obtained that for any vector z , $z^T SS z = az^T SS_i z \geq 0$, $a > 0$. As $X = X^T$ is a symmetric matrix, there is $z^T SX Sz = az^T SX S_i z$, which indicates that $V(t) = aV_i(t)$. This gives the proof. $\square$

Inserting the control policies (15) into the nonlinear MJS (8), it yields

$$\begin{aligned} \dot{x}(t) &= \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} \left[A_{l,i} x(t) + B_{l,i} K x(t) + E_{l,i} \mathcal{L} x(t) \right] \\ &= \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} \left[A_{l,i} + B_{l,i} \sum_{j=1}^{r_l} h_{l,j} K_{l,j} \right. \\ &\quad \left. + E_{l,i} \sum_{j=1}^{r_l} h_{l,j} L_{l,j} \right] x(t) \\ &= \sum_{l=1}^s \sum_{i=1}^{r_l} \sum_{j=1}^{r_l} \zeta_l h_{l,i} h_{l,j} \left[A_{l,i} + B_{l,i} K_{l,j} + E_{l,i} L_{l,j} \right] x(t) \\ &= [\bar{A} + \bar{B}K + \bar{E}\mathcal{L}] x(t) \end{aligned} \quad (17)$$

where $\bar{A} = \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} A_{l,i}$, $\bar{B} = \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} B_{l,i}$, $\bar{E} = \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} E_{l,i}$, $l \in \mathcal{S}$.

As the saddle points exist, based on the optimal control theory, we suppose that the overall and in i th rule performance indexes can be expressed by the symmetric positive definite quadratic form as $J(x(t), \eta_t = l) = x^T(t) P^l x(t)$, $J_i(x(t), \eta_t = l) = x^T(t) P_i^l x(t)$ with $P^l P_i^l = P_i^l P^l$, and assume $a \in (0, 2)$ such that $J(x, \eta_t) < 2J_i(x, \eta_t)$, $\forall i \in \mathcal{I}$ in the model, then the control policies can be derived with the principle of optimality.

Theorem 1. Considering the MJS (8) in i th rule, the solution matrix in quadratic performance index $J_i(x(t), \eta_t)$ satisfies the following stochastic coupled ARE:

$$\begin{aligned} &[A_{l,i} + B_{l,i} K_{l,i} + E_{l,i} L_{l,i}]^T P_i^l \\ &\quad + P_i^l [A_{l,i} + B_{l,i} K_{l,i} + E_{l,i} L_{l,i}] \\ &= - \sum_{k=1}^s \pi_{lk} P_i^k - Q - K_{l,i}^T K_{l,i} + \gamma^2 L_{l,i}^T L_{l,i} \end{aligned} \quad (18)$$

where matrices $P_i^l, P_i^k, i \in \mathcal{I}, l, k \in \mathcal{S}$ are symmetric positive definite. Besides, the corresponding mode-dependent feedback controls are $u_{l,i}(t) = K_{l,i} x(t) = -B_{l,i}^T P_i^l x(t)$ and $w_{l,i}(t) = L_{l,i} x(t) = \frac{1}{\gamma^2} E_{l,i}^T P_i^l x(t)$.

Proof. For a hybrid stochastic process $\{(\eta_t, x(t)), t \in [t_0, T]\}$, the weak infinitesimal operator $\mathfrak{L}$ can be considered as the deterministic derivative by nature stochastic analog in the function space $[t_0, T] \times \mathcal{S} \times \mathbb{R}^n$ [17]. Hence, for the MJS (8) in i th rule, assuming that the performance index $J_i(x(t), \eta_t)$, $i \in \mathcal{I}$ is a stochastic quadratic function as $J_i(x(t), \eta_t = l) = x^T(t) P_i^l x(t)$, one has:

$$\begin{aligned} \mathfrak{L} J_i(x, k) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} [\mathbf{E}\{J_i(x(t + \Delta t), \eta_{t+\Delta t}) | x(t), \eta_t = l\} \\ &\quad - J_i(x(t), \eta_t = l)] \\ &= x^T(t) [(A_{l,i} + B_{l,i} K_{l,i} + E_{l,i} L_{l,i})^T P_i^l + P_i^l \\ &\quad \times (A_{l,i} + B_{l,i} K_{l,i} + E_{l,i} L_{l,i}) + \sum_{k=1}^s \pi_{lk} P_i^k] x(t) \\ &= -x^T(t) [Q + K_{l,i}^T K_{l,i} - \gamma^2 L_{l,i}^T L_{l,i}] x(t) \\ &= -\mathcal{U}(x, u_{l,i}, w_{l,i}) \end{aligned} \quad (19)$$

which indicates the equation (18). And it can further obtain

$$\begin{aligned} x^T(t) [A_{l,i}^T P_i^l + P_i^l A_{l,i} + \sum_{k=1}^s \pi_{lk} P_i^k] x(t) \\ + 2x^T(t) [P_i^l B_{l,i} u_{l,i}(t) + P_i^l E_{l,i} w_{l,i}(t)] \\ = -x^T(t) Q x(t) - u_{l,i}^T(t) u_{l,i}(t) + \gamma^2 w_{l,i}^T(t) w_{l,i}(t). \end{aligned} \quad (20)$$

Thus by using Bellman's optimization principle, derivative equation (20) with respect to policies u_i and w_i respectively, there are $u_{l,i}(t) = K_{l,i} x(t) = -B_{l,i}^T P_i^l x(t)$ and $w_{l,i}(t) = L_{l,i} x(t) = \frac{1}{\gamma^2} E_{l,i}^T P_i^l x(t)$, which gives the feedback control policies. The proof is thus completed. $\square$

Based on the game theory, the policies are derived from their objectives, respectively. It means that the policy making in each player does not rely on policies of other players, thus, they are independent from each other.

IV. IRL BASED RESILIENT CONTROL ALGORITHM AND STABILITY ANALYSIS FOR MJSS WITH COMPLETELY UNKNOWN TRANSITION PROBABILITY

In the section, a novel resilient control IRL algorithm is proposed for the nonlinear MJS by solving the control policies of zero-sum games. In the light of the adaptive learning ability, the data based IRL technique is utilized to obtain the overall resilient control and the stability of states under the algorithm is proved by Lyapunov theory.

A. Data Based IRL Resilient Control Algorithm for MJSSs with Completely Unknown Transition Probability

According to the optimal control design and nonlinear system (8), the overall saddle point solution can hardly be obtained in a progress, thus, by using the local logical model, the following data based IRL resilient control algorithm is designed without transition probability information for partially unknown MJSSs.

Algorithm 1 Data-Based IRL Resilient Control Algorithm for MJSSs with Completely Unknown Transition Probability

Initialization

Select a sequence $P_i(0) = (P_i^1(0), \dots, P_i^l(0), \dots, P_i^s(0))$ for every rule, initial admissible feedback matrices $K_{l,i}(0)$ and $L_{l,i}(0)$. Let $0 < \Delta t < \bar{\Delta}$ be a small time constant, step $\lambda = 1$ and the learning process be conducted as time t

Procedure

- 1: **while** $\max_{l,i} \|P_i^l(\lambda + 1) - P_i^l(\lambda)\| > \epsilon$, $\epsilon > 0$ is a small constant **do**
- 2: **for** for all $\eta_t = l \in S$ **do**
- 3: **for** for all $i \in \mathcal{I}$ **do**
- 4: Learn the quadratic function $x^T(t)P_i^l(\lambda)x(t)$ from

$$x^T(t)P_i^l(\lambda)x(t) = \int_t^{t+\Delta t} \mathcal{U}(x, u_{l,i}, w_{l,i})d\tau + x^T(t+\Delta t)P_i^l(\lambda)x(t+\Delta t) \quad (21)$$

- 5: Solve the decoupled control matrices by

$$\begin{aligned} K_{l,i}(\lambda + 1) &= -B_{l,i}^T P_i^l(\lambda); \\ L_{l,i}(\lambda + 1) &= \frac{1}{\gamma^2} E_{l,i}^T P_i^l(\lambda) \end{aligned} \quad (22)$$

- 6: Update the feedback control policies as

$$\begin{aligned} u_{l,i}(t) &= K_{l,i}(\lambda + 1)x(t); \\ w_{l,i}(t) &= L_{l,i}(\lambda + 1)x(t) \end{aligned} \quad (23)$$

- 7: **end for**
- 8: **end for**
- 9: Set $\lambda = \lambda + 1$

- 10: **end while**

- 11: Use the overall control policies with

$$u(t) = \mathcal{K}x(t); w(t) = \mathcal{L}x(t) \quad (24)$$

$$\text{where } \mathcal{K} = \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} K_{l,i} \text{ and } \mathcal{L} = \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} L_{l,i}$$

End Procedure

By using the proposed Algorithm 1, the complex stochastic coupled AREs in zero-sum games are converted to IRL solutions, which involves an easy computational process and a simple solving framework. Besides, the main computation is solving the parameter equation (21), and the least square method can achieve the solution easily, with a very small computational load.

Lemma 3. Suppose the symmetric positive definite matrix P_i^l is solved by using the integral equation (21), $\Delta t > 0$ is a small

constant, then for every rule $i \in \mathcal{I}$, the solution of equation (21) equals to solving the coupled AREs (18).

Proof. Since $J_i(x, k) = x^T(t)P_i^l x(t)$ and using equation (21), there is

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \left[\mathbf{E} \{ J_i(x(t + \Delta t), \eta_{(t+\Delta t)}) | x(t), \eta_t = l \} \right. \\ & \quad \left. - J_i(x(t), \eta_t = l) \right] \\ &= \lim_{\Delta t \rightarrow 0} \left(\sum_{k=1, k \neq l}^s (\Delta t) \pi_{lk} [x^T(t + \Delta t) P_i^l x(t + \Delta t) \right. \\ & \quad \left. - x^T(t) P_i^l x(t)] + (1 - \sum_{k=1, k \neq l}^s (\Delta t) \pi_{lk}) \right. \\ & \quad \left. \times [x^T(t + \Delta t) P_i^l x(t + \Delta t) - x^T(t) P_i^l x(t)] \right) \quad (25) \\ &\triangleq [x^T(t + \Delta t) P_i^l x(t + \Delta t) - x^T(t) P_i^l x(t)] \\ &= \int_t^{t+\Delta t} -[x^T(\tau) Q x(\tau) + u_{l,i}^T(\tau) w_{l,i}(\tau) \\ & \quad - \gamma^2 w_{l,i}^T(\tau) w_{l,i}(\tau)] d\tau \\ &= - \int_t^{t+\Delta t} \mathcal{U}(x(\tau), u_{l,i}(\tau), w_{l,i}(\tau)) d\tau. \end{aligned}$$

Based on the Newton-Leibniz formula, it yields

$$\begin{aligned} & [x^T(t + \Delta t) P_i^l x(t + \Delta t) - x^T(t) P_i^l x(t)] \\ &= \int_t^{t+\Delta t} x^T(\tau) \left[(A_{l,i} + B_{l,i} K_{l,i} + E_{l,i} L_{l,i})^T P_i^l \right. \\ & \quad \left. + P_i^l (A_{l,i} + B_{l,i} K_{l,i} + E_{l,i} L_{l,i}) + \sum_{k=1}^s \pi_{lk} P_i^k \right] x(\tau) d\tau \\ &= \int_t^{t+\Delta t} -x^T(\tau) [Q + K_{l,i}^T K_{l,i} - \gamma^2 L_{l,i}^T L_{l,i}] x(\tau) d\tau \\ &= - \int_t^{t+\Delta t} \mathcal{U}(x(\tau), u_{l,i}(\tau), w_{l,i}(\tau)) d\tau. \quad (26) \end{aligned}$$

Combining equations (25) and (26), it indicates that the solved matrix P_i^l in (21) is equal to the solution of (18). This gives the proof. $\square$

Remark 2. Note that the coupled ARE (18) is decoupled and solved by using the developed data based IRL resilient control algorithm. Compared to some general control methods of nonlinear MJSSs, this design is novel and preponderant as that some essential measured information, matrices $A_{l,i}$ and transition probability π_{lk} , can be completely unknown. It means that, based on the state data, the method can reduce the modeling requirements and eliminate the dependence of some system knowledge efficiently as these information can hardly be measured or obtained accurately in practical applications.

B. Stability Analysis of the Data Based IRL Resilient Control Algorithm

As a kind of key learning methods in the artificial intelligence field, the novel IRL resilient control algorithm for nonlinear MJSSs is developed without using information of the system matrices $A_{l,i}$ and stationary transition probability

π_{lk} , $l, k \in \mathcal{S}$. The local optimal policies provide an overall stabilizing solution, which satisfies the H_∞ performance by the following theorem.

Theorem 2. Consider the overall performance index (13) with quadratic form $J(x, \eta_t = l) = x^T(t)P^l x(t)$, where $P^l > 0$ is symmetric, x_0 and η_0 are initial state and mode, respectively, then, the system (17) has the H_∞ performance (12) if the proposed overall control (24) from Algorithm 1 is used.

Proof. For the nonlinear MJS (17), using the overall stochastic quadratic Lyapunov function $J(x, \eta_t = l) = x^T(t)P^l x(t)$ and operator $\mathfrak{L}$, one has

$$\begin{aligned} \mathfrak{L}J(x, l) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} [\mathbf{E}\{J(x(t + \Delta t), \eta_{t+\Delta t}) | x(t), \eta_t = l\} \\ &\quad - J(x(t), \eta_t = l)] \\ &= x^T(t) \left[(\bar{A} + \bar{B}\bar{K} + \bar{E}\bar{\mathcal{L}})^T P^l \right. \\ &\quad \left. + P^l (\bar{A} + \bar{B}\bar{K} + \bar{E}\bar{\mathcal{L}}) + \sum_{k=1}^s \pi_{lk} P^k \right] x(t) \end{aligned} \quad (27)$$

where the control policies are from equation (24).

According to the optimal principle, there exists a symmetric positive definite matrix P^l satisfying the following equation

$$\begin{aligned} x^T(t) &\left[(\bar{A} + \bar{B}\bar{K} + \bar{E}\bar{\mathcal{L}})^T P^l \right. \\ &\quad \left. + P^l (\bar{A} + \bar{B}\bar{K} + \bar{E}\bar{\mathcal{L}}) + \sum_{k=1}^s \pi_{lk} P^k \right] x(t) \\ &= -x^T(t) [Q + \bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}] x(t) \end{aligned} \quad (28)$$

where $\bar{K} = -\bar{B}^T P^l$ and $\bar{\mathcal{L}} = \frac{1}{\gamma^2} \bar{E}^T P^l$.

By using the aforementioned Lemma 2 and assumption after equation (17), there exists an $\alpha \in (0, 2)$ satisfying

$$\begin{aligned} x^T(t) &[\bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}] x(t) \\ &= x^T(t) [P^l \bar{B} \bar{B}^T P^l - P^l \bar{E} \frac{1}{\gamma^2} \bar{E}^T P^l] x(t) \\ &= x^T(t) [P^l (\bar{B} \bar{B}^T - \bar{E} \frac{1}{\gamma^2} \bar{E}^T) P^l] x(t) \\ &= \alpha x^T(t) \left[\sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} [P^l (\bar{B} \bar{B}^T - \bar{E} \frac{1}{\gamma^2} \bar{E}^T) P_i^l] \right] x(t) \\ &= \alpha x^T(t) \left[\sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} (P^l \bar{B} \bar{B}^T P_i^l) \right. \\ &\quad \left. - \sum_{l=1}^s \sum_{i=1}^{r_l} \zeta_l h_{l,i} (P^l \bar{E} \frac{1}{\gamma^2} \bar{E}^T P_i^l) \right] x(t) \\ &= \alpha x^T(t) [\bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}] x(t). \end{aligned} \quad (29)$$

Then combining the equation (28), the equation (27) can

further become

$$\begin{aligned} \mathfrak{L}J(x, l) &= x^T(t) [(\bar{A}^T P^l - \bar{K}^T \bar{K} + \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) \\ &\quad + (P^l \bar{A} - \bar{K}^T \bar{K} + \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) + \sum_{k=1}^s \pi_{lk} P^k] x(t) \\ &\quad - x^T(t) [(\bar{A}^T P^l - \bar{K}^T \bar{K} + \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) \\ &\quad + (P^l \bar{A} - \bar{K}^T \bar{K} + \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) + \sum_{k=1}^s \pi_{lk} P^k] x(t) \\ &\quad - x^T(t) [Q + \bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}] x(t) \\ &= -2x^T(t) [(\bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) - (\bar{K}^T \bar{K} \\ &\quad - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}})] x(t) - x^T(t) [Q + \bar{K}^T \bar{K} \\ &\quad - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}] x(t) \\ &= -\frac{1}{\alpha} [2x^T(t) (\bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) x(t) - 2\alpha x^T(t) \\ &\quad \times (\bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) x(t) + \alpha x^T(t) (Q + \bar{K}^T \bar{K} \\ &\quad - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}) x(t)] \\ &= -(1/\rho) x^T(t) [Q_1 + \bar{K}^T \bar{K} - \gamma^2 \bar{\mathcal{L}}^T \bar{\mathcal{L}}] x(t) \end{aligned} \quad (30)$$

where $\rho = \frac{\alpha}{2-\alpha}$ and $Q_1 = \rho Q$.

Note that, when the attacking signal $w(t) \equiv \mathbf{0}$ in the equation (30), there is $\mathfrak{L}J(x, l) \leq 0$, which ensures the stochastically asymptotic stability of states in the closed-loop system according to Definition 1. Integrating both sides of equation (30) in a time interval $[0, T]$, we have

$$\begin{aligned} \mathbf{E}\{ \int_0^T \rho \mathfrak{L}J(x(t), \eta_t) dt | x_0, \eta_0 \} &+ \mathbf{E}\{ \int_0^T x^T(t) \\ &\times [\rho Q + P^l \bar{B} \bar{B}^T P^l] x(t) dt \\ &- \gamma^2 \int_0^T x^T(t) [P^l \bar{E} \frac{1}{\gamma^4} \bar{E}^T P^l] x(t) dt | x_0, \eta_0 \} \\ &= \mathbf{E}\{ \int_0^T \rho \mathfrak{L}J(x(t), \eta_t) dt | x_0, \eta_0 \} + \mathbf{E}\{ \int_0^T [x^T(t) Q_1 x(t) \\ &+ u^T(t) u(t)] dt | x_0, \eta_0 \} - \gamma^2 \int_0^T [w^T(t) w(t)] dt = 0 \end{aligned} \quad (31)$$

then taking the limit as $T \rightarrow \infty$, there is

$$\begin{aligned} \mathbf{E}\{ \int_0^\infty [x^T(t) Q_1 x(t) + u^T(t) u(t)] dt | x_0, \eta_0 \} \\ \leq \gamma^2 \int_0^\infty w^T(t) w(t) dt \end{aligned} \quad (32)$$

which matches the Definition 2. Hence the stochastic stability is obtained, which completes the proof. $\square$

V. SIMULATION RESULT

In this section, a typical multi-mode robot arm system in many industrial applications is presented to verify the designed data based IRL resilient control algorithm. Let us consider the dynamic equation as [16], [38]:

$$\ddot{\theta}(t) = -\frac{J M g}{L} \sin(\theta(t)) - \frac{D(t)}{L} \dot{\theta}(t) + \frac{1}{L} \mu(t) \quad (33)$$

where the angular position $\theta(t)$ denotes a system state, the length J of the arm is 0.5, the uncertain coefficient $D(t)$ of viscous friction is selected with its nominal value $D = 2$ here, the acceleration g of gravity uses 9.81, the control is $\mu(t)$, the moment L of the inertia and the mass M of the payload will lead different system modes.

Then, under the condition $\theta(t) \in (-179.4270^\circ, 179.4270^\circ)$, one can rewrite the nonlinear term $\sin(\theta(t))$ as [38]:

$$\sin(\theta(t)) = h_1(\theta(t)) \cdot \theta(t) + h_2(\theta(t)) \cdot \beta \cdot \theta(t) \quad (34)$$

where $\beta = 10^{-2}/\pi$. By solving the equation above, we can further calculate the membership functions as:

$$h_1(\theta(t)) = \begin{cases} \frac{\sin(\theta(t)) - \beta \cdot \theta(t)}{\theta(t) \cdot (1 - \beta)}, & \theta(t) \neq 0 \\ 1, & \theta(t) = 0 \end{cases}, \quad (35)$$

and $h_2(\theta(t)) = 1 - h_1(\theta(t))$.

Define $x(t) = [\theta(t), \dot{\theta}(t)]^T$, then, the Markov jump uncertain nonlinear system will be rewritten as the following revised T-S logical model:

Mode 1:

IF parameters $M = 1$ and $L = 1$

THEN

Plant Rule 1:

IF state $\theta(t)$ is “around 0 rad”,

THEN $\dot{x}(t) = (A_{1,1} + \tilde{A}_{1,1})x(t) + (B_{1,1} + \tilde{B}_{1,1})\mu(t)$

Plant Rule 2:

IF state $\theta(t)$ is “around $-\pi$ or π rad”,

THEN $\dot{x}(t) = (A_{1,2} + \tilde{A}_{1,2})x(t) + (B_{1,2} + \tilde{B}_{1,2})\mu(t)$

Mode 2:

IF parameters $M = 2$ and $L = 2$

THEN

Plant Rule 1:

IF state $\theta(t)$ is “around 0 rad”,

THEN $\dot{x}(t) = (A_{2,1} + \tilde{A}_{2,1})x(t) + (B_{2,1} + \tilde{B}_{2,1})\mu(t)$

Plant Rule 2:

IF state $\theta(t)$ is “around $-\pi$ or π rad”,

THEN $\dot{x}(t) = (A_{2,2} + \tilde{A}_{2,2})x(t) + (B_{2,2} + \tilde{B}_{2,2})\mu(t)$

Mode 3:

IF parameters $M = 3$ and $L = 3$

THEN

Plant Rule 1:

IF state $\theta(t)$ is “around 0 rad”,

THEN $\dot{x}(t) = (A_{3,1} + \tilde{A}_{3,1})x(t) + (B_{3,1} + \tilde{B}_{3,1})\mu(t)$

Plant Rule 2:

IF state $\theta(t)$ is “around $-\pi$ or π rad”,

THEN $\dot{x}(t) = (A_{3,2} + \tilde{A}_{3,2})x(t) + (B_{3,2} + \tilde{B}_{3,2})\mu(t)$

where $x_1(t) = \theta(t)$, $x_2(t) = \dot{\theta}(t)$, and $A_{1,1} = \begin{bmatrix} 0 & 1 \\ -gL & -D \end{bmatrix}$,

$$A_{2,1} = \begin{bmatrix} 0 & 1 \\ -gL & -0.5D \end{bmatrix}, A_{3,1} = \begin{bmatrix} 0 & 1 \\ -gL & -0.25D \end{bmatrix},$$

$$A_{1,2} = \begin{bmatrix} 0 & 1 \\ -gL/\pi & -D \end{bmatrix}, A_{2,2} = \begin{bmatrix} 0 & 1 \\ -gL/\pi & -0.5D \end{bmatrix},$$

$$A_{3,2} = \begin{bmatrix} 0 & 1 \\ -gL/\pi & -0.25D \end{bmatrix}, B_{3,1} = B_{3,2} = \begin{bmatrix} 0 \\ 0.25 \end{bmatrix},$$

$$B_{2,1} = B_{2,2} = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}, B_{1,1} = B_{1,2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \text{ the uncertain}$$

control gain attack matrix is $\Delta K = \hat{H}F(t)\hat{D}$, where $\hat{H} = 5$, $\hat{D} = [0.1 \ 0.2]$, and $F(t) = 0.8\sin(t)$; the uncertain system attack matrices are $[\tilde{A}_{l,i}(t) \ \tilde{B}_{l,i}(t)] = H_{l,i}F(t) [D_{1l,i} \ D_{2l,i}]$, where $D_{11,i} = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.02 \end{bmatrix}$,

$$D_{12,i} = \begin{bmatrix} -0.01 & 0 \\ 0 & 0.02 \end{bmatrix}, \quad D_{13,i} = \begin{bmatrix} 0.01 & 0 \\ 0 & -0.02 \end{bmatrix},$$

$$D_{21,i} = \begin{bmatrix} -0.01 \\ 0 \end{bmatrix}, D_{22,i} = \begin{bmatrix} 0.01 \\ 0 \end{bmatrix}, D_{23,i} = \begin{bmatrix} 0.02 \\ 0 \end{bmatrix}, H_{1,i} = 0.01,$$

$$H_{2,i} = 0.02, H_{3,i} = 0.01; \text{ and } \Pi = \begin{bmatrix} -0.3 & 0.25 & 0.05 \\ 0.1 & -0.2 & 0.1 \\ 0.03 & 0.07 & -0.1 \end{bmatrix} \text{ is}$$

the transition-probability-rate matrix.

Note that under the actuator and sensor attacks, the novel data based IRL resilient control algorithm for the MJS runs without the information of stationary transition probability π_{lk} , $l, k \in \mathcal{S}$ and dynamic matrices $A_{l,i}$. Thus the transition-probability-rate matrix Π can be completely avoided during the solving process by the proposed Algorithm 1.

Fig. 1. State trajectories of MJS states under attacks via designed data based IRL resilient control algorithm

Select the initial system state $x(0) = [1, -1]^T$, matrix $Q = \text{diag}\{10, 10\}$ for the system evolution, and $\gamma = 0.2$ in the simulation. Hence by using the proposed resilient control algorithm for thus a nonlinear system, the dynamic trajectories are presented in the Fig. 1, where one can find that the system state trajectories converge to origin as desired. Moreover, the modes of the MJS during the solving process are also depicted in Fig. 1.

The updating procedure of the solution matrices are presented in Fig. 2, together with the converging matrix parameters of P_i^l , where $i \in \mathcal{I}$ and $l \in \mathcal{S}$, both the convergence of Algorithm 1 and the stability of the dynamics under the designed control scheme are demonstrated by the simulation results.

Fig. 2. Parameter Evolution of the obtained positive definite matrices P_i^l during the solving process by the proposed IRL algorithm

Without requiring the system matrices $A_{l,i}$ and stationary transition probability π_{lk} for each transmission during the process, the proposed IRL resilient control algorithm learns the mode-dependent optimal solution for every rule successfully, and the positive definite solution matrices from state data are obtained as:

$$P_1^1 = \begin{bmatrix} 14.8348 & 1.4085 \\ 1.4085 & 2.1524 \end{bmatrix}, P_2^1 = \begin{bmatrix} 13.8675 & 2.6841 \\ 2.6841 & 2.5177 \end{bmatrix},$$

$$P_1^2 = \begin{bmatrix} 22.4075 & 1.0010 \\ 1.0010 & 4.1930 \end{bmatrix}, P_2^2 = \begin{bmatrix} 14.0839 & 3.1460 \\ 3.1460 & 5.6563 \end{bmatrix},$$

$$P_1^3 = \begin{bmatrix} 50.0480 & 1.0174 \\ 1.0174 & 10.0602 \end{bmatrix}, P_2^3 = \begin{bmatrix} 36.7458 & 3.2564 \\ 3.2564 & 23.1731 \end{bmatrix}.$$

And it can be found that $P_i^l P_j^k > 0$ for all $l, k \in \mathcal{S}$ and $i, j \in \mathcal{I}$, as proved in Lemma 1, which means that the product of any two symmetric positive definite matrices is still positive definite.

Fig. 3. Trajectories of the overall resilient control policy

According to the solutions from the zero-sum games, the overall resilient feedback control policy is also achieved in the Algorithm 1 and presented in Fig. 3, where the gain matrices of controller in rules are obtained as

$$K_{1,1} = [-1.4085, -2.1524],$$

$$K_{1,2} = [-2.6841, -2.5177],$$

$$K_{2,1} = [-0.5005, -2.0965],$$

$$K_{2,2} = [-1.5730, -2.8281],$$

$$K_{3,1} = [-0.2544, -2.5150],$$

$$K_{3,2} = [-0.8141, -5.7933].$$

VI. CONCLUSION

In this paper, a data-driven IRL control algorithm has been developed for nonlinear MJSs without any information of transition probability. By regarding the attacking signal as an opponent's policy, the resilient control problem has been transformed into a set of zero-sum games. Hence, the optimal policies in rules of two players satisfy the saddle point and can be obtained by solving the stochastic coupled AREs. Combining the IRL technique with the T-S logical system, the new algorithm has been proposed to achieve the solution, and it can run without using any information about the stationary transition probability. Moreover, the system matrices in rules can also be partially unknown as the new architecture requires less transmission or computation during the learning process. The stochastic stability has been proved by Lyapunov theory and the designed algorithm has been applied in a typical multi-mode robot arm application, where the effectiveness and practicality have been demonstrated by the simulation results. In our further work, we will attempt to improve the proposed approach to solve more other complicated control

problems, such as resilient filtering in power production, fixed-time resilient control, etc., which serves a security defense system in industrial implementations.

REFERENCES

- [1] K. Doya, H. Kimura, M. Kawato, "Neural mechanisms of learning and control," *IEEE Control Systems Magazine*, vol. 21, no. 4, pp. 42-54, Aug. 2001.
- [2] C.L.P. Chen, B. Wang, "Random-positioned license plate recognition using hybrid broad learning system and convolutional networks," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi:10.1109/TITS.2020.3011937, 2020.
- [3] F. Wang, H. Zhang, D. Liu, "Adaptive dynamic programming: an introduction," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 39-47, May. 2009.
- [4] A. Al-Tamimi, F.L. Lewis, M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 38, no. 4, pp. 943-949, Aug. 2008.
- [5] T. Bian, Y. Jiang, Z.P. Jiang, "Adaptive dynamic programming for stochastic systems with state and control dependent noise," *IEEE Trans. Autom. Contr.*, vol. 61, no. 12, pp. 4170-4175, Dec. 2016.
- [6] D. Liu, Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 621-634, Mar. 2014.
- [7] C. Mu, Y. Zhang, H. Jia, H. He, "Energy-storage-based intelligent frequency control of microgrid with stochastic model uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1748-1758, Mar. 2020.
- [8] K. Zhang, H. Zhang, Y. Cai, R. Su, "Parallel optimal tracking control schemes for mode-dependent control of coupled Markov jump systems via integral RL method," *IEEE Trans. Autom. Sci. Engin.*, vol. 17, no. 3, pp. 1332-1342, Jul. 2020.
- [9] D. Liu, Q. Wei, "Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 779-789, Apr. 2013.
- [10] P. Shi, X. Su, F. Li, "Dissipativity-based filtering for fuzzy switched systems with stochastic perturbation," *IEEE Trans. Autom. Contr.*, vol. 61, no. 6, pp. 1694-1699, Jun. 2016.
- [11] X. Xie, D. Yue, C. Peng, "Relaxed real-time scheduling stabilization of discrete-time takagi-sugeno fuzzy systems via an alterable-weights-based ranking switching mechanism," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3808-3819, Dec. 2018.
- [12] H.N. Wu, Z.P. Wang, "Observer-based H_∞ sampled-data fuzzy control for a class of nonlinear parabolic PDE systems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 454-473, Apr. 2018.
- [13] Y. Wang, L. Zheng, H. Zhang, W. Zheng, "Fuzzy observer-based repetitive tracking control for nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 10, pp. 2401-2415, Oct. 2020.
- [14] Z.G. Wu, S. Dong, P. Shi, H. Su, T. Huang, R. Lu, "Fuzzy-model-based nonfragile guaranteed cost control of nonlinear Markov jump systems," *IEEE Trans. Cybern.*, vol. 47, no. 8, pp. 2388-2397, Aug. 2017.
- [15] H. Shen, F. Li, H. Yan, H.R. Karimi, H. Lam, "Finite-time event-triggered H_∞ control for T-S fuzzy Markov jump systems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3122-3135, Oct. 2018.
- [16] H.N. Wu, K. Cai, "Mode-independent robust stabilization for uncertain Markovian jump nonlinear systems via fuzzy control," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 36, no. 3, pp. 509-519, Jun. 2005.
- [17] N.S.D. Arrifano, V.A. Oliveira, "Robust H_∞ fuzzy control approach for a class of Markovian jump nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 738-754, Dec. 2006.
- [18] S. Tong, Y. Li, Y. Li, Y. Liu, "Observer-based adaptive fuzzy backstepping control for a class of stochastic nonlinear strict-feedback systems," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 41, no. 6, pp. 1693-1704, Dec. 2011.
- [19] K. Zhang, H. Zhang, Y. Mu, C. Liu, "Decentralized tracking optimization control for partially unknown fuzzy interconnected systems via reinforcement learning method," *IEEE Trans. Fuzzy Syst.*, to be published, doi:10.1109/TFUZZ.2020.2966418, 2020.
- [20] Y. Wang, W.X. Zheng, H. Zhang, "Dynamic event-based control of nonlinear stochastic systems," *IEEE Trans. Autom. Contr.*, vol. 62, no. 12, pp. 6544-6551, Dec. 2017.
- [21] M.S. Mahmoud, *Resilient Control of Uncertain Dynamical Systems*, LNCIS 303, Springer: Verlag Berlin Heidelberg, 2004.
- [22] K.D. Pham, *Resilient Controls for Ordering Uncertain Prospects*, Springer, 2014.
- [23] X. Zhang, Q.L. Han, X. Ge, et al. "Networked control systems: A survey of trends and techniques," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 1, pp. 1-17, Jan. 2020.
- [24] J. Hu, J. Shen, D. Lee, "Resilient stabilization of switched linear control systems against adversarial switching," *IEEE Trans. Autom. Contr.*, vol. 62, no. 8, pp. 3820-3834, Aug. 2017.
- [25] Z. Cao, Y. Niu, J. Song, "Finite-time sliding-mode control of Markovian jump cyber-physical systems against randomly occurring injection attacks," *IEEE Trans. Autom. Contr.*, vol. 65, no. 3, pp. 1264-1271, Mar. 2020.
- [26] L. An, G.H. Yang, "Improved adaptive resilient control against sensor and actuator attacks," *Inform. Scienc.*, vol. 432, pp. 145-156, Jan. 2018.
- [27] B. Ning, Q.L. Han, Q. Lu, "Fixed-time leader-following consensus for multiple wheeled mobile robots," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4381-4392, Oct. 2020.
- [28] X. Jin, W.M. Haddad, T. Yucelen, "An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems," *IEEE Trans. Autom. Contr.*, vol. 62, no. 11, pp. 6058-6064, Nov. 2017.
- [29] X. Huang, J. Dong, "Reliable control policy of cyber-physical systems against a class of frequency-constrained sensor and actuator attacks," *IEEE Trans. Cybern.*, vol. 48, no. 12, pp. 3432-3439, Dec. 2018.
- [30] B. Ning, Q.L. Han, Z. Zuo, "Practical fixed-time consensus for integrator-type multi-agent systems: A time base generator approach," *Automatica*, vol. 105, pp. 406-414, Jul. 2019.
- [31] M. Zhu, S. Martinez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Trans. Autom. Contr.*, vol. 59, no. 3, pp. 804-808, Mar. 2014.
- [32] X.M. Zhang, Q.L. Han, X. Ge, L. Ding, "Resilient control design based on a sampled-data model for a class of networked control systems under denial-of-service attacks," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3616-3626, Aug. 2020.
- [33] J.L. Zhang, X. Chen, G. Gu, "State consensus for discrete-time multi-agent systems over time-varying graphs," *IEEE Trans. Autom. Contr.*, vol. 66, no. 1, pp. 346-353, Jan. 2021.
- [34] Y. Yuan, Z. Wang, P. Zhang, H. Liu, "Near-optimal resilient control strategy design for state-saturated networked systems under stochastic communication protocol," *IEEE Trans. Cybern.*, vol. 48, no. 8, pp. 3155-3167, Aug. 2019.
- [35] F. Pasqualetti, F. Dorfler, F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *IEEE Contr. Syst. Mag.*, vol. 35, no. 1, pp. 110-127, Feb. 2015.
- [36] Y. Yuan, P. Zhang, Z. Wang, L. Guo, "On resilient strategy design of multi-tasking optimal control for state-saturated systems with nonlinear disturbances: The time-varying case," *Automatica*, vol. 107, pp. 138-145, Sep. 2019.
- [37] Z. Zhang, Y. Niu, J. Song, "Input-to-state stabilization of interval type-2 fuzzy systems subject to cyberattacks: An observer-based adaptive sliding mode approach," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 190-203, Jan. 2020.
- [38] M. He, J. L., "Resilient guaranteed cost control for uncertain T-S fuzzy systems with time-varying delays and Markov jump parameters," *ISA Trans.*, vol. 88, no. 12, pp. 12-22, May 2019.
- [39] W. Chen, D. Ding, H. Dong, G. Wei, "Distributed resilient filtering for power systems subject to denial-of-service attacks," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 49, no. 8, pp. 1688-1697, Aug. 2019.
- [40] D. Ding, Z. Wang, Q.L. Han, "A set-membership approach to event-triggered filtering for general nonlinear systems over sensor networks," *IEEE Trans. Autom. Contr.*, vol. 65, no. 4, pp. 1792-1799, Apr. 2020.
- [41] C. Chen, F.L. Lewis, S. Xie, H. Modares, Z. Liu, S. Zuo, A. Davoudi, "Resilient adaptive and H_∞ controls of multi-agent systems under sensor and actuator faults," *Automatica*, vol. 102, pp. 19-26, Apr. 2019.
- [42] S.M. LaValle, *Planning Algorithms*, Cambridge: Cambridge University Press, 2006.
- [43] H. Li, D. Liu, D. Wang, "Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 3, pp. 706-714, Jul. 2014.
- [44] X. Zhong, H. He, D. Wang, Z. Ni, "Model-free adaptive control for unknown nonlinear zero-sum differential game," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1633-1646, May 2018.
- [45] Q. Wei, D. Liu, Q. Lin, R. Song, "Adaptive dynamic programming for discrete-time zero-sum games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 957-969, Apr. 2018.