Event-Driven Off-Policy Reinforcement Learning for Control of Interconnected Systems

Vignesh Narayanan, *Member, IEEE*, Hamidreza Modares, *Senior Member, IEEE*, Sarangapani Jagannathan, *Fellow, IEEE*, and Frank L. Lewis, *Life Fellow, IEEE*

***Abstract*—In this article, we introduce a novel approximate optimal decentralized control scheme for uncertain input-affine nonlinear-interconnected systems. In the proposed scheme, we design a controller and an event-triggering mechanism (ETM) at each subsystem to optimize a local performance index and reduce redundant control updates, respectively. To this end, we formulate a noncooperative dynamic game at every subsystem in which we collectively model the interconnection inputs and the event-triggering error as adversarial players that deteriorate the subsystem performance and model the control policy as the performance optimizer, competing against these adversarial players. To obtain a solution to this game, one has to solve the associated Hamilton–Jacobi–Isaac (HJI) equation, which does not have a closed-form solution even when the subsystem dynamics are accurately known. In this context, we introduce an event-driven off-policy integral reinforcement learning (OIRL) approach to learn an approximate solution to this HJI equation using artificial neural networks (NNs). We then use this NN approximated solution to design the control policy and event-triggering threshold at each subsystem. In the learning framework, we guarantee the Zeno-free behavior of the ETMs at each subsystem using the exploration policies. Finally, we derive sufficient conditions to guarantee uniform ultimate bounded regulation of the controlled system states and demonstrate the efficacy of the proposed framework with numerical examples.**

***Index Terms*—Decentralized control, differential game, event-triggered learning, off-policy learning.**

I. Introduction

THE CONTROL of large-scale interconnected systems is an active area of research, and its application extends from engineering systems, such as smart grids, cyber-physical systems, manufacturing plants, traffic networks, and large-scale robotic systems to biomedical systems, such as neuronal networks, cancer models, and gene regulatory networks, to name a few. Several efficient control methods have been studied and reported in the literature for dealing with both the stabilization and the tracking problems associated with such interconnected systems when the subsystem dynamics are accurately known [1]–[7].

In this context, robust decentralized controllers have also been proposed (e.g., [8] and [9]), which used the small-gain theorem [10] to ensure that the controlled subsystems are input-to-state stable (ISS) and collectively lead to the overall system stability. In addition to stabilization, performance optimization is also essential for efficiently operating an interconnected system, as the deterioration of the transient performance in one subsystem might propagate to the other subsystems through their interconnections. To ensure the desired performance, optimal controllers have been developed for interconnected systems (e.g., see [2], [11] and the references therein).

On the other hand, approximate dynamic programming (ADP) with reinforcement learning (RL) [12], [13] has been extensively and successfully used to learn approximate optimal control policies for systems with uncertain nonlinear dynamics. For a single and stand alone system, the implementation of RL-based optimal adaptive control schemes has been established, and several results are available in the literature (see [14]–[18] and the references therein). However, for an interconnected system with uncertain dynamics, relatively fewer such results are available [19]–[21]. For instance, an adaptive optimal robust control scheme for large-scale interconnected systems based on the small-gain theorem was presented in [8], and decentralized approximate optimal control schemes were reported in [19] and [20].

Typically, adaptive and learning-based control schemes for interconnected systems are computationally intensive, and the presence of a communication network in the control-feedback loop increases the associated communication cost. Therefore, to reduce redundant computations and for better communication resource utilization, event-triggered feedback control frameworks have been proposed [4], [21]–[25]. The efforts in [23]–[25] proposed event-triggered controllers for a centralized system using the ADP approach. For interconnected systems, approximate optimal distributed controllers have been proposed (see [26] and the references therein) for optimizing the performance of the overall system with aperiodic feedback. However, the existing approaches consider the controller and the ETM design problems independently, and a

Manuscript received December 30, 2019; accepted April 15, 2020. This work was supported in part by NSF under Grant ECCS 1406533 and Grant CMMI 1547042. This article was recommended by Associate Editor D. Liu. *(Corresponding author: Vignesh Narayanan.)*

Vignesh Narayanan and Sarangapani Jagannathan are with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: vnxv4@umsystem.edu).

Hamidreza Modares is with the Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824 USA.

Frank L. Lewis is with the Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76118 USA.

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2020.2991166

unified decentralized approximate optimal control scheme in the event-triggered feedback framework for co-designing controllers and the ETMs for a large-scale interconnected system has not been reported in the literature.

In this article, we propose an optimization problem to co-design the control policy and the event-triggering instants at each subsystem. In particular, in our formulation, we define a cost functional at each subsystem using the subsystem states and control inputs, and introduce additional terms to model the effect of the interconnection inputs and the event-triggering error. Since the interconnection inputs and the event-triggering error degrade the subsystem performance, these additional terms are modeled as adversarial inputs, while the control policy is modeled to restore the subsystem performance. This leads to a novel game-theoretic scheme wherein the control policy tries to minimize a desired subsystem cost while competing against a team of two players. In doing so, the control policy explicitly accounts for the effect of the event-triggering error and the interconnections in the subsystem performance. A natural benefit of the proposed approach is that the optimal solution for the maximizing player, when used as the event-triggering threshold in the ETM, potentially elongates the interevent time without deteriorating the subsystem performance.

The solution to the proposed game can be obtained by solving a Hamilton–Jacobi–Isaacs (HJI) equation [27]–[29], which does not have a closed-form solution. To solve the HJI equation without requiring the complete knowledge of the subsystem dynamics, an RL-based controller with neural-network (NN) approximators is used at each subsystem. However, the commonly used on-policy RL schemes [12], [15] cannot be directly used to learn a solution to the proposed game because the actions of the adversaries (i.e., the interconnection inputs and the event-triggering error) cannot be assigned and updated arbitrarily to facilitate on-policy learning. This demands developing an off-policy RL [16], [30]–[32]-based scheme to obtain a solution to the proposed game.

The off-policy learning allows to separate the policies (both control and adversaries) that are applied to the system (also referred as the behavior policy), and the policies that are being evaluated and learned (also referred as the target/learning policy). Therefore, we develop an off-policy integral RL (OIRL) approach to learn a solution for the formulated game. Specifically, this off-policy approach will learn an approximate solution to the HJI equation at each subsystem, which will then be used to design the controller and the ETM for each subsystem. Due to the NN adaptation and interconnection terms, it is possible for the ETM to exhibit the Zeno behavior [22]. We demonstrate this possibility with an example, and then develop an event-triggering condition using the bounds of the exploration signal to both satisfy the persistence of excitation requirement, and to ensure the Zeno-free behavior of the ETM. Finally, sufficient conditions for local uniform ultimate bounded regulation of the overall system are determined for three cases: when the (behavior) policy applied to the system is—1) the optimal policy; 2) the adaptively learned greedy policy; and 3) an exploration policy.

Thus, the contributions of this article include the: 1) development of an approximate optimal decentralized control scheme using a game-theoretic formulation; 2) a unified design approach to obtain controllers and ETMs; 3) derivation of an event-triggering scheme using the exploration policy to ensure Zeno-free behavior; and 4) deriving sufficient conditions and the demonstration of closed-loop stability in the presence of uncertain dynamics with greedy and nongreedy policies.

In this article, we use standard mathematical notations; $\mathbb{R}$ and $\mathbb{N}$ denote the set of real and natural numbers, respectively. In the analysis, the Frobenius norm is used for matrices and the Euclidean norm is used for vectors. In the equations, the functional dependencies are not explicitly mentioned unless required.

II. Background

In this section, we begin by introducing the dynamics of the interconnected system considered in this article, followed by a brief background on the event-triggered feedback framework, ISS, and optimal control.

A. System Description

Consider an interconnected system composed of N subsystems, each with the dynamics given by

$$\dot{x}_i(t) = f_i(x_i) + g_i(x_i)u_i(t) + \sum_{j=1, j\neq i}^{N} \Delta_{ij}\big(x_i, x_j\big) \tag{1}$$

where $x_i(0) = x_{i0}$, $x_i \in \Omega_{ix} \subseteq \mathbb{R}^{n_i}$ is the state vector of the ith subsystem, $u_i : \Omega_{iu} \to \mathbb{R}^{m_i}$ is the feedback control policy, $f_i : \Omega_{ix} \to \mathbb{R}^{n_i}$ and $g_i : \Omega_{ix} \to \mathbb{R}^{n_i \times m_i}$ are nonlinear maps representing internal dynamics and input gain for the ith subsystem, and Δ_{ij} represents the interconnection between the ith and the jth subsystem. The sets Ω_{ix} and Ω_{iu} are subsets in $\mathbb{R}^{n_i}$ that describe a local neighborhood of the equilibrium point of interest, and without loss of generality, we consider the stabilization problem (using feedback) about the equilibrium point at origin. Since the control policy considered in this article is a feedback policy, we will also use the notation $u_i(x_i)$ in place of $u_i(t)$.

The following assumptions are used in the analysis presented in this article.

Assumption 1: The subsystems defined in (1) are stabilizable and the entire state vector is measurable [29]. Moreover, the order of each subsystem is known and the effect of computational delays and losses in the feedback channel are negligible.

Assumption 2: The interconnection function satisfies $\Delta_{ij}(x_i, x_j) = \Delta_{ij}(x_i)x_j$. There exist positive constants g_{im} and g_{iM}, such that $g_{im} \leq \|g_i\| \leq g_{iM}$, uniformly in Ω_{ix}, for $i = 1, \ldots, N$. The nonlinear maps f_i, g_i, and Δ_{ij} are locally Lipschitz continuous on compact sets.

The assumption on the interconnection dynamics (Assumption 2) does not restrict the application of the proposed control scheme. We present a numerical example to demonstrate the applicability of the proposed scheme for interconnected systems by relaxing Assumption 2, that is, when $\Delta_{ij}(x_i)x_j$ is replaced with $\Delta_{ij}(x_i, x_j)$.

In the event-triggered controller, the feedback is available to the controller at discrete sampling instants that are uniquely determined by an ETM, which is co-located with the sensors. These discrete sampling instants at the ith subsystem are represented as a monotonically increasing sequence of time instants $\{t_k^i\}_{k\in\{0,\mathbb{N}\}}$, with $t_0^i = 0$. Due to the lack of continuous feedback, the controllers are implemented with the most recent feedback. Therefore, the control input will be of the form $u_i(x_i(t)) = u_i(\breve{x}_i(t))$, where $\breve{x}_i(t) = x_i(t_k^i)$ for $t \in [t_k^i, {}^{i}_{k+1})$, $\forall k \in \{0, \mathbb{N}\}, i = 1, 2, \ldots, N$. This sporadic feedback will result in an error $e_i(t) = \breve{x}_i(t) - x_i(t)$ with $e_i(t) = 0$ at $t = t_k^i$ for each k and i. The subsystem dynamics in the event-triggered feedback framework are given by

$$\dot{x}_i(t) = f_i(x_i) + g_i(x_i)u_i(x_i) + \sum_{j=1, j\neq i}^{N} \Delta_{ij}(x_i)x_j + g_i(x_i)\hat{d}_{i0} \tag{2}$$

where $\hat{d}_{i0} = u_i(\breve{x}_i) - u_i(x_i)$. The overarching goal of this article is to develop a Zeno-free event-triggered control scheme for the uncertain interconnected system such that a desired performance function is optimized at each subsystem given in (2). Therefore, we will introduce the notions of stability, event triggering, and optimality next.

B. Stability and Optimality

The notion of ISS is reviewed first, followed by a brief review of event triggering and L_2 optimality conditions.

Definition 1 [34]: Consider the interconnected system (2) and assume that for each subsystem there exists a function $J_i : \mathbb{R}^{n_i} \to \mathbb{R}_+$ which is proper, positive definite, and Lipschitz continuous on $\mathbb{R}^{n_i}$ such that $J_i(0) = 0$. The function J_i is called an input-to-state practically stable (ISpS) Lyapunov function for the ith subsystem if there exist a function μ_i, a positive constant $\bar{z}_i \geq 0$, $\gamma_{ij} \in \mathbb{K}_\infty \cup \{0\}$ $\forall j \in \{1, \ldots, N\}, j \neq i, \phi_i \in \mathbb{K} \cup \{0\}$, and a positive-definite function α_i^k such that

$$J_i(x_i) \geq \mu_i(\gamma_{i1}(J_1(x_1)), \ldots, \gamma_{iN}(J_N(x_N)), \phi_i(\|e_i\|) + \bar{z}_i)$$

implies $\dot{J}_i \leq -\alpha_i^k(\|x_i\|)$, where $\mu_i : \mathbb{R}^{N+1} \to \mathbb{R}$, is a strictly positive function in its domains except at the origin, strictly increasing, and is radially unbounded; $\mathbb{K} = \{\hat{f} : \mathbb{R}_+ \to \mathbb{R}_+, \hat{f}$ is continuous, strictly increasing, and $\hat{f}(0) = 0\}$; $\mathbb{K}_\infty = \{\hat{f} \in \mathbb{K}, \hat{f}$ is unbounded$\}$ and $\mathbb{R}_+$ is the positive real line. When $\bar{z}_i = 0$, the Lyapunov function is ISS. The functions γ_{ij} and ϕ_i are called ISS Lyapunov gains.

A common design practice for finding an event-triggering condition that guarantees the stability of the system is to enforce

$$\phi_i(\|e\|) \leq \vartheta_i \alpha_i^\kappa(\|x_i\|) + \beta_i \tag{3}$$

in the interevent periods, where $0 \leq \vartheta_i < 1, \beta_i \geq 0$. This condition can be derived based on the stability analysis of the system during the interevent time [22]. Here, the error function due to event-triggered feedback ϕ_i is bounded by the sum of a dynamic threshold function $\vartheta_i \alpha_i^\kappa(\|x_i\|)$ and a constant β_i [22].

Definition 2 [22]: The trigger mechanism is called a relative trigger mechanism if [in (3)] $0 < \vartheta_i < 1$, and $\beta_i = 0$; an absolute trigger mechanism if $\vartheta_i = 0, \beta_i > 0$; and a mixed trigger mechanism if $0 < \vartheta_i < 1, \beta_i > 0$.

Next, to incorporate optimality, let the performance output to be controlled at each subsystem be defined as

$$\|\zeta_i(t)\|^2 = Q_i(x_i) + u_i^T(t)R_i u_i(t) \tag{4}$$

where $Q_i(.) \geq 0, R_i > 0$ such that $Q_i(0) = 0$. Now, observe that the subsystem dynamics (1) can be represented as

$$\dot{x}_i(t) = F_i(x_i, u_i) + \bar{\Delta}(x_i)d_i \tag{5}$$

where $F_i(x_i, u_i)$ represents the controlled dynamics $f_i + g_i u_i$, and d_i is an augmented vector of the adversarial inputs (interconnection inputs) to the ith subsystem with the interconnection gain $\bar{\Delta}_i$, each with appropriate dimensions.

Definition 3 [31]: Each subsystem as in (5) is said to have an L_2 gain less than or equal to σ_i from the interconnections to the controlled output (4), if for any initial state $x_i(0)$ and for an admissible control $u_i(t)$, the response of each subsystem corresponding to all $d_i \in L_2[0, \infty)$ satisfies

$$\int_t^\infty \|\zeta_i(\tau)\|^2 d\tau \leq \sigma_i^2 \int_t^\infty \|d_i(\tau)\|^2 d\tau + \bar{\kappa}(x_i(0)) \tag{6}$$

where σ_i represents the amount of attenuation from the interconnection input to the defined performance output, $\bar{\kappa}$ is a bounded positive semidefinite function, and L_2 is a set of square-integrable functions.

Having reviewed the notions of the L_2 gain condition, event-triggering conditions, and the definition of ISS, in the next section, we present our control design framework.

III. Controller Design

In this section, first, the proposed control scheme is introduced. Then, we present learning-based design strategies for realizing the proposed control scheme. Finally, artificial neural-networks (ANNs)-based implementation procedure and Zeno-freeness of the ETM are discussed.

A. Proposed Control Scheme—Game-Theoretic Formulation

Let the output to be controlled at each subsystem be defined as in (4). From Definition 3, each subsystem is said to have an L_2 gain less than or equal to σ_i, if for any initial state x_{i0} and control u_i, the response of each subsystem [with adversarial interconnection inputs, $x_j(t)$] satisfies

$$\int_t^\infty \|\zeta_i(\tau)\|^2 d\tau \leq \sigma_i^2 \int_t^\infty \sum_{j=1, j\neq i}^{N} \left\|x_j(\tau)\right\|^2 d\tau. \tag{7}$$

Using (7), an infinite horizon cost function for the ith subsystem can be defined as

$$J_i(\cdot) = \int_t^\infty \left[Q_i(x_i) + u_i^T R_i u_i - \sigma_i^2 \sum_{j=1, j\neq i}^{N} x_j^T x_j \right] d\tau. \tag{8}$$

The event-triggering error, in addition to the interconnection inputs can adversely affect the subsystem performance.

Therefore, they can be augmented to the cost, and the cost function can be redefined as

$$J_i(x_i, x_j, u_i) = \int_t^\infty \Bigg[Q_i(x_i) + u_i^T R_i u_i - \sigma_i^2 \sum_{\substack{j=1,\\ j\neq i}}^{N} x_j^T x_j - \sigma_{i0}^2 \hat{d}_{i0}^T \hat{d}_{i0} \Bigg] d\tau \quad (9)$$

where $\sigma_{i0} > 0$.

Remark 1: Note that minimizing (9) under the system dynamic constraint (2) is equivalent to finding a control policy at the ith subsystem that satisfies the bounded L_2 gain condition. However, directly solving this optimization problem is highly nontrivial as the adversarial inputs (i.e., the interconnection inputs x_j and the event-triggering error $\hat{d}_{i0}$) are not just dependent on the states of the ith subsystem. The interconnection inputs are the states of the neighboring subsystems with their own dynamics, and $\hat{d}_{i0}$ is dependent on the sampling instants.

Remark 2: In order to ensure a desired performance (4) at each subsystem, we propose an approximation-based learning scheme, where we will develop a solution to an optimization problem with the objective function (9) under the system dynamic constraint (2) by treating the adversarial inputs as independent players in a zero-sum game. Since the adversarial inputs acting on the subsystems cannot be directly modified for the purpose of learning the optimal state-dependent value function $V_i^*(x_i)$, we will present an off-policy learning scheme (in Section III-C) to learn an approximate optimal value function using which the performance of the controlled system (2) can be quantified.

In order to derive an approximation-based learning scheme, we will consider the following dynamics:

$$\dot{x}_i(t) = f_i(x_i) + g_i(x_i)u_i(x_i) + \sum_{\substack{j=1,\\ j\neq i}}^{N} \Delta_{ij}(x_i)d_j + g_i(x_i)d_{i0} \quad (10)$$

where $d_{i0}, d_j \in L_2[0, \infty)$ represent the adversarial inputs to the ith subsystem dynamics, and $x_i(0) = x_{i0}$. By defining the value function as $V_i(x_i)$ for any admissible policies (u_i, d_j, d_{i0}) [13] and taking the first derivative of the value function using the revised cost function given by

$$J_i(x_i, d_j, u_i) = \int_t^\infty [Q_i(x_i) + u_i^T R_i u_i - \sigma_i^2 \sum_{j=1, j\neq i}^{N} d_j^T d_j - \sigma_{i0}^2 d_{i0}^T d_{i0}] d\tau \quad (11)$$

along the system dynamics (10), we obtain the following Hamiltonian:

$$H_i = Q_i(x_i) + u_i^T R_i u_i - \sigma_i^2 \sum_{j=1, j\neq i}^{N} d_j^T d_j - \sigma_{i0}^2 d_{i0}^T d_{i0} + V_{ix}^T \Bigg[f_i(x_i) + g_i(x_i)u_i(x_i) + g_i(x_i)d_{i0} + \sum_{\substack{j=1\\ j\neq i}}^{N} \Delta_{ij}(x_i)d_j \Bigg] \quad (12)$$

where $V_{ix} = \partial V_i / \partial x_i$. Applying the stationarity condition [13], [29], that is, $\partial H_i / \partial u_i = 0$ and $\partial H_i / \partial d_j = 0$, gives the optimal control, the theoretical worst case interaction inputs, and the threshold policy as

$$u_i^* = -\frac{1}{2} R_i^{-1} g_i^T V_{ix}^* \quad (13)$$

$$d_j^* = \frac{1}{2\sigma_i^2} \Delta_{ij}^T V_{ix}^*, \ j = 1, 2, \ldots, N, j \neq i \quad (14)$$

$$d_{i0}^* = \frac{1}{2\sigma_{i0}^2} g_i^T V_{ix}^*, \ i = 1, 2, \ldots, N \quad (15)$$

where d_j^* and d_{i0}^* are the optimal adversarial inputs that can be injected into the ith subsystem (10). The optimal policy d_{i0}^* reveals a threshold for the event-triggering error to design the event-triggering instants for the system (2), and u_i^* provides an approximate optimal decentralized control policy for the subsystem in (2). Note that since the cost function (11) includes d_j and d_{i0} for $j = 1, \ldots, N, j \neq i$, which model the effect of both the interconnection input and the event-triggering error, this event-triggering condition and control policy explicitly accounts for the effects of adversarial inputs in the system performance.

Remark 3: The zero-sum game problem has a unique solution $[V_i^*(x_i)]$ under certain conditions, locally in the neighborhood of the origin, if the Nash condition holds [27]. Specifically, for a linear system, the infinite horizon zero-sum game admits a unique saddle point solution under certain conditions as stated in [27]. However, for a nonlinear system, the infinite horizon zero-sum game may not have a global solution. Moreover, even when a local solution exists, the solution V_i^* may be smooth only under the stabilizability and zero-state detectability conditions [35], [36]. In this article, we assume that such a local optimal solution exists.

Moving on, it should be noted that the solution to the proposed game can be derived by solving the associated HJI equation which does not have a closed-form solution. To obtain the solution to the game when the system dynamics are uncertain, a novel event-driven OIRL scheme is proposed to design approximate optimal controllers such that the controller learns the solution to the game online without *a priori* knowledge of the system dynamics. Finally, the optimal policies designed for the system (10) will be utilized to design the control and the ETM for (1), and the corresponding stability results will be derived.

B. Adaptive Optimal Controller—On-Policy Learning

By substituting the control policy (13), and the adversarial policies (14) and (15) in the Hamiltonian (12), the following HJI equation is obtained:

$$H_i^*\left(u_i^*, d_j^*, d_{i0}^*\right) \triangleq Q_i(x_i) - \frac{1}{4} V_{ix}^{*T} g_i R_i^{-1} g_i^T V_{ix}^* + \frac{1}{4\sigma_{i0}^2} V_{ix}^{*T} g_i g_i^T V_{ix}^* + \frac{1}{4\sigma_i^2} V_{ix}^{*T} \left(\sum_{j=1, j\neq i}^{N} \Delta_{ij} \Delta_{ij}^T \right) V_{ix}^* + V_{ix}^{*T} f_i(x_i(t)) = 0. \quad (16)$$

It can be shown that the control solution obtained by solving the HJI equation (16) will ensure that the controlled system (10) will satisfy the L_2 gain condition.

Lemma 1: Assume that the optimal value function for the ith subsystem (10), that is, $V_i^*(x_i) \in C^1$, a positive semidefinite solution to the HJI equation, exists for each subsystem. Then, $u_i^* = -(R_i^{-1}/2)g_i^T V_{ix}^*$, makes the closed-loop subsystem to have L_2 gain less than or equal to σ_i.

Proof: The proof for Lemma 1 follows the results in [31]. ∎

Since a closed-form solution to (16) does not exist, an on-policy RL-based approximate solution can obtained. However, after a brief introduction of on-policy learning, it will be revealed that the on-policy method cannot be used to solve the proposed game.

To begin with, differentiate $V_i^*(x_i)$ along the ith subsystem dynamics (10) with the optimal policies to obtain

$$\dot{V}_i^*(.) = V_{ix}^{*^T}\left[f_i(x_i) + g_i(x_i)\left[u_i^* + d_{i0}^*\right] + \sum_{\substack{j=1\\ j\neq i}}^{N} \Delta_{ij}(x_i) d_j^*\right].$$

Using the fact that $H_i^* = 0$ from (16), we obtain

$$\dot{V}_i^* = -Q_i(.) - u_i^{*T} R_i u_i^* + \sigma_i^2 \sum_{\substack{j=1\\ j\neq i}}^{N} d_j^{*T} d_j^* + \sigma_{i0}^2 d_{i0}^{*T} d_{i0}^*.$$

Integrating both sides in the interval $(t, t+T)$, for $T > 0$, yields the integral Bellman equation [13], [15], given by

$$\begin{aligned} V_i^*(x_i(t+T)) - V_i^*(x_i(t)) &= \int_t^{t+T} \Bigg(- Q_i(x_i) - u_i^{*T} R_i u_i^* \\ &\quad + \sigma_i^2 \sum_{j=1, j\neq i}^{N} d_j^{*T} d_j^* + \sigma_{i0}^2 d_{i0}^{*T} d_{i0}^* \Bigg) d\tau. \end{aligned} \quad (17)$$

The integral Bellman equation is also a consistency condition based on which the optimal policies can be learned in a forward-in-time manner. Let V_i be the estimate of the optimal value function V_i^*, and let the corresponding estimates of optimal policies be u_i, d_j, and d_{i0}. Using the estimates in (17), we obtain

$$\begin{aligned} E_i = \int_t^{t+T} \Bigg(Q_i(x_i) + u_i^T R_i u_i - \sigma_i^2 \sum_{j=1, j\neq i}^{N} d_j^T d_j \\ - \sigma_{i0}^2 d_{i0}^T d_{i0} \Bigg) d\tau + V_i(x_i(t+T)) - V_i(x_i(t)) \end{aligned} \quad (18)$$

where E_i is the TD error/Bellman error [13]. By updating V_i with an objective of reducing this error, V_i converges to V_i^* under certain conditions [15]. To realize such a learning scheme, the policies (u_i, d_j, d_{i0}) should be applied to the system, and the resulting change in the state (and the one-step cost or reward) should be used to calculate the Bellman error. The RL-based ADP schemes in which the policies that are being learned are the same as the policies that are applied to the system (behavior policy) for calculating the Bellman error are called on-policy RL schemes. However, it is desired that an exploratory control policy be used during the learning process for attaining optimality. Also, since $d_j \neq x_j$, and $d_{i0} \neq \hat{d}_{i0}$, the traditional on-policy RL-ADP schemes cannot be used to obtain the solution to (16), which can then be used for controlling the system (1). Next, we present an off-policy learning scheme that can mitigate this shortcoming.

C. Solution Using Off-Policy Learning

To develop an off-policy learning scheme, define the estimate of the approximate optimal value function as V_i^l and the associated policies as (u_i^l, d_j^l, d_{i0}^l), where the superscript l denotes the learning step with $l = 0, 1, \ldots$. The actual interconnection inputs $[x_j(t)]$ and the event-triggering error $[\hat{d}_{i0}(t)]$, which enters the ith subsystem, are different from d_j^l and d_{i0}^l. Similarly, if an exploratory control policy u_i is applied to the system, which is different from the estimated approximate optimal control policy u_i^l, the differences between the behavior policies, u_i, x_j, and $\hat{d}_{i0}$, and the learned policies, u_i^l, d_j^l, and d_{i0}^l, should be explicitly considered in the learning process. Adding and subtracting the estimates of the approximate optimal policies (u_i^l, d_j^l, d_{i0}^l) in the subsystem dynamics (2), we obtain

$$\begin{aligned} \dot{x}_i(t) = f_i(x_i) + g_i(x_i)\left(u_i^l + d_{i0}^l\right) + g_i(x_i)\left(u_i - u_i^l\right) \\ + g_i\left(\hat{d}_{i0} - d_{i0}^l\right) + \sum_{\substack{j=1\\ j\neq i}}^{N}\left[\Delta_{ij} d_j^l(t) + \Delta_{ij}(x_j(t) - d_j^l(t))\right]. \end{aligned} \quad (19)$$

Define the learning policies obtained using the stationarity condition similar to (13) but with the approximated value function, as

$$\begin{aligned} u_i^{l+1} &= -\frac{1}{2} R_i^{-1} g_i^T V_{ix}^l, \quad d_j^{l+1} = \frac{1}{2\sigma_i^2} \Delta_{ij}^T V_{ix}^l \\ d_{i0}^{l+1} &= \frac{1}{2\sigma_{i0}^2} g_i^T V_{ix}^l. \end{aligned} \quad (20)$$

To develop the integral equation to evaluate the temporal difference error using the off-policy scheme, differentiate $V_i^l(x_i)$ along the subsystem dynamics (19) to obtain

$$\begin{aligned} \dot{V}_i^l = V_{ix}^{l^T}\Bigg[f_i(x_i) + g_i(x_i)\left(u_i^l + d_{i0}^l\right) + \sum_{\substack{j=1\\ j\neq i}}^{N} \Delta_{ij} d_j^l \\ + g_i\left(u_i - u_i^l\right) + g_i\left(\hat{d}_{i0} - d_{i0}^l\right) + \sum_{\substack{j=1\\ j\neq i}}^{N} \Delta_{ij}\left(x_j - d_j^l\right) \Bigg]. \end{aligned}$$

Using the Hamiltonian (16) for the system (19) (with V_i^l, u_i^l, d_j^l, and d_{i0}^l), we obtain

$$\begin{aligned}\dot{V}_i^l = &-Q_i - u_i^{lT} R_i u_i^l + \sigma_i^2 \sum_{j=1, j\neq i}^{N} d_j^{lT} d_j^l + \sigma_{i0}^2 d_{i0}^{lT} d_{i0}^l \\ &- 2u_i^{l+1^T} R_i \left(u_i - u_i^l\right) + 2\sigma_i^2 \sum_{\substack{j=1 \\ j\neq i}}^{N} d_j^{l+1^T} \left(x_j - d_j^l\right) \\ &+ 2\sigma_{i0}^2 d_{i0}^{l+1^T} \left(\hat{d}_{i0} - d_{i0}^l\right). \end{aligned} \tag{21}$$

Integrating both sides in the interval $(t, t+T)$ yields the off-policy integral Bellman equation, given by

$$\begin{aligned}&V_i^l(x_i(t+T)) - V_i^l(x_i(t)) \\ &= 2\int_t^{t+T} \left(-u_i^{l+1^T} R_i \left(u_i - u_i^l\right) + \sigma_i^2 \sum_{\substack{j=1 \\ j\neq i}}^{N} d_j^{l+1^T} \left(x_j - d_j^l\right) \right. \\ &\qquad \left. + \sigma_{i0}^2 d_{i0}^{l+1^T} \left(\hat{d}_{i0} - d_{i0}^l\right) \right) d\tau \\ &\quad + \int_t^{t+T} \left(-Q_i - u_i^{lT} R_i u_i^l + \sigma_i^2 \sum_{\substack{j=1, \\ j\neq i}}^{N} d_j^{lT} d_j^l + \sigma_{i0}^2 d_{i0}^{lT} d_{i0}^l \right) d\tau. \end{aligned} \tag{22}$$

Lemma 2 [31]: The solution obtained for the optimal value function to the IRL Bellman equation (22) and the solution obtained for the optimal value function using the HJI equation (16) are same. Furthermore, the control policies and the interconnection policies evaluated using the off-policy IRL Bellman equation and the HJI equation are the same.

Theorem 1: The online event-based off-policy IRL scheme that uses (20) and (22) converges to optimal policies (13)–(15) and the value function satisfies the HJI equation (16) for each $i = 1, \ldots, N$.

Proof: Lemma 2 establishes that the off-policy IRL scheme and the standard on-policy algorithm converges to the same value function and the policies. Therefore, both the on-policy and the off-policy schemes have the same convergence properties. The detailed convergence results are available in [31].

In the following theorem, it is demonstrated that for a stabilizing event-triggering condition and an optimal control policy, the closed-loop system admits an ISS Lyapunov function. ■

Theorem 2 (Ideal Case): Consider the interconnected system (2). Let Assumptions 1 and 2 hold. Let an optimal control policy u_i^* as in (13) be applied to each subsystem, and let an event be triggered on violation of the condition

$$\Xi_i \|e_i\|^2 \leq \lambda_i \|x_i\|^2 \tag{23}$$

where $\Xi_i = L_{u_i}^2 \|R_i\|$, $1 > \lambda_i > 0$. Then, the ith subsystem and the overall system are ISS when Q_i, R_i, and σ_i are chosen such that $\bar{q}_i > 2\|R_i\|L_{u_i}^2 + \lambda_i \hat{\sigma}_i^2 (N-1)$, where $L_{u_i} > 0$, $\bar{q}_i > 0$ is the smallest singular value of the positive-definite matrix q_i satisfying $Q_i(x_i) = x_i^T q_i x_i$ and $\hat{\sigma}_i = \max\{\sigma_1, \ldots, \sigma_N\}$.

Proof: See the supplementary material. ■

Next, we present an ANN-based implementation of the proposed OIRL scheme.

D. NN Learning and the Role of Exploration in Event Triggering

In this section, we present an ANN-based implementation strategy for the proposed control scheme and derive sufficient conditions for system stability.

For the ease of exposition, define for each $i = 1, \ldots, N$, $X_i, D_i, \hat{D}_i^{l+1} \in \mathbb{R}^{N_j}$ as augmented vectors of $x_j, d_j, \hat{d}_j^l$, for $j = 1, \ldots, N$, $j \neq i$ and $N_j = \sum_{j=1, j\neq i}^{N} n_j$, respectively. Using the universal approximation theorem [37], the optimal value function and the associated optimal policies can be represented using parameterized NNs as $V_i^*(x_i) = W_{i1}^T \phi_{i1}(x_i) + \varepsilon_{i1}(x_i)$, $u_i^*(x_i) = W_{i2}^T \phi_{i2}(x_i) + \varepsilon_{i2}(x_i)$, $D_i^*(x_i) = W_{i3}^T \phi_{i3}(x_i) + \varepsilon_{i3}(x_i)$, and $d_{i0}^*(x_i) = W_{i03}^T \phi_{i03}(x_i) + \varepsilon_{i03}(x_i)$, where $W_{i\bullet}$ are NN weights, $\phi_{i\bullet}$ are activation functions, and $\varepsilon_{i\bullet}$ are bounded reconstruction errors with bounds denoted as $\varepsilon_{i\bullet M}$, each with appropriate dimensions. The estimate of the optimal value function and the optimal policies are defined as $\hat{V}_i^l(x_i) = \hat{W}_{i1}^T \phi_{i1}(x_i)$, $\hat{u}_i^{l+1}(x_i) = \hat{W}_{i2}^T \phi_{i2}(x_i)$, $\hat{D}_i^{l+1}(x_i) = \hat{W}_{i3}^T \phi_{i3}(x_i)$, and $\hat{d}_{i0}^{l+1}(x_i) = \hat{W}_{i03}^T \phi_{i03}(x_i)$, where $\hat{W}_{i\bullet}$ are the NN weight estimates, each with appropriate dimensions. Assume R_i for each i to be a diagonal matrix with entries $R_{i,k}$ for $k = 1, \ldots, m_i$.

Define $\nu_{i1}^l = u_i - \hat{u}_i^l$, $\nu_{i2}^l = X_i - \hat{D}_i^l$, and $\nu_{i3}^l = \hat{d}_{i0} - \hat{d}_{i0}^l$, and let $\nu_{il,k}$ for $l = 1, 2, 3$ denote the kth component of the vector ν_{il}. Use the OIRL Bellman equation (22), with estimated values and policies, to obtain the temporal difference error as

$$\begin{aligned}\mathrm{E}_i^l = &\hat{W}_{i1}^T \left[\phi_{i1}\left(x_i\left(t_{k+1}^i\right)\right) - \phi_{i1}\left(x_i\left(t_k^i\right)\right)\right] \\ &+ \int_{t_k^i}^{t_{k+1}^i} \left(Q_i + u_i^{lT} R_i u_i^l - \sigma_i^2 \sum_{\substack{j=1, \\ j\neq i}}^{N} d_j^{lT} d_j^l - \sigma_{i0}^2 d_{i0}^{lT} d_{i0}^l \right) d\tau \\ &+ 2\int_{t_k^i}^{t_{k+1}^i} \left(\sum_{k=1}^{m_i} \hat{W}_{i2,k}^T \phi_{i2} R_{i,k} \nu_{i1,k}^l - \sigma_i^2 \sum_{k=1}^{N_j} \hat{W}_{i3,k}^T \phi_{i3} \nu_{i2,k}^l \right. \\ &\qquad \left. - \sigma_{i0}^2 \sum_{k=1}^{n_i} \hat{W}_{i03,k}^T \phi_{i03} \nu_{i3,k}^l \right) d\tau. \end{aligned} \tag{24}$$

Define $\Delta\phi = \phi_{i1}(x_i(t_{k+1}^i)) - \phi_{i1}(x_i(t_k^i))$, $\hat{W}_i = [\hat{W}_{i1}^T, \hat{W}_{i2,1}^T, \ldots, \hat{W}_{i2,m_i}^T, \hat{W}_{i3,1}^T, \ldots, \hat{W}_{i3,N_j}^T, \hat{W}_{i03,1}^T, \ldots, \hat{W}_{i03,n_i}^T]^T$

$$\Phi_i = \begin{bmatrix} \Delta\phi \\ \int_{t_k^i}^{t_{k+1}^i} -2\phi_{i2}(x_i) R_{i,1} \nu_{i1,1}^l d\tau \\ \vdots \\ 2\sigma_i^2 \int_{t_k^i}^{t_{k+1}^i} \phi_{i3}(x_i) \nu_{i2,1}^l) d\tau \\ \vdots \\ 2\sigma_{i0}^2 \int_{t_k^i}^{t_{k+1}^i} \phi_{i03}(x_i) \nu_{i3,1}^l \\ \vdots \end{bmatrix}$$

and

$$Y_i = \int_{t_k^i}^{t_{k+1}^i} \left(-\|\zeta_i\|^2 + \sigma_i^2 \sum_{\substack{j=1, \\ j\neq i}}^{N} d_j^{l^T} d_j^l + \sigma_{i0}^2 d_{i0}^{lT} d_{i0}^l \right) d\tau.$$

To bring the TD error to its minimum value, rewrite the IRL Bellman equation as

$$E_i^l = \hat{W}_i^T \Phi_i - Y_i. \tag{25}$$

Note that this learning problem for obtaining a suitable set of NN weights satisfying the Bellman equation is a linear regression problem when the ANNs are selected as random vector functional link networks [37]. Next, we present two results. First, it will be demonstrated that using the greedy policy (from the learning scheme), the closed-loop system composed of the system state and weight estimation errors is locally uniformly ultimately bounded (UUB).

Theorem 3 (Greedy Policy): Consider the interconnected system (2). Let Assumptions 1 and 2 hold, and let V_i^* be the solution of the HJI equation (16). Let $u_i(x_i) = \hat{W}_{i2}^T \phi_{12}(x_i)$ be applied to each subsystem, and let an event be triggered on violation of the condition $L_{ui}\|e_i\| \leq \|d_{i0}\|$ with $L_{ui} > 0$ and $d_{i0} = \hat{W}_{i03}^T \phi_{i03}(x_i)$. Let the NN weights be initialized such that the resulting initial policy is admissible, and be updated using the learning rule

$$\dot{\hat{W}}_i = -T_i \frac{\Phi_i}{(1+\Phi_i{}^T \Phi_i)^2} E_i^{l^T} \tag{26}$$

where $T_i > 0, E_i^l$ is the Bellman error (25). Then, the ith subsystem, the overall system state vector, and the weight estimation errors are UUB as $k \to \infty$ when, $Q_i, R_i > 0$ and σ_i are chosen such that $\bar{q}_i > (N-1)\sigma_i^2$, where $\bar{q}_i > 0$ is the smallest singular value of the positive-definite matrix q_i satisfying $Q_i(x_i) = x_i^T q_i x_i$, and the bounds are defined as $B_{iM}^k = 8\lambda_M((\|R_i\|) + 4\sigma_i^2)\varepsilon_{iM}^2 + B_{\tilde{W}M}$ with $\|\varepsilon_{i.}\| \leq \varepsilon_{iM}$ and $\lambda_M(.)$ denoting the maximum eigenvalue operator. In addition, when the regression vector is persistently exciting (PE), and as $l, k \to \infty$, $V_i^l \to V_i^*$, the control policies $u_i^l \to u_i^*$ and the event-triggering threshold $d_{i0}^l \to d_{i0}^*$.

Proof: See the supplementary material. ∎

Remark 4: It can be observed from Theorem 3 that when the greedy policy is applied to the system, the weight estimation error and the system state vector remain bounded, and the bounds are functions of the design parameters R_i and σ_i, and the approximation error ε_i. By the appropriate design of NN, the approximation error can be reduced [37], and by choosing small R_i, σ_i, and T_i that satisfy the conditions derived in Theorem 3, the bounds for the regulation error can be made small. Furthermore, as the number of events increases, more data points along the system trajectory are collected to improve the weight estimation error.

Remark 5: Note that the Bellman error is computed with the latest states at each event-triggering instant, and it is used in the weight update rule. In the inter-event period, the weights are updated similar to the hybrid learning approach [21], to improve the estimated weights. The NN weight update rule (26) gives $\hat{V}_i^l$, $\hat{u}_i^{l+1}$, $\hat{D}_i^{l+1}$, $\hat{d}_{i0}^{l+1}$ after each event, and the update rule can be replaced with the concurrent learning rule or experience replay strategy, to overcome the requirement of persistency of the excitation condition [29]. However, since the objective of this article is to develop a decentralized learning control framework, we report our learning scheme with the update rule as in (26).

Existence of the Zeno Behavior: The event-triggering condition derived in Theorems 2 and 3 is a relative trigger condition (Definition 2). Since the interconnection inputs are influencing the subsystem dynamics, the relative trigger mechanism may introduce Zeno behavior. This can also happen if an additive external disturbance is injected to the system as shown in [22]. For example, for nonzero x_{i0} and for $t \in [0, t_1^i)$, if we have

$$-f_i - g_i u_i(x_i) - x_{i0} - \sum_{\substack{j=1 \\ j\neq i,p}}^{N} \Delta_{ij}(x_i) x_j = \Delta_{ip} x_p \tag{27}$$

for some $p \in \{1, \ldots, N\}, p \neq i$, there will be an accumulation point, resulting in the Zeno behavior. To see this, consider the event-triggering error dynamics in the interevent period, that is, using the definition, $e_i(t) = \breve{x}_i(t) - x_i(t)$, we have $\dot{e}_i(t) = -\dot{x}_i(t)$. Substituting the system dynamics, and using (27), we have $\dot{x}_i(t) = -x_{i0}$ which on integration reveals $x_i(t) = (1-t)x_{i0}$. Using the definition of the event-triggering error, we have for $t_0^i < t < t_1^i$, $e_i(t) = x_{i0} - (1-t)x_{i0} = t x_{i0}$. Using the relative event-trigger condition, we obtain an event when $\|e_i(t)\| \leq P_i\|x_i(t)\|$ is violated, where $P_i > 0$. Therefore, at $t = t_1^i$, one has $t_1^i \|x_{i0}\| = P_i(1-t_1^i)\|x_{i0}\|$, which can be simplified to get $t_1^i = [P_i/(1+P_i)] = 1 - [1/(1+P_i)] < 1$. Similarly, one can repeat this process [22] to obtain $t_k^i = 1 - [1/((1+P_i)^k)]$. This shows that there exists a limit for the sequence of event times at $t = 1$. Therefore, the ETM will exhibit Zeno behavior.

On the other hand, for the learning scheme, the behavior policies applied to the system need to be PE. This ensures that the data collected for the learning process contain the necessary information about the system. The PE condition can also be viewed as the condition which promotes "exploration" in the RL algorithm. Therefore, the control policy that is injected to the system is designed as

$$\hat{u}_i(t) = u_i(x_i(t)) + \delta_i(t) \tag{28}$$

where u_i is the greedy policy and $\delta_i(t)$ is a bounded signal, and it is an exploration signal that ensures that the PE condition is satisfied.

In the next theorem, sufficient conditions for the overall system stability are derived when the control policy is $\hat{u}_i(t)$.

Theorem 4 (Mixed Event Triggering): Consider the interconnected system (2). Let Assumptions 1 and 2 hold, and let V_i^* be the solution to the HJI equation (16). Let the explorative control policy (28) be applied to each subsystem, and let an event be triggered on violation of the condition

$$\left\| u_i\left(\breve{x}_i\right) - u_i(x_i) \right\| \leq \|d_{i0}(x_i)\| + \delta_{2M}$$

where $0 < \|\delta_2(t)\| \leq \delta_{2M}, \ \forall t \in \mathbb{R}_+$, and $\delta_2(t)$ are a bounded exploration signal. Let the NN weights be initialized such that the resulting initial policy is admissible, and be updated using the weight update rule (26). Then, the ith subsystem and the overall system, weight estimation errors are UUB when Q_i, R_i,

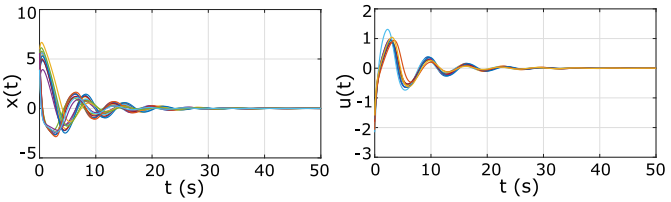

Fig. 1. Example 1: state (left) and control (right) trajectories and design parameters satisfying conditions in Theorem 2, case 1 ($b = 1$).

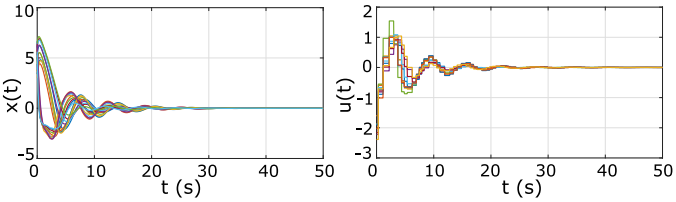

Fig. 2. Example 1: state (left) and control (right) trajectories and event-triggering condition satisfying (23), case 1 ($\beta = 1$).

and σ_i are chosen such that $\bar{q}_i > \sigma_i^2(N-1)$, and the bounds are obtained as $B_{iM}^k = 8\lambda_M(\|R_i\|)(\varepsilon_{i2M}^2 + \delta_{1M}) + 4\sigma_i^2(\varepsilon_{i03M}^2 + \delta_{2M}) + B_{\tilde{W}M}$. Furthermore, as $l, k \to \infty$, $V_i^l \to V_i^*$, the control policies $u_i^l \to u_i^*$ and the event-triggering threshold $d_{i0}^l \to d_{i0}^*$.

Proof: See the supplementary material. ■

Remark 6: Note that in the event-triggering condition, the exploration signal term $\delta_2(t)$ determines the minimum positive interevent time. Specifically, in the event-triggering condition, since $\|d_{i0}\| \geq 0$, we have $\|u_i(\breve{x}_i) - u_i(x_i)\| \leq \delta_{2M} \leq \|d_{i0}(x_i)\| + \delta_{2M}$. This ensures that the events are separated by the time it takes for the event-triggering error in the control to evolve from zero until $\|u_i(\breve{x}_i) - u_i(x_i)\| - \delta_{2M}$, which is a positive constant [22]. Furthermore, note that the bounds for the steady-state regulation errors are obtained as a function of the design parameters R_i, σ_i, δ_i, and T_i, and the approximation error ε_i. Since all these parameters are user defined, they can be chosen appropriately to reduce this bound.

Next, the simulation results for the proposed OIRL are presented.

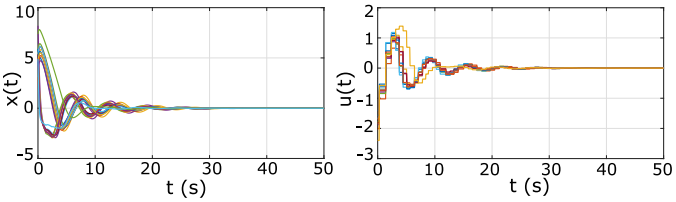

Fig. 3. Example 1: (case 2) state (left) and control (right) trajectories and event-triggering condition in Theorem 3.

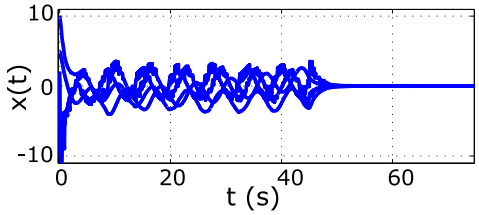

Fig. 4. Example 2: controlled state trajectories.

IV. Simulation Results

The proposed optimal adaptive control schemes are evaluated in this section using two examples.

Example 1: We considered a network of interconnected Van der Pol oscillators with a ring configuration with two types of coupling functions [38]. Specifically, we considered the dynamics

$$\ddot{x}_i(t) + \epsilon\left(x_i^2(t) - 1\right) + x_i(t) + \alpha u_i(t) = \beta(x_{i-1} - 2x_i + x_{i+1}),\ \ i = 1, \ldots, N \tag{29}$$

where $\epsilon > 0$. We considered ten interconnected oscillators, $\alpha, \beta = 1$ (case 1), and $\alpha = \sin(x_i)$, $\beta(.) = \sin(.)$ (case 2). Note that the input gain is constant, and the interconnection is a linear function of x_i for case 1 ($\alpha, \beta = 1$) while for case 2, it is a bounded nonlinear function. Furthermore, the interconnections are of the form $\Delta_{ij}(x_i, x_j)$ satisfying Assumptions 1 and 2 (for case 2, $\Delta_{ij}(.)$ satisfies the assumptions provided in [39]).

The design parameters were chosen as $R_i = 10$, $Q_i = \left[\begin{smallmatrix} 1 & 0.6 \\ 0.4 & 1 \end{smallmatrix}\right]$, $T_i = 10$, and $\sigma_i = 0.75$ for each subsystem. The approximate optimal cost function corresponding to each subsystem is estimated to be of the form $V_i^*(x_i) = W_1 x_{i1}^2 + W_2 x_{i2}^2 + W_3 x_{i1}x_{i2} + W_4 x_{i1}x_{i2}^2 + W_5 x_{i1}^2 x_{i2}$ with $W_k = \{2.5093, 2.6061, -0.9625, 0.4725, -0.1746\}$ for $k = 1, \ldots, 5$. The training is terminated, that is, the PE condition is removed once the Bellman error converged to a small bounded set with radius 10^{-3}. The state and control trajectories corresponding to this controller are plotted in Fig. 1.

Furthermore, using the design parameters, the event-triggering condition was obtained as $\|e_i\|^2 \leq 0.9\|x_i\|^2$. The resulting state and control trajectories are recorded in Fig. 2 for case 1 where $\alpha, \beta = 1$.

Without changing the design parameters, the control scheme was implemented for case 2, which corresponds to the case when $\alpha = \sin(.)$ and $\beta(x_i, x_j) = \sin(-2x_i + x_{i-1} + x_{i+1})$. Similar to case 1, the proposed controller stabilized the system with a similar control effort (Fig. 3).

Example 2: In this example, we considered the load frequency control application of a three area power system. The states of the power system model at each subsystem under consideration are—frequency change, incremental change in output power of the generator, change in the governor valve position, incremental change in integral control, and tie-line power deviation. For the detailed dynamics of the system considered see [11].

The controller design parameters were chosen as $R_1 = 0.01, R_2 = 0.01, q_i = 20$, $\sigma_i = 0.6$, and $T_i = 10$. For the value function, a single layer NN with inputs of the form $x_i^2, x_i x_j$ for $j \neq i$ were used, and for the policy approximators x_i was used as input. For the exploration policy, a sinusoidal signal was used.

Fig. 4 shows the state trajectories of the subsystems. The initial oscillations are induced by the sinusoidal exploratory signal. The event error versus the threshold function and the interevent time are seen in Figs. 5 and 6, respectively. These demonstrate the functioning of the ETM designed with positive interevent times.

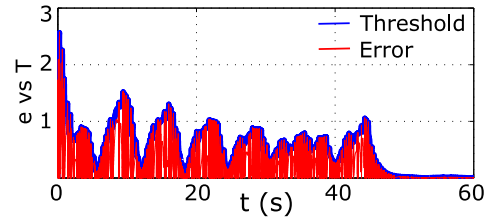

Fig. 5. Example 2: event-triggering error (e) versus threshold (T) at an ETM (subsystem 2).

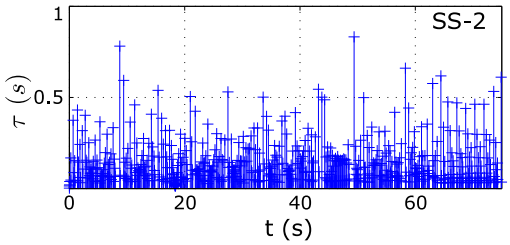

Fig. 6. Example 2: interevent times at subsystem 2 (SS-2).

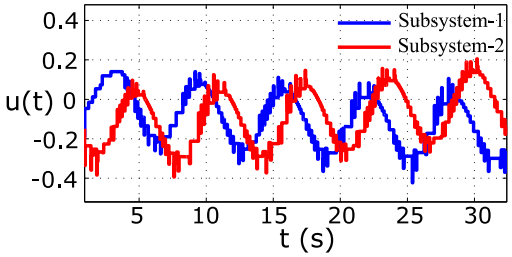

Fig. 7. Example 2: event-triggered behavior policy.

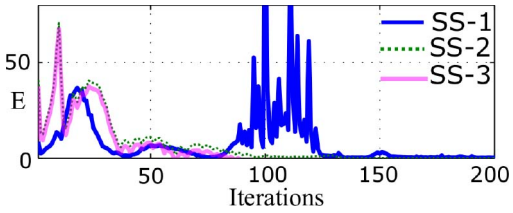

Fig. 8. Example 2: convergence of the Bellman error (E). SS-1, 2, 3 corresponds to subsystems 1, 2, and 3, respectively.

The event-triggered policy applied to the system during the learning phase is seen in Fig. 7. Due the event-based implementation, the control policy is piecewise continuous. The results from this example show that the stability of the system and convergence of the learning scheme (Fig. 8) are achieved as expected from the theoretical analysis.

Next, to analyze the performance of the proposed unified design, we considered an isolated nonlinear subsystem [18]. The design parameters were chosen similar to that in [18] for comparative analysis. Specifically, we compared the proposed unified design with the event-triggered control scheme presented in [23] for an isolated subsystem and the results are summarized in Table I. In Table I, the proposed method is abbreviated as PM, and the event-triggered NN-based approximate optimal control is abbreviated as ET-NN.

From the comparative analysis and the simulations presented in this section, it is observed that the proposed approach presents an alternative method for the emulation-based design of event-triggered controllers which is inclusive of the effects of external inputs in the form of interconnections (or disturbance for the isolated system). In the unified framework for designing both the controller and the event-triggering mechanism (ETM), the attenuation parameter and penalty matrices can be chosen to improve the performance of the system in the interevent period.

TABLE I
ANALYSIS OF THE APPROXIMATE OPTIMAL CONTROL SCHEME

σ	Avg. IET in s		Cumulative cost		Events	
	ET-NN	P.M	ET-NN	P.M	ET-NN	P.M
0.90	0.0350	0. 0685	2.074e+5	1.442e+5	290	120
0.925	0.0357	0.0675	2.085e+5	1.530e+5	280	130
0.95	0.0362	0.0744	2.100e+5	1.602e+5	270	140
0.975	0.0369	0.0583	2.110e+5	1.654e+5	260	160
1	0.0604	0.0562	2.114e+5	1.704e+5	260	170

V. CONCLUSION

In this article, we proposed a novel optimization problem at each subsystem of an interconnected system to co-design the subsystem control policy and the event-triggering conditions. Due to the natural connection between the proposed optimization problem and the differential game theory, we adopted a game-theoretic framework to solve this optimization problem, and to design controllers and event-triggering instants at each subsystem, which resulted in the UUB stability of the interconnected system. The NN approximation-based controller effectively regulated each subsystem in the interconnected system using event-triggered feedback and relaxed the requirement of an accurate knowledge of the system dynamics. In addition to the stability, the proposed scheme ensured that the performance of both the controller and the ETM was optimized as the event-triggering threshold designed using the Nash solution elongated the interevent time without deteriorating the system performance.

REFERENCES

[1] D. D. Šiljak, *Large-Scale Dynamic Systems : Stability and Structure*, vol. 2. New York, NY, USA: North-Holland, 1978,
[2] L. Bakule, "Decentralized control: An overview," *Annu. Rev. Control*, vol. 32, no. 1, pp. 87–98, 2008.
[3] D. Xue, A. Gusrialdi, and S. Hirche, "Robust distributed control design for interconnected systems under topology uncertainty," in *Proc. Amer. Control Conf. (ACC)*, Washington, DC, USA, 2013, pp. 6541–6546.
[4] X. Wang and M. D. Lemmon, "Event-triggering in distributed networked control systems," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 586–601, Mar. 2011.
[5] M. Kubisch, H. Karl, A. Wolisz, L. C. Zhong, and J. Rabaey, "Distributed algorithms for transmission power control in wireless sensor networks," in *Proc. IEEE Wireless Commun. Netw. (CNC 2003)*, vol. 1. New Orleans, LA, USA, 2003, pp. 558–563.
[6] C. De Persis, R. Sailer, and F. Wirth, "Parsimonious event-triggered distributed control: A zeno free approach," *Automatica*, vol. 49, no. 7, pp. 2116–2124, 2013.
[7] B. Luo, D. Liu, T. Huang, and D. Wang, "Model-free optimal tracking control via critic-only Q-learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 10, pp. 2134–2144, Oct. 2016.

[8] W. Gao, Y. Jiang, Z.-P. Jiang, and T. Chai, "Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming," *Automatica*, vol. 72, pp. 37–45, Oct. 2016.

[9] D. Liu, D. Wang, and H. Li, "Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 418–428, Feb. 2014.

[10] Z. Jiang, A. R. Teel, and L. Praly, "Small-gain theorem for ISS systems and applications," *Math. Control Signals Syst. (MCSS)*, vol. 7, no. 2, pp. 95–120, 1994.

[11] M. T. Alrifai, M. F. Hassan, and M. Zribi, "Decentralized load frequency controller for a multi-area interconnected power system," *Int. J. Elect. Power Energy Syst.*, vol. 33, no. 2, pp. 198–209, 2011.

[12] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, vol. 1. Cambridge, MA, USA: MIT Press, 1998.

[13] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. Hoboken, NJ, USA: Wiley, 2012.

[14] T. Bian and Z.-P. Jiang, "Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design," *Automatica*, vol. 71, pp. 348–360, Sep. 2016.

[15] D. Vrabie and F. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Netw.*, vol. 22, no. 3, pp. 237–246, 2009.

[16] B. Luo, H.-N. Wu, T. Huang, and D. Liu, "Data-based approximate policy iteration for affine nonlinear continuous-time optimal control design," *Automatica*, vol. 50, no. 12, pp. 3281–3290, 2014.

[17] R. Kamalapurkar, H. Dinh, S. Bhasin, and W. E. Dixon, "Approximate optimal trajectory tracking for continuous-time nonlinear systems," *Automatica*, vol. 51, pp. 40–48, Jan. 2015.

[18] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems using an online Hamilton-Jacobi-Isaacs formulation," in *Proc. 49th IEEE Conf. Decis. Control (CDC)*, Atlanta, GA, USA, Dec. 2010, pp. 3048–3053.

[19] S. Huang, K. K. Tan, and T. H. Lee, "Decentralized control of a class of large-scale nonlinear systems using neural networks," *Automatica*, vol. 41, no. 9, pp. 1645–1649, 2005.

[20] S. Mehraeen and S. Jagannathan, "Decentralized optimal control of a class of interconnected nonlinear discrete-time systems by using online Hamilton–Jacobi–Bellman formulation," *IEEE Trans. Neural Netw.*, vol. 22, no. 11, pp. 1757–1769, Nov. 2011.

[21] V. Narayanan and S. Jagannathan, "Event-triggered distributed approximate optimal state and output control of affine nonlinear interconnected systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2846–2856, Jul. 2018.

[22] D. N. Borgers and W. P. M. H. Heemels, "Event-separation properties of event-triggered control systems," *IEEE Trans. Autom. Control*, vol. 59, no. 10, pp. 2644–2656, Oct. 2014.

[23] A. Sahoo, H. Xu, and S. Jagannathan, "Approximate optimal control of affine nonlinear continuous-time systems using event-sampled neurodynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 639–652, Mar. 2017.

[24] X. Zhong and H. He, "An event-triggered ADP control approach for continuous-time system with unknown internal states," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 683–694, Mar. 2017.

[25] D. Wang and D. Liu, "Neural robust stabilization via event-triggering mechanism and adaptive learning technique," *Neural Netw.*, vol. 102, pp. 27–35, Jun. 2018.

[26] N. Vignesh, *Event-Triggered Near Optimal Adaptive Control of Interconnected Systems*. Rolla, MO, USA: Missouri Univ. Sci. Technol., 2017.

[27] T. Başar and P. Bernhard, *H_∞ Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. New York, NY, USA: Springer, 2008.

[28] I. Petersen, "Disturbance attenuation and H^∞optimization: A design method based on the algebraic riccati equation," *IEEE Trans. Autom. Control*, vol. 32, no. 5, pp. 427–429, May 1987.

[29] K. G. Vamvoudakis, H. Modares, B. Kiumarsi, and F. L. Lewis, "Game theory-based control system algorithms with real-time reinforcement learning: how to solve multiplayer games online," *IEEE Control Syst. Mag.*, vol. 37, no. 1, pp. 33–52, Feb. 2017.

[30] B. Luo, H.-N. Wu, and T. Huang, "Off-policy reinforcement learning for H_∞ control design," *IEEE Trans. Cybern.*, vol. 45, no. 1, pp. 65–76, Jan. 2015.

[31] H. Modares, F. L. Lewis, and Z.-P. Jiang, "H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2550–2562, Oct. 2015.

[32] Y. Jiang and Z.-P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.

[33] S. N. Dashkovskiy, B. S. Rüffer, and F. R. Wirth, "Small gain theorems for large scale systems and construction of ISS Lyapunov functions," *SIAM J. Control Optim.*, vol. 48, no. 6, pp. 4089–4118, 2010.

[34] M. Aliyu, *Nonlinear H-Infinity Control, Hamiltonian Systems and Hamilton-Jacobi Equations*. Boca Raton, FL, USA: CRC, 2017.

[35] A. J. Van Der Schaft, "L_2-gain analysis of nonlinear systems and nonlinear state-feedback H_∞ control," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 770–784, Jun. 1992.

[36] K. G. Vamvoudakis and F. L. Lewis, "Online solution of nonlinear two-player zero-sum games using synchronous policy iteration," in *Proc. 49th IEEE Conf. Decis. Control (CDC)*, Atlanta, GA, USA, Dec. 2010, pp. 3040–3047.

[37] F. Lewis, S. Jagannathan, and A. Yesildirak, *Neural Network Control of Robot Manipulators and Non-Linear Systems*. Boca Raton, FL, USA: CRC, 1998.

[38] M. A. Barrón and M. Sen, "Synchronization of four coupled van der pol oscillators," *Nonlinear Dyn.*, vol. 56, no. 4, p. 357, 2009.

[39] Y. Guo, D. J. Hill, and Y. Wang, "Nonlinear decentralized control of large-scale power systems," *Automatica*, vol. 36, no. 9, pp. 1275–1289, 2000.

[40] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.

Vignesh Narayanan (Member, IEEE) received the B.Tech. degree from SASTRA University, Thanjavur, India, the M.Tech. degree from the National Institute of Technology Kurukshetra, Kurukshetra, India, in 2012 and 2014, respectively, and the Ph.D. degree from the Missouri University of Science and Technology, Rolla, MO, USA, in 2017.

He is currently working as a Postdoctoral Research Associate with Washington University in St. Louis, St. Louis, MO. His research interests include dynamical systems, neural networks, and learning and adaptation in systems theory.

Hamidreza Modares (Senior Member, IEEE) received the B.Sc. degree from Tehran University, Tehran, Iran, in 2004, the M.Sc. degree from the Shahrood University of Technology, Shahrood, Iran, in 2006, and the Ph.D. degree from the University of Texas at Arlington (UTA), Arlington, TX, USA, in 2015.

From 2006 to 2009, he was with the Shahrood University of Technology as a Senior Lecturer. From 2015 to 2016, he was a Faculty Research Associate with UTA. From 2016 to 2018, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA. He is currently an Assistant Professor with the Department of Mechanical Engineering, Michigan State University, East Lansing, MI, USA. He has authored several journal and conference papers on the design of optimal controllers using reinforcement learning. His current research interests include cyber-physical systems, machine learning, distributed control, robotics, and renewable energy microgrids.

Dr. Modares was a recipient of the Best Paper Award from the 2015 IEEE International Symposium on Resilient Control Systems, the Stelmakh Outstanding Student Research Award from the Department of Electrical Engineering, UTA, in 2015, and the Summer Dissertation Fellowship from UTA in 2015. He is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Sarangapani Jagannathan (Fellow, IEEE) received the bachelor's degree in electrical engineering from Anna University, Chennai, India, in 1986, the M.S. degree in embedded systems and robotics from the University of Saskatchewan, Saskatoon, SK, Canada, in 1989, and the Ph.D. degree from the University of Texas, TX, USA, in 1994.

He is with the Missouri University of Science and Technology, Rolla, MO, USA (formerly, University of Missouri–Rolla), where he is a Rutledge-Emerson Distinguished Professor of electrical and computer engineering. He served as a Site Director for the NSF Industry/University Cooperative Research Center on Intelligent Maintenance Systems for 13 years. He has coauthored with his students 171 peer-reviewed journal articles, 287 refereed IEEE conference articles, several book chapters, authored/co-edited six books, received 21 U.S. patents, one patent defense publication, and several pending. He graduated 29 Doctoral and 31 M.S. thesis students, and his total funding is in excess of $17.5 million with over $10 million toward his shared credit from federal and industrial entities. He was a co-editor for the IET book series on control from 2010 to 2013 and currently serves many editorial boards including IEEE SYSTEMS, MAN AND CYBERNETICS. His research interests include neural network control, adaptive event-triggered control/secure human–cyber–physical systems, prognostics, and autonomous systems/robotics.

Prof. Jagannathan received many awards, including the 2018 IEEE CSS Transition to Practice Award, the 2007 Boeing Pride Achievement Award, the 2000 NSF Career Award, the 2001 Caterpillar Research Excellence Award, and has been on the organizing committees of several IEEE conferences. He is a fellow of the National Academy of Inventors, the Institute of Measurement and Control, U.K., and the Institution of Engineering and Technology, U.K.

Frank L. Lewis (Life Fellow, IEEE) received the bachelor's degree in physics/electronics engineering and the M.S.E.E. degree from Rice University, Houston, TX, USA, in 1971, the M.S. degree in aeronautical engineering from the University of West Florida, Pensacola, FL, USA, in 1977, and the Ph.D. degree from Georgia Tech, Atlanta, GA, USA, in 1981.

He is currently the Moncrief-O'Donnell Chair with the University of Texas at Arlington (UTA) Research Institute, Fort Worth, TX. He has authored seven U.S. patents, numerous journal special issues, 420 journal papers, 20 books, including the textbooks *Optimal Control, Aircraft Control and Simulation: Dynamics, Controls Design, and Autonomous Systems*, *Optimal Estimation: With an Introduction to Stochastic Control Theory*, and *Robot Manipulator Control: Theory and Practice*. He works in feedback control, intelligent systems, cooperative control systems, and nonlinear systems.

Dr. Lewis received the Fulbright Research Award, the NSF Research Initiation Grant, the ASEE Terman Award, the International Neural Network Society Gabor Award, the U.K. Institute of Measurement and Control Honeywell Field Engineering Medal, the IEEE Computational Intelligence Society Neural Networks Pioneer Award, the AIAA Intelligent Systems Award, and the AACC Ragazzini Award. He is a fellow of the National Academy of Inventors, IFAC, AAAS, and U.K. Institute of Measurement and Control, a Professional Engineer in Texas, a U.K. Chartered Engineer, a UTA Distinguished Scholar Professor, and a UTA Distinguished Teaching Professor Engineer.