Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 12, No. 2, 2024, 149-161.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Event-Triggered Optimal Adaptive Leader-Follower Consensus Control for Unknown Input-Constrained Discrete-Time Nonlinear Systems

Zahra Jahan, Abbas Dideban* and Farzaneh Tatari

Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran.

Article Info

Article History:
Received 21 November 2023
Revised 08 April 2024
Accepted 05 June 2024

DOI:10.22044/jadm.2024.13779.2500

Keywords:
Adaptive Optimal Control, Event-Triggering Scheme, Optimal Leader-follower Consensus, Reinforcement Learning, Neural Networks.

**Corresponding author:*
adideban@semnan.ac.ir (A. Dideban).

Abstract

This paper introduces an adaptive optimal distributed algorithm based on event-triggered control to solve multi-agent discrete-time zero-sum graphical games for unknown nonlinear constrained-input systems with external disturbances. Based on the value iteration heuristic dynamic programming, the proposed algorithm solves the event-triggered coupled Hamilton-Jacobi-Isaacs equations assuming unknown dynamics to develop distributed optimal controllers and satisfy leader-follower consensus for agents interacting on a communication graph. The algorithm is implemented using the actor-critic neural network, and unknown system dynamics are approximated using the identifier network. Introducing and solving nonlinear zero-sum discrete-time graphical games in the presence of unknown dynamics, control input constraints and external disturbances, differentiate this paper from the previously published works. Also, the control input, external disturbance, and the neural network's weights are updated aperiodic and only at the triggering instants to simplify the computational process. The closed-loop system stability and convergence to the Nash equilibrium are proven. Finally, simulation results are presented to confirm theoretical findings.

1. Introduction

Due to the numerous applications of distributed control of multi-agent systems (MAS) in various engineering fields, researchers have become interested in it in the last few decades [1, 2].

The leaderless and leader-follower consensus problems are two types of consensus issues that distributed control approaches have recently addressed [3-5]. In the leader-follower consensus [6-8], a favorite in this article, the states of all agents synchronize to a leader state, whereas in the leaderless consensus [9, 10], all agents synchronize to a standard value. Leader-follower consensus in discrete-time (DT) MASs has been the subject of numerous studies [11, 12]. There is no guarantee that the studies mentioned above will be optimal. However, the game-theoretic framework can be utilized to accomplish this.

The optimal multi-agent control problems, in which each agent tries to optimize their performance index and obtain their optimal policy, are well suited for research in game theory [13]. Since many real-world MASs problems have external disturbances, solving multi-agent games with unknown external disturbances is critical because neglecting this can lead to performance degradation and instability. Graphical games [14], first developed for continuous time (CT) systems, are used to solve optimally distributed leader-follower consensus problems in linear [14-16] and nonlinear [17] systems, where every follower's performance index, actions, and tracking error dynamics are dependent on neighbors local information. Zero-sum nonlinear differential graphical games of CT systems, which considered the existence of external disturbances, were added to these games [18, 19]. Several studies have been done in DT games on two-player zero-sum linear and nonlinear games [20-22]. Additionally, linear

N-player DT graphical games have been studied by [23, 24].
To solve multi-agent games that include external disturbances, it is necessary to solve the coupled Hamilton-Jacobi-Isaacs (HJI) equations. However, solving these nonlinear equations is either impossible or extremely challenging. Consequently, these equations can be solved using approximation-based approaches.
Reinforcement learning (RL) algorithms [25] have many applications in various problems [26] and have recently been developed as effective approximation-based techniques for solving multi-agent games [27, 28].
The work [20] employed heuristic dynamic programming (HDP) techniques to solve nonlinear zero-sum games with known dynamics. Reference [23] introduced a value iteration (VI) algorithm to solve linear DT graphical games under systems' known dynamics. Moreover, most physical systems are nonlinear, with higher-order dynamics that are difficult to model precisely. Also, saturation nonlinearity exists in many actuators and should be considered in the obtained controller design [29-31] since ignoring it can cause performance degradation or system instability. Therefore, solving multi-player games under unknown dynamics and input constraints is very important.
Most of the references mentioned above use time-based methods that perform calculations at all instants. This process is time-consuming and increases the complexity of calculations. Due to the significant reduction in calculations, the event-triggered control method has recently attracted much attention from researchers [32-34]. Event-triggered schemes for DT systems without considering optimal control problems in [35] are presented. Reference [36] employed an HDP-based event-triggered adaptive control method for unknown nonlinear DT systems. Reference [37] deals with event-triggered adaptive control for DT zero-sum games and these references have not considered graphical games and input constraints.
Several issues discriminate DT zero-sum graphical games from DT zero-sum [20, 21, 22] and DT graphical games [23, 24]. In DT zero-sum graphical games, distributed optimal controllers are developed to satisfy optimal leader-follower consensus for agents interacting on a communication graph. Every follower's performance index, actions, and local error dynamics are dependent on neighbors' local information. But in zero-sum games, the communication of agents is not under the graph topology and the influence of neighboring agents is not considered. They do not consider the leader-follower synchronization problem and are leaderless. On the other hand, solving HJI equations is required in the DT zero-sum game to determine the worst-case disturbance and the best controller. In DT zero-sum graphical games, coupled HJI equations should be solved to develop distributed optimal controllers to satisfy leader-follower consensus for agents interacting on a communication graph, while solving coupled HJI equations due to the existence of graphical coupling terms is quite different and rather difficult than solving HJI equations. Furthermore, DT zero-sum graphical games engage external disturbance as an opposite player against the controller for every agent and need to solve coupled HJI equations which are different from Hamilton-Jacobi-Belman equations in graphical games, which neglect the existence of external disturbance.
We are aware that no results have been reported on DT zero-sum graphical games that can solve the N-player leader-follower synchronization problem for unknown dynamics nonlinear agents with input constraints and external disturbances. Also, DT graphical games are not solved by the event-triggered method.
This paper introduces nonlinear multi-agent DT zero-sum graphical games. It suggests a new adaptive optimal distributed algorithm based on event-triggered control to solve these games for constrained-input unknown nonlinear systems with external disturbances. The following are the article's main contributions:

- The first is the introduction of DT zero-sum graphical games for nonlinear DT MASs that use the local information of neighbor agents to achieve the leader-follower synchronization problem.
- A VI HDP algorithm based on event-triggered control for solving multi-agent DT zero-sum graphical games in an online and distributed fashion is proposed.
- Using the event-triggered method and performing calculations non-periodically, which leads to a reduction of calculations and execution time.
- The presented algorithm solves the games under the assumption of unknown dynamics where every agent's unknown dynamics are identified using an identifier.
- Constraints on control inputs and external disturbance existence are considered to make the proposed algorithm more applicable to real-world problems.
- Finally, The stability of the closed-loop system and its convergence to the game's approximate Nash equilibrium is demonstrated.

Structure: This paper has the following sections. In the next section, the required knowledge of graphs and problem formulation is explained. Then, a multi-player DT zero-sum graphical game based on event-triggered with control input constraints and unknown dynamics is proposed. The presented graphical game's event-triggered Nash solution with proof of its convergence and the proposed event-triggered optimal distributed algorithm with unknown dynamics and external disturbances are presented afterward. The identifier employed for every agent unknown dynamics is described and an actor-critic structure used for the proposed algorithm is presented. The simulation study is presented, and then the conclusions are drawn.

2. Graphs and Problem Formulation

This section presents background on graphs, formulation of optimal leader-follower consensus control, and the triggering mechanism.

2.1. Graphs

The directed graph $Gr=(P,\Sigma)$ provides a description of the interactive topology for N agents' information exchange, where a set of graph nodes is $P=\{p_0,.....,p_N\}$ and a set of graph edges is $\sum \subset P\times P$. $C=[c_{ij}]\in R^{N\times N}$ represents an adjacency matrix for the graph such that $c_{ij}>0$ if $(p_j,p_i)\in\sum$, otherwise $c_{ij}=0$ where (p_j,p_i) means that agent i can get information from agent j but not necessarily vice versa. The list of node p_i's neighbors is shown with $N_i=\{p_j:(p_j,p_i)\in\sum\}$. $Q=diag\{q_i\}$ denotes the in-degree matrix of Gr with $q_i=\sum_{j\in N_i}c_{ij}$. The Laplacian matrix of the graph is given by $L=Q-C$. The leader agent is denoted by 0 and information can be transmitted from the leader to its neighbors. The leader pinning matrix is indicated by $A=diag\{a_i\}\in R^{N\times N}$, where $a_i\geq 0$ is the pinning gain. If the agent i is connected to the leader, it is non-zero, otherwise it is equal to zero. Note that the leader is connected to at least one of the follower nodes. A graph has a spanning tree if there is a directed path from an agent called the root to all other agents. In this article, it is assumed that the graph has a spanning tree.

2.2. Problem Formulation

On the communication graph Gr with N follower agents, the local dynamic of agent i is as follows

$$s_i(t+1)=f_i(s_i(t))+g_i(s_i(t))u_i(t)+h_i(s_i(t))\omega_i(t) \tag{1}$$

where $s_i(t)\in R^n$, $u_i(t)\in R^m$ and $\omega_i(t)\in R^k$ are the state, control input, and external disturbance vector of agent i, respectively. $f_i(s_i)\in R^n$, $g_i(s_i)\in R^{n\times m}$ and $h_i(s_i)\in R^{n\times k}$ are the drift, input, and disturbance dynamics of the agent i, respectively, which are all considered unknown in our developments. The leader dynamic s_0, indicating a reference trajectory, is as follows

$$s_0(t+1)=f_0(s_0(t)) \tag{2}$$

f_0 is the state matrix of the leader.

Synchronizing all follower agents' states to the leader is the goal of the leader-follower consensus. The local error of agent i [38] is defined as

$$\mathrm{E}_i(t)=\sum_{j\in N_i}c_{ij}(s_j(t)-s_i(t))+a_i(s_0(t)-s_i(t)) \tag{3}$$

The network local error for all agents is as

$$\mathrm{E}(t)=-((L+A)\otimes I_n)(s(t)-\bar{s}_0(t)) \tag{4}$$

where $\bar{s}_0(t)=\left[s_0^T(t),\ldots,s_0^T(t)\right]^T\in R^{nN}$, $s=\left[s_1^T,\ldots,s_N^T\right]^T\in R^{nN}$, $\mathrm{E}=\left[\mathrm{E}_1^T,\ldots,\mathrm{E}_N^T\right]^T\in R^{nN}$. I_n is the unit matrix of $n\times n$ and $\otimes$ represents Kronecker multiplication.

The disagreement error vector is as

$$\theta(t)=s(t)-\bar{s}_0(t)\in R^{nN} \tag{5}$$

If a root node is connected to the leader and the graph has a spanning tree, $(L+A)$ is nonsingular [38].

Remark1. It is demonstrated in [23] that if $(L+A)$ is nonsingular, the disagreement error vector is bounded as

$$\|\theta(t)\|\leq\|\mathrm{E}(t)\|/\underline{\sigma}(L+A) \tag{6}$$

where $\underline{\sigma}(.)$ is a matrix's smallest singular value. Therefore, leader-follower synchronization can be achieved by keeping the local error small.

For simplicity, $s_i(t)$ is written as s_{it} from now on, and other variables are considered similarly.

We describe a series of incremental time instants $\{t_m^i\}_{m=0}^{\infty}$ for each agent i to create an event-triggered control situation, where t_m^i is the m-th sampling instant with $t_0^i=0$ and $t_m^i<t_{m+1}^i$. To reduce the computational complexity, in a new sampled state $t_0^i,t_1^i,....$, weights of the neural

networks, the control input, and the external disturbance are updated.
Sampling instants are obtained in violation of the following condition for agent i at $t = t_m^i$.

$$\|e_i(t)\| \le \varepsilon_i \quad (7)$$

where ε_i is the threshold, and $e_i(t)$ is the event-triggered error as follows

$$e_i(t) = \mathrm{E}_i(t_m^i) - \mathrm{E}_i(t), \forall t \in [t_m^i, t_{m+1}^i) \quad (8)$$

Where $E_i(t_m^i)$ and $E_i(t)$ are the local error in the sampled state and current state, respectively.

Remark 2. In the sampled state $t = t_m^i$, event-triggered error $e_i(t)$ is reset to zero. Meanwhile, the control input and external disturbance of agent i are adjusted in line with the new measurement and remain constant in $[t_m^i, t_{m+1}^i)$, i.e $u_i(t) = u_i(t_m^i), \omega_i(t) = \omega_i(t_m^i), \forall t \in [t_m^i, t_{m+1}^i)$.

Assumption 1. For each agent i, positive constants D_i exist, such that

$$\|e_i(t+1)\| \le D_i \|e_i(t)\| + D_i \|\mathrm{E}_i(t)\| \quad (9)$$

Theorem 1. For DT nonlinear systems, the event-triggered error must satisfy the following condition:

$$\|e_i(t)\| \le \varepsilon_i = \frac{1-(2D_i)^{t-t_m^i}}{1-2D_i} D_i \|\mathrm{E}_i(t_m^i)\| \quad (10)$$

Proof. Inspired by [36] for each $t \in [t_m^i, t_{m+1}^i)$ it is obvious that $e_i(t+1) = \mathrm{E}_i(t_m^i) - \mathrm{E}_i(t+1)$, so that $\|e_i(t+1)\| \le \|\mathrm{E}_i(t+1)\|$ and so we have

$$\|e_i(t+1)\| \le D_i \|e_i(t)\| + D_i \|\mathrm{E}_i(t)\| \quad (11)$$

Then, by substituting (8) in (11), we obtain

$$\|e_i(t+1)\| \le 2D_i \|e_i(t)\| + D_i \|\mathrm{E}_i(t_m^i)\| \quad (12)$$

Therefore

$$\begin{aligned} \|e_i(t)\| &\le 2D_i \|e_i(t-1)\| + D_i \|\mathrm{E}_i(t_m^i)\| \\ &\le 2D_i(2D_i \|e_i(t-2)\| + D_i \|\mathrm{E}_i(t_m^i)\|) \\ &+ D_i \|\mathrm{E}_i(t_m^i)\| \\ &\ldots\ldots \\ &\le (2D_i)^{t-t_m^i} \|e_i(t_m^i)\| + (2D_i)^{t-t_m^i-1} D_i \|\mathrm{E}_i(t_m^i)\| \\ &+ \ldots. + D_i \|\mathrm{E}_i(t_m^i)\| \end{aligned} \quad (13)$$

Considering the initial conditions and $e_i(t_m^i) = 0$ simplifying (13), we have

$$\|e_i(t)\| \le \left(\frac{1-(2D_i)^{t-t_m^i}}{1-2D_i} D_i \|\mathrm{E}_i(t_m^i)\| = \varepsilon_i \right) \quad (14)$$

Therefore, the proof is complete.

By using Equations (1) and (3) the event-triggered local error of agent i for $t \in [t_m^i, t_{m+1}^i)$ is obtained as

$$\begin{aligned} \mathrm{E}_{i(t+1)} &= \sum_{j \in N_i} c_{ij} (f_j(s_{jt}) - f_i(s_{it})) \\ &+ a_i (f_0(s_{0t}) - f_i(s_{it})) \\ &- (q_i + a_i)(g_i(s_{it}) u_{i(t_m^i)} + h_i(s_{it}) \omega_{i(t_m^i)}) \\ &+ \sum (c_{ij}(g_j(s_{jt}) u_{j(t_m^j)} + h_j(s_{jt}) \omega_{j(t_m^j)}) \end{aligned} \quad (15)$$

To obtain leader-follower synchronization, a distributed controller for each player is proposed to minimize Equation (15) for $\omega_i \ne 0$, assuming that the dynamics of the system are unknown.

3. DT Zero-Sum Graphical Game by Event-Triggered Control

In this section, we introduce nonlinear DT zero-sum graphical games with disturbance and control constraints. The definition of these games is based on local error (15) and by introducing the local performance index.

Considering constant control input and constant disturbance between trigger instants for $t \in [t_m^i, t_{m+1}^i)$, the local performance index of agent i is as follows

$$\begin{aligned} J_i &= \sum_{\cup_m [t_m^i, t_{m+1}^i) = [0,\infty)} \sum_{t=t_m^i}^{t_{m+1}^i} U_i(\mathrm{E}_{it}, u_{i(t_m^i)}, u_{-i(t_m^i)}, \omega_{i(t_m^i)}, \omega_{-i(t_m^i)}) \\ &= \sum_{\cup_m [t_m^i, t_{m+1}^i) = [0,\infty)} \sum_{t=t_m^i}^{t_{m+1}^i} \frac{1}{2} \begin{pmatrix} \mathrm{E}_{it}^T O_i \mathrm{E}_{it} + W(u_{i(t_i^m)}) \\ + \sum_{j \in N_i} W(u_{j(t_j^m)}) - \gamma^2 \omega_{i(t_i^m)}^T T_{ii} \omega_{i(t_i^m)} \\ - \gamma^2 \sum_{j \in N_i} \omega_{j(t_j^m)}^T T_{ij} \omega_{j(t_i^m)} \end{pmatrix} \end{aligned} \quad (16)$$

where $O_i > 0 \in R^{n \times n}$, $T_{ii} > 0 \in R^{k \times k}$, $T_{jj} > 0 \in R^{k \times k}$, $\gamma > 0$ is a prescribed constant and $W(.) > 0$. The control and disturbance of the neighbors of agent i are $u_{-i} = \{u_j \mid j \in N_i\}$ and $\omega_{-i} = \{\omega_j \mid j \in N_i\}$, respectively.

For each player, the following nonquadratic functional [39] is used to take into account the control input constraints

$$W(u_{it}) = 2\int_0^{u_i} \phi^{-T}(\overline{Y}^{-1} x) \overline{Y} y dx \quad (17)$$

where $y > 0$ is a diagonal positive definite matrix, $\phi^{-1}(u_{it}) = [\psi^{-1}(u_{it}^1) \psi^{-1}(u_{it}^2) \ldots. \psi^{-1}(u_{it}^b)]^T$, $x \in R^m$, $\phi \in R^m$, where u_{it}^z is the z-th element of the vector u_{it}, $z = 0, \ldots, b$. $\psi(.)$ is a monotonic odd bounded function satisfying $|\psi(.)| \le 1$ and its first derivative is bounded. $\bar{Y}$ is a bound for actuators. In this paper, $\psi(.) = \tanh(.)$.

Each agent 's value function is defined as

$$V_i(\mathrm{E}_{it}) = \sum_{d=t}^{\infty} U_i(\mathrm{E}_{id}, u_{id}, u_{-id}, \omega_{id}, \omega_{-id}) \tag{18}$$

3.1 Bounded L2-Gain Synchronization Problem

Gaining a control policy u_{it} is desirable for solving the synchronization problem of DT zero-sum graphical games when $\omega_{it} = 0$ and it satisfies the following bounded L_2-gain condition for a given $\gamma > \gamma^*$ when $\omega_{it} \ne 0$ for all players

$$\sum_{\cup_m [t_m^i, t_{m+1}^i) = [0,\infty)} \sum_{t=t_m^i}^{t_m^{i+1}} \begin{pmatrix} \mathrm{E}_{it}^T O_i \mathrm{E}_{it} + W(u_i(t_m^i)) \\ + \sum_{j \in N_i} W(u_j(t_m^j)) \end{pmatrix} \tag{19}$$

$$\le \gamma^2 \sum_{t \in (t_m^i, t_{m+1}^i, \dots, \infty)} \begin{pmatrix} \omega_{i(t_m^i)}^T T_{ii} \omega_{i(t_m^i)} \\ + \sum_{j \in N_i} \omega_{j(t_m^j)}^T T_{ij} \omega_{j(t_m^j)} \end{pmatrix} + \beta(\mathrm{E}_{i0})$$

for several bounded functions β such that $\beta(0) = 0$. Let γ^* is the smallest amount that satisfies the above bounded L_2-gain condition.

3.2. Event-Triggered Bellman Equation for DT Zero-Sum Graphical Games

Considering the control input and disturbance at the trigger instants and the first difference of Equation (18), the Bellman equation of each agent i for $t \in [t_m^i, t_{m+1}^i)$ is obtained as follows

$$V_i(\mathrm{E}_{it}) = U_i(\mathrm{E}_{it}, u_{i(t_m^i)}, u_{-i(t_m^i)}, \omega_{i(t_m^i)}, \omega_{-i(t_m^i)}) + V_i(\mathrm{E}_{i(t+1)}), V_i(0) = 0 \tag{20}$$

The game's purpose is to find a unique saddle point $(u_{it}^o, \omega_{it}^o)$ for every agent such that

$$V_i^o(\mathrm{E}_{ik}) = \min_{u_i} \max_{\omega_i} \begin{pmatrix} U_i(\mathrm{E}_{it}, u_{i(t_m^i)}, u_{-i(t_m^i)}, \omega_{i(t_m^i)}, \omega_{-i(t_m^i)}) \\ + V_i^o(\mathrm{E}_{i(t+1)}) \end{pmatrix} \tag{21}$$

In our event-triggered control method, the event-triggered optimal control and disturbance policies are only updated when the events are triggered. As a result, they are updated using the sampled state $s_{i(t_m^i)}$ and $\mathrm{E}_{i(t_m^i)}$ rather than the actual state s_{it} and E_{it} when $t \in [t_m^i, t_{m+1}^i)$. Therefore, the bounded event-triggered optimal control and event-triggered worst disturbance policies for each agent i can be respectively obtained by minimizing and maximizing Equation (20) concerning control input u_{it} and disturbance ω_{it} as follows

$$\begin{aligned} u_i^o(t) &= u_i^o(t_m^i) \\ &= \arg\min_{u_{it}} \begin{pmatrix} U_i(\mathrm{E}_{it}, u_{i(t_m^i)}, u_{-i(t_m^i)}, \omega_{i(t_m^i)}, \omega_{-i(t_m^i)}) \\ + V_i^o(\mathrm{E}_{i(t+1)}) \end{pmatrix} \\ &= \bar{Y}\phi\left(\left(\bar{Y}y\right)^{-1}\left(q_i + a_i\right) g_i^T(s_{i(t_m^i)}) \nabla V_i^o\left(\mathrm{E}_{i(t_m^i+1)}\right)\right) \end{aligned} \tag{22}$$

$$\begin{aligned} \omega_i^o(t) &= \omega_i^o(t_m^i) \\ &= \arg\max_{\omega_{it}} \begin{pmatrix} U_i(\mathrm{E}_{it}, u_{i(t_m^i)}, u_{-i(t_m^i)}, \omega_{i(t_m^i)}, \omega_{-i(t_m^i)}) \\ + V_i^o(\mathrm{E}_{i(t+1)}) \end{pmatrix} \\ &= -\frac{1}{\gamma^2}(q_i + a_i) T_{ii}^{-1} h_{it}^T(s_{i(t_m^i)}) \nabla V_i^o\left(\mathrm{E}_{i(t_m^i+1)}\right) \end{aligned} \tag{23}$$

where $\nabla V_i(\mathrm{E}_{i(t+1)}) = \partial V_i(\mathrm{E}_{i(t+1)}) / \partial \mathrm{E}_{i(t+1)}$.

Substituting Equations (22) and (23) into Bellman Equation (20), the following event-triggered coupled Bellman optimality equations are gained

$$V_i^o(\mathrm{E}_{it}) = \frac{1}{2} \begin{pmatrix} \mathrm{E}_{it}^T O_{ii} \mathrm{E}_{it} \\ + W(u_{i(t_m^i)}^o) + \sum_{j \in N_i} W(u_{j(t_m^j)}^o) \\ -\frac{1}{\gamma^2}(q_i + a_i)^2 \nabla V_i^o(\mathrm{E}_{i(t_m^i+1)})^T \\ \times h_i(s_{i(t_m^i)}) T_{ii}^{-1} h_i^T(s_{i(t_m^i)}) \nabla V_i^o(\mathrm{E}_{i(t_m^i+1)}) \\ -\frac{1}{\gamma^2} \sum_{j \in N_i} (q_j + a_j)^2 \nabla V_j^o(\mathrm{E}_{j(t_m^j+1)})^T \\ \times h_j(s_{j(t_m^j)}) T_{jj}^{-1} T_{ij} T_{jj}^{-1} h_j^T(s_{j(t_m^j)}) \nabla V_j^o(\mathrm{E}_{j(t_m^j+1)}) \end{pmatrix} + V_i^o(\mathrm{E}_{i(t+1)}), V_i^o(0) = 0 \tag{24}$$

3.3. Event-Triggered Coupled Hamilton-Jacobi-Isaacs Equation

Based on (15) and (18) each agent's Hamiltonian function for $t \in [t_m^i, t_{m+1}^i)$ is defined as

$$\begin{aligned} &H_i\left(E_{it}, \nabla V_i\left(E_{i(t+1)}\right), u_{i(t_m^i)}, u_{-i(t_m^i)}, \omega_{i(t_m^i)}, \omega_{-i(t_m^i)}\right) = \\ &\nabla V_i\left(E_{i(t+1)}\right)^T \begin{pmatrix} \sum_{j \in N_i} c_{ij}(f_j(s_{jk}) - f_i(s_{ik})) \\ + a_i(f_0(s_0) - f_i(s_{ik})) \\ -(q_i + a_i)(g_i(s_{it}) u_{i(t_m^i)} + h_i(s_{ik}) \omega_{i(t_m^i)}) \\ + \sum_{j \in N_i} (c_{ij}(g_j(s_{jt}) u_{j(t_m^j)} + h_j(s_{jt}) \omega_{j(t_m^j)}) \end{pmatrix} \\ &+ U_i(\varepsilon_{it}, u_{i(t_m^i)}, u_{-i(t_m^i)}, \omega_{i(t_m^i)}, \omega_{-i(t_m^i)}) = 0, V_i(0) = 0 \end{aligned} \tag{25}$$

By employing the $\frac{\partial H_i}{\partial u_{it}} = 0$ and $\frac{\partial H_i}{\partial \omega_{it}} = 0$, the bounded optimal control policy and the worst disturbance are obtained for each agent i respectively as follows:

$$\begin{aligned} u_{it}^* &= u_{i(t_m^i)}^* \\ &= \arg\min_{u_{it}}(H_i) \\ &= \bar{Y}\phi\left(\left(\bar{Y}y\right)^{-1}(q_i+a_i)\, g_i^T(s_{i(t_m^i)})\nabla V_i^*\left(\mathrm{E}_{i(t_m^i+1)}\right)\right) \end{aligned} \tag{26}$$

$$\begin{aligned} \omega_{it}^* &= \omega_{i(t_m^i)}^* \\ &= \arg\max_{\omega_{it}}(H_i) \\ &= -\frac{1}{\gamma^2}(q_i+a_i)T_{ii}^{-1}h_i^T(s_{i(t_m^i)})\nabla V_i^*\left(\mathrm{E}_{i(t_m^i+1)}\right) \end{aligned} \tag{27}$$

Substituting Equation (26) and Equation (27) into Equation (25), one obtains the following event-triggered coupled DT HJI equations

$$\begin{aligned} &H_i\left(\mathrm{E}_{it}, \nabla V_i^*\left(\mathrm{E}_{i(t+1)}\right), u_{i(t_m^i)}^*, u_{-i(t_m^i)}^*, \omega_{i(t_m^i)}^*, \omega_{-i(t_m^i)}^*\right) = \\ &\nabla V_i^*\left(\mathrm{E}_{i(t+1)}\right)^T \\ &\times\begin{pmatrix} \sum_{j\in N_i} c_{ij}\left(f_j(s_{jt}) - f_i(s_{it})\right) + a_i\left(f_0(s_0) - f_i(s_{it})\right) \\ -(q_i+a_i)\left(\bar{Y}g_i(s_{it})\right)\phi\begin{pmatrix}\left(\bar{Y}y\right)^{-1}(q_i+a_i) \\ \times g_i^T(s_{it_m^i})\nabla V_i^*\left(\mathrm{E}_{i\left(t_m^i+1\right)}\right)\end{pmatrix} \\ -h_i(s_{it})\frac{1}{\gamma^2}(q_i+a_i)T_{ii}^{-1}h_i^T(s_{it_m^i})\nabla V_i\left(\mathrm{E}_{i\left(t_m^i+1\right)}\right) \\ +\sum_{j\in N_i} c_{ij}\left(\bar{Y}g_j(s_{jt})\right)\phi\begin{pmatrix}\left(\bar{Y}y\right)^{-1}(q_j+a_j) \\ \times g_j^T(s_{jt_m^j})\nabla V_j^*\left(\mathrm{E}_{j\left(t_m^j+1\right)}\right)\end{pmatrix} \\ -\frac{1}{\gamma^2}h_j(s_{jt})(q_j+a_j)T_{jj}^{-1}h_j^T(s_{jt_m^j})\nabla V_j^*\left(\mathrm{E}_{j\left(t_m^j+1\right)}\right) \end{pmatrix} \\ &+\frac{1}{2}\begin{pmatrix} \mathrm{E}_{it}^T O_i \mathrm{E}_{it} + W\left(u_{i(t_m^i)}^*\right) + \sum_{j\in N_j} W\left(u_{j(t_m^j)}^*\right) \\ -\frac{1}{\gamma^2}(\mathrm{q}_i+\mathrm{a}_i)^2\nabla V_i^{*T}\left(\mathrm{E}_{i\left(t_m^i+1\right)}\right) \\ \times h_i(s_{it_m^i})T_{ii}^{-1}h_i^T(s_{it_m^i})\nabla V_i^*\left(\mathrm{E}_{i\left(t_m^i+1\right)}\right) \\ -\frac{1}{\gamma^2}\sum_{j\in N_i}(q_j+a_j)^2\nabla V_j^{*T}\left(\mathrm{E}_{j\left(t_m^j+1\right)}\right) \\ \times h_j(s_{jt_m^j})T_{jj}^{-1}\mathrm{T}_{ij}T_{jj}^{-1}h_j^T(s_{jt_m^j})\nabla V_j^*\left(\mathrm{E}_{j\left(t_m^j+1\right)}\right) \end{pmatrix} = 0 \end{aligned} \tag{28}$$

The obtained event-triggered bounded optimal control policy and the worst disturbance require the solution of the event-triggered coupled Bellman optimality equations (24) or event-triggered DT HJI equations (28). These equations are equivalent [23] and difficult or impossible to solve. Therefore, the VI algorithm is employed to approximate the solutions.

4. Event-Triggered Nash Equilibrium of DT Zero-Sum Graphical Game

To solve the multi-agent DT graphical games for unknown nonlinear constrained-input systems in the presence of external disturbances, one needs to find the Nash equilibrium solutions by solving the event-triggered coupled HJI equations. Therefore, in the next step, the Nash equilibrium condition is defined.

Definition 1 The N-player games have an event-triggered global Nash equilibrium solution if for all agents

$$\begin{aligned} &J_i\left(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}, \omega_{-it}^*\right) \le J_i\left(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*\right) \\ &\le J_i\left(\mathrm{E}_{it}, u_{it}, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*\right) \end{aligned} \tag{29}$$

Theorem 2. Let $V_i^*(\varepsilon_{it}) < 0$ satisfies Equation (24) or Equation (28). Let for $t \in [t_m^i, t_{m+1}^i)$ the condition $(\mathrm{E}_{it}^T O_{ii}\mathrm{E}_{it} + W(u_{it_m^i}) + \sum_{j\in N_i} W(u_{jt_m^j}) \le \gamma^2(\omega_{it_m^i}^T T_{ii}\omega_{it_m^i} + \sum_{j\in N_i}\omega_{jt_m^j}^T T_{ij}\omega_{jt_m^j})$ holds. Let control and disturbance policies are respectively given by Equation (26) and Equation (27) in terms of V_i^*. Let the communication graph has a spanning tree and $a_i \ne 0$ for at least one agent. Then, the local error (15) are asymptotically stable and all agents' states synchronize to the leader state.

Proof. Consider $-V_i^*(\mathrm{E}_{it}) > 0$ as Lyapunov function for (15), and from (20) we have

$$\begin{aligned} &-(V_i^*(\mathrm{E}_{i(t+1)}) - V_i^*(\mathrm{E}_{it})) \\ &= U_i^*(\mathrm{E}_{it}, u_{i(t_m^i)}^*, u_{-i(t_m^i)}^*, \omega_{i(t_m^i)}^*, \omega_{-i(t_m^i)}^*) < 0 \end{aligned} \tag{30}$$

Thus, the dynamics of the local error (15) are asymptotically stable and according to Remark1, all agents' states synchronize to the leader state.

In the following theorem, it is proved that solutions to the event-triggered coupled DT HJI equations provide Nash equilibrium solutions and solve the DT zero-sum graphical game.

Theorem 3. let $V_i^*(\mathrm{E}_{it})$ satisfies Equation (24) or Equation (28) such that the dynamics of the local error (15) are asymptotically stable. Let control and disturbance policies are respectively given by Equation (26) and Equation (27) in terms of V_i^*. Then, all agents constitute a Nash equilibrium, inequality (29) is satisfied and the optimal performance index of each agent i is expressed as

$$J_i^*\left(\mathrm{E}_{ih}, u_{ih}^*, u_{-ih}^*, \omega_{ih}^*, \omega_{-ih}^*\right) = V_i^*(\mathrm{E}_{ih})$$

Proof. According to Theorem 1, the closed-loop system (Equation (15)) is asymptotically stable and $\mathrm{E}_i(\infty) \to 0$. Therefore $V_i^*(\mathrm{E}_i(\infty)) = 0$ and

$$J_i\left(\mathrm{E}_{ih}, u_{ih}, u_{-ih}, \omega_{ih}, \omega_{-ih}\right) = V_i^*(\mathrm{E}_i(\infty)) + \sum_{t=h}^{\infty} U_i(\mathrm{E}_{ih}, u_{ih}, u_{-ih}, \omega_{ih}, \omega_{-ih}) \tag{31}$$

Rearranging Equation (31) results in

$$J_i\left(\{\mathrm{E}_{ih}, u_{ih}, u_{-ih}, \omega_{ih}, \omega_{-ih}\}\right) = V_{ih}^*(\mathrm{E}_{it}) + \sum_{t=h}^{\infty}(U_i(\mathrm{E}_{it}, u_{it}, u_{-it}, \omega_{it}, \omega_{-it}) - U_i^*(E_{it}, u_{it}^*, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*)) \tag{32}$$

Assume that $u_{-id} = u_{-id}^*$, $\omega_{id} = \omega_{id}^*$, $\omega_{-id} = \omega_{-id}^*$ for the local performance index. Then, one has

$$J_i\left(\{\mathrm{E}_{ih}, u_{ih}, u_{-ih}^*, \omega_{ih}^*, \omega_{-ih}^*\}\right) = V_{ih}^*(\mathrm{E}_{it}) + \sum_{t=h}^{\infty}(U_i(\mathrm{E}_{it}, u_{it}, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*) - U_i^*(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*)) \tag{33}$$

The optimal control policy minimizes the function of the local performance index. Thus, it is clear that

$$\sum_{t=h}^{\infty} U_i(E_{it}, u_{it}, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*) - \sum_{t=h}^{\infty} U_i(E_{it}, u_{it}^*, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*) = \sum_{t=h}^{\infty} W(u_{it}) - W(u_{it}^*) \geq 0 \tag{34}$$

Therefore, according to Equation (33) and Equation (34), we obtain

$$J_i\left(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*\right) \leq J_i\left(\mathrm{E}_{it}, u_{it}, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*\right) \tag{35}$$

Then, for each function $V_i(\mathrm{E}_{il}), i \in N$, we can write Equation (16) as

$$\begin{aligned} J_i\left(E_{ih}, u_{ih}, u_{-ih}, \omega_{ih}, \omega_{-ih}\right) &= \frac{1}{2}\sum_{t=0}^{M} E_{it}^T O_{ii} E_{it} + W(u_{it}) + \sum_{j \in N_i} W(u_{jt}) \\ &- \gamma^2 \omega_{it}^T T_{ii} \omega_{it} - \gamma^2 \sum_{j \in N_i} \omega_{jt}^T T_{ij} \omega_{jt} + V_i(\mathrm{E}_{it}) \\ &+ \sum_{t=0}^{\infty} \nabla V_i(\mathrm{E}_{i(t+1)})^T \times \begin{pmatrix} \sum_{j \in N_i} c_{ij}(f_j(s_{jt}) - f_i(s_{it})) \\ a_i(f_0(s_0) - f_i(s_{it})) \\ -(q_i + a_i)(g_i(s_{it})u_{it} + h_i(s_{it})\omega_{it}) \\ + \sum_{j \in N_i} (c_{ij}(g_j(u_{jt})u_{jt} + h_j(s_{jt})\omega_{jt}) \end{pmatrix} \end{aligned} \tag{36}$$

Let now that V_i^* satisfies the DT HJI (28), u_i^* and ω_i^* are respectively the optimal policies and the disturbance given by Equation (26) and Equation (27), one can write Equation (36) as

$$\begin{aligned} J_i\left(\delta_{ih}, u_{ih}, u_{-ih}, \omega_{ih}, \omega_{-ih}\right) = \; & V_i^*(E_{it}) + \frac{1}{2}(W(u_{it}) - W(u_{it}^*)) \\ & + \frac{1}{2}(\sum_{j \in N_i} W(u_{jt}) - W(u_{jt}^*)) \\ & + \nabla V_i^T(E_{i(t+1)}) \sum_{j \in N_i} c_{ij} g_j(s_{jt})(u_{jt}^* - u_{jt}) \\ & + \nabla V_i^T(E_{i(t+1)})(q_i + a_i)(g_i(s_{it})(u_{it}^* - u_{it}) \\ & - \frac{1}{2}\gamma^2 \sum_{j \in N_i} (\omega_{jt} - \omega_{jt}^*)^T T_{ij}(\omega_{jt} - \omega_{jt}^*) \\ & - \frac{1}{2}\gamma^2 \left((\omega_{it} - \omega_{it}^*) T_{ij}(\omega_{it} - \omega_{it}^*)\right) \\ & + \nabla V_i^T(E_{i(t+1)}) \sum_{j \in N_i} c_{ij} h_j(s_{jt})(\omega_{jt}^* - \omega_{jt}) \\ & - \gamma^2 \sum_{j \in N_i} \left(\omega_{jt}^*\right)^T T_{ij}(\omega_{jt} - \omega_{jt}^*) \end{aligned} \tag{37}$$

Assigning $u_{jt} = u_{jt}^*$, $\omega_{jt} = \omega_{jt}^*$, $\forall j \in N_i$ and $u_{it} = u_{it}^*$, one has

$$J_i\left(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}, \omega_{-it}^*\right) = V_i^*(\mathrm{E}_{it}) - \frac{1}{2}\gamma^2 \sum_{t=0}^{\infty} (\omega_{it} - \omega_{it}^*)^T T_{ii}(\omega_{it} - \omega_{it}^*) \tag{38}$$

Therefore, it is shown that the left inequality of (29) is satisfied

$$J_i\left(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}, \omega_{-it}^*\right) \leq J_i\left(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*\right) \tag{39}$$

Setting $\omega_{it} = \omega_{it}^*$ in Equation (38) yields

$$J_i\left(\mathrm{E}_{it}, u_{it}^*, u_{-it}^*, \omega_{it}^*, \omega_{-it}^*\right) = V_i^*(\mathrm{E}_{it}) \tag{40}$$

Therefore, the proof is complete.

5. Event-Triggered Value Iteration HDP Algorithm for DT Zero-Sum Graphical Games

In this part, an online event-triggered value iteration HDP algorithm based on event-triggered Bellman equations is presented to obtain the answers of the DT zero-sum graphical game. Algorithm 1 solves the event-triggered coupled Bellman optimality equations to obtain optimal values, control policies, and disturbance policies.

Algorithm 1
Event-Triggered Value Iteration Algorithm for DT Zero-Sum Graphical Games

(Initialization). Give arbitrary initial control and disturbance policies and values for all agents.
(Policy evaluation). Solve the following equation for $t \in [t_m^i, t_{m+1}^i)$ to obtain V_i^{l+1}

$$V_i^{l+1}(E_{it}) = U_i(E_{it}, u^l_{i(t_m^i)}, u^l_{-i(t_m^i)}, \omega^l_{i(t_m^i)}, \omega^l_{-i(t_m^i)}) \quad (41)$$
$$+V_i^l(\mathrm{E}_{i(t+1)})$$

(Policy improvement). Update the disturbance and control policies at the triggering instants by the following equations

$$u_{it}^{l+1} = u_{i(t_m^i)}^{l+1} \quad (42)$$
$$= \bar{Y}\phi\left(\left(\bar{Y}y\right)^{-1}\left(q_i + a_i\right) g_i^T(s_{i(t_m^i)})\nabla V_i\left(E_{i\left(t_m^i+1\right)}\right)^{l+1}\right)$$

$$\omega_{it}^{l+1} = \omega_{i(t_m^i)}^{l+1} \quad (43)$$
$$= -\frac{1}{\gamma^2}(q_i + a_i)T_{ii}^{-1}h_i^T(s_{i(t_m^i)})\nabla V_i(E_{i(t_m^i+1)})^{l+1}$$

For all i, when $\left\|V_i^{l+1}(E_{it}) - V_i^l(E_{it})\right\| \le \alpha$ end. α is a small constant.
l denotes the iteration index.

6. Neural Network-Based System Approximation

In this study, the drift, input, and disturbance dynamics of each agent are assumed to be unknown and we approximated unknown system dynamics using the neural-network-based identification technique as follows

$$s_i(t+1) = W_{is}^T\varphi_{is}(v_{is}^T z_{is}(t)) + \varepsilon_{is}(t) \quad (44)$$

where $\varphi_{is}(.)$ is the activation function and it is presumed to be bounded, $\varepsilon_{is}(t)$ is the NN estimation error, $z_{is}(t) = [s_{it}^T \quad u_{it}^T \quad w_{it}^T]^T \in R^d$ is the NN input with $d = n+m+k$, v_{is}^T and W_{is}^T denote the ideal weight matrix between the input layer and the hidden layer, the hidden layer and the output layer, respectively.

In the system identification process, v_{is}^T is assumed to be constant and only W_{is}^T is adjusted. Hence, the identifier network output is described as

$$\hat{s}_i(t+1) = \hat{W}_{is}^T\varphi_{is}(Z_{is}(t)) \quad (45)$$

where $Z_{is}(t) = v_{is}^T z_{is}(t)$, $\hat{s}_i$ is the estimated state vector and $\hat{W}_{is}$ is the estimation of W_{is}.
The approximation error for system identification is defined as

$$e_{is} = \hat{s}_i(t+1) - s_i(t+1) \quad (46)$$
$$= \hat{W}_{is}^T\varphi_{is}(Z_{is}(t)) - s_i(t+1)$$

The squared approximation error is defined as follows

$$E_{is} = \frac{1}{2}(e_{is})^T e_{is} \quad (47)$$

The identifier weights are updated using the gradient descent rule as

$$\hat{W}_{is}^{(l+1)T} = \hat{W}_{is}^{lT} \quad (48)$$
$$-\hat{\mu}_{is}\varphi_{is}(Z_{is}(t))(\hat{W}_{is}^{lT}\varphi_{is}(Z_{is}(t)) - s_i^l(t+1))^T$$

where $0 < \hat{\mu}_{is} < 1$ denotes the learning rate of the identifier network.
After the learning process in the NN, the identifier weight matrix will converge to a certain value $\hat{W}_{ism}^T$. Then, the identifier network output is expressed as

$$\hat{s}_i\left(t+1\right) = \hat{f}_i\left(s_i\left(t\right)\right) + \hat{g}_i\left(s_i\left(t\right)\right)u_{i(t_m^i)} \quad (49)$$
$$+\hat{h}_i(s_{it})\omega_{i(t_m^i)} = \hat{W}_{ism}^T\varphi_{is}(v_{is}^T z_{is}(t))$$

where $\hat{f}_i\left(s_i\left(t\right)\right)$, $\hat{g}_i\left(s_i\left(t\right)\right)$ and $\hat{h}_i(s_i(t))$ are the estimations of $f_i\left(s_i\left(t\right)\right)$, $g_i\left(s_i\left(t\right)\right)$ and $h_i(s_i(t))$, respectively.

7. Event-Triggered-Based Actor-Critic Learning for Unknown DT Zero-Sum Graphical Games

In this section, the algorithm (1) will be implemented using an actor-critic framework that is based on event-triggered. In order to carry out the policy evaluation (41) and estimate the optimal value function, the critic network is designed for each agent. The actor approximators are constructed to perform the policy improvements Equations (42) and (43), respectively, which estimate the bounded event-triggered optimal control and disturbance policies. Additionally, using a neural network, each agent's unknown dynamics is approximated.

7.1. Actor-Critic Approximators and Tuning

For each agent i, The control and disturbance policies are estimated using actor approximators $\hat{u}_i(.\big|\hat{W}_{ia})$ and $\hat{\omega}_i\left(.\big|\hat{W}_{id}\right)$, respectively. Also, the optimal value function is estimated using the critic approximator $\hat{V}_i\left(.\middle|\hat{W}_{ic}\right)$ so that

$$\hat{u}_{it}\left(\hat{W}_{ia}\right)=\hat{W}_{ia}^{T}\varphi_{iu}(t) \tag{50}$$

$$\hat{\omega}_{it}\left(\hat{W}_{id}\right)=\hat{W}_{id}^{T}\varphi_{id}(t) \tag{51}$$

$$\hat{V}_{it}\left(\hat{W}_{ic}\right)=\hat{W}_{ic}^{T}\varphi_{ic}(t) \tag{52}$$

where $\hat{W}_{ia}$, $\hat{W}_{id}$ are the estimated actor weights for control and disturbance, respectively and $\hat{W}_{ic}$ is the estimated critic weight. φ_{iu}, φ_{id} and φ_{ic} are the activation functions for the control actor, disturbance actor, and critic, respectively. All activation functions are assumed to be bounded, i.e., $\|\varphi_{iu}(.)\|\le\varphi_{ium}$, $\|\varphi_{id}(.)\|\le\varphi_{idm}$ and $\|\varphi_{ic}(.)\|\le\varphi_{icm}$. Each agent's activation function is equal to the local error vector of that agent and its neighbor.

The actor network's control input error for $t\in[t_m^i,t_{m+1}^i)$ can be defined as

$$e_{ia}=\hat{u}_{i(t_m^i)}\left(\hat{W}_{ia}\right)-\tilde{u}_{i(t_m^i)}=\hat{W}_{ia}^{T}\varphi_{iu}(t_m^i)-\tilde{u}_{i(t_m^i)} \tag{53}$$

The control input $\tilde{u}_{i(t_m^i)}$ is given as

$$\tilde{u}_{i(t_m^i)}=\bar{Y}\varphi\left(\left(\bar{Y}y\right)^{-1}\left(q_i+a_i\right)\hat{g}_i^{T}(s_{i(t_m^i)})\nabla\hat{V}_i\left(E_{i\left((t_m^i)+1\right)}\right)\right) \tag{54}$$

By using (52), (54) can be written as follows

$$\tilde{u}_{i(t_m^i)}=\bar{Y}\varphi\left(\left(\bar{Y}y\right)^{-1}\left(q_i+a_i\right)\hat{g}_i^{T}(s_{i(t_m^i)})G_i\hat{W}_{ic}^{T}\right) \tag{55}$$

where $G_i=[0....[I]_{ii}....0]\in\mathbb{R}^{n\times nN_{ij}}$. N_{ij} is the total number of each agent i and its neighbors.

The actor network's squared error for control input is as

$$E_{ia}=\frac{1}{2}\left(e_{ia}\right)^{T}e_{ia} \tag{56}$$

The actor weights for control input are updated using the gradient descent rule as

$$\hat{W}_{ia}^{(1+l)T}=\hat{W}_{ia}^{lT}-\mu_{ia}\left(\hat{W}_{ia}^{lT}\varphi_{iu}(t_m^i)-\tilde{u}_{i(t_m^i)}^{l}\right)(\varphi_{iu}(t_m^i))^{T} \tag{57}$$

where $0<\mu_{ia}<1$ denotes the actor network learning rate for control input.

Similarly, the actor network's disturbance error can be described as

$$e_{id}=\hat{\omega}_{i(t_m^i)}\left(\hat{W}_{id}\right)-\tilde{\omega}_{i(t_m^i)}=\hat{W}_{id}^{T}\varphi_{id}(t_m^i)-\tilde{\omega}_{i(t_m^i)} \tag{58}$$

The disturbance $\tilde{\omega}_{i(t_m^i)}$ can be defined as

$$\tilde{\omega}_{i(t_m^i)}=-\frac{1}{\gamma^2}(q_i+a_i)T_{ii}^{-1}\hat{h}_i^{T}(s_{i(t_m^i)})\nabla\hat{V}_i(E_{i(t_m^i+1)}) \tag{59}$$

By using (52), (59) can be written as follows

$$\tilde{\omega}_{i(t_m^i)}=-\frac{1}{\gamma^2}(q_i+a_i)T_{ii}^{-1}\hat{h}_i^{T}(s_{i(t_m^i)})\Delta_i\hat{W}_{ic}^{T}) \tag{60}$$

where $\Delta_i=[0....[I]_{ii}....0]\in\mathbb{R}^{k\times kN_{ij}}$.

The following is a definition of the disturbance actor's squared error

$$E_{id}=\frac{1}{2}\left(e_{id}\right)^{T}e_{id} \tag{61}$$

To update the actor weights for the disturbance, the gradient descent rule is employed as follows

$$\hat{W}_{id}^{(1+l)T}=\hat{W}_{id}^{lT}-\mu_{id}\left(\hat{W}_{id}^{lT}\varphi_{id}(t_m^i)-\tilde{\omega}_{i(t_m^i)}^{l}\right)(\varphi_{id}(t_m^i))^{T} \tag{62}$$

where $0<\mu_{id}<1$ is the learning rate for the disturbance.

The objective value function $\tilde{V}_{it}$ is given by

$$\tilde{V}_{it}=\frac{1}{2}\begin{pmatrix}E_{it}^{T}O_iE_{it}+W\left(\hat{u}_{i(t_m^i)}^{l}\right)+\sum_{j\in N_j}W\left(\hat{u}_{j(t_m^i)}^{l}\right)\\ -\gamma^2\hat{\omega}_{i(t_m^i)}^{T}T_{ii}\hat{\omega}_{i(t_m^i)}-\gamma^2\sum_{j\in N_i}\hat{\omega}_{j(t_m^i)}^{T}T_{ij}\hat{\omega}_{j(t_m^i)}\end{pmatrix}+\hat{V}_{i(t+1)}^{l} \tag{63}$$

The following is the critic network's error

$$e_{ic}=\tilde{V}_{it}-\hat{V}_{it}\left(\hat{w}_{ic}\right) \tag{64}$$

The squared error for the critic is defined as follows

$$E_{ic}=\frac{1}{2}\left(e_{ic}\right)^{T}e_{ic} \tag{65}$$

The critic weights are updated using the gradient descent rule.

$$\hat{W}_{ic}^{(1+l)T}=\hat{W}_{ic}^{lT}-\mu_{ic}\left(\hat{W}_{ic}^{lT}\varphi_{ic}(t_m^i)-\tilde{V}_{i(t_m^i)}^{l}\right)(\varphi_{ic}(t_m^i))^{T} \tag{66}$$

where $0<\mu_{ic}<1$ denotes the critic network learning rate.

Finally, Algorithm 2 is presented for actor-critic network weights online tuning of unknown DT zero-sum graphical games.

Algorithm 2
Event-triggered value iteration HDP algorithm Implementation using an actor-critic structure

1. Start the weights for the actors and identifiers at random and critic with zero.
2. Initialize the initial state $s_i(0)$ for all agents and $s_0(0)$ for the leader agent randomly.
3. setting D_i for each agent
4. Do Loop (l iterations)
   - Calculate the local error E_{it} on the system trajectory by (3).
   - Calculate the event-triggered error e_{it} by Equation (8)
   - Calculate the threshold ε_i for each agent by Equation (14)
   - Calculate the estimated state $\hat{s}^l_{i(t+1)}$ by Equation (45)
   - For each agent i, if $\|e_i(n)\| \le \varepsilon_i$ then $l \leftarrow l+1$ and go to step 4, otherwise:

- Calculate control policies $\hat{u}^l_{it}$ by Equation (50)

- Calculate disturbance policies $\hat{\omega}^l_{it}$ by Equation (51)

- Calculate the local error $\varepsilon^l_{i(t+1)}$ (3) using the estimated states

- Calculate the value function $\hat{V}^l_{i(t+1)}$ by Equation (52)

- Update the critic weights

$$\hat{W}_{ic}^{(1+l)T} = \hat{W}_{ic}^{lT} - \mu_{ic}\left(\hat{W}_{ic}^{lT}\varphi_{ic}(t_m^i) - \tilde{V}^l_{i(t_m^i)}\right)(\varphi_{ic}(t_m^i))^T$$

where $\tilde{V}_{it}$ is gained by Equation (63)

- Update the actor weights

$$\hat{W}_{ia}^{(1+l)T} = \hat{W}_{ia}^{lT} - \mu_{ia}\left(\hat{W}_{ia}^{lT}\varphi_{iu}(t_m^i) - \tilde{u}^l_{i(t_m^i)}\right)(\varphi_{iu}(t_m^i))^T$$

$$\hat{W}_{id}^{(1+l)T} = \hat{W}_{id}^{lT} - \mu_{id}\left(\hat{W}_{id}^{lT}\varphi_{id}(t_m^i) - \tilde{\omega}^l_{i(t_m^i)}\right)(\varphi_{id}(t_m^i))^T$$

- Update the identifier weights

$$\hat{W}_{is}^{(l+1)T} = \hat{W}_{is}^{lT} - \hat{\mu}_{is}\varphi_{is}(Z_{is}(t))(\hat{W}_{is}^{lT}\varphi_{is}(Z_{is}(t)) - s_i^l(t+1))^T$$

- For all i, when $\left\|\hat{V}_i^{l+1}(E_{it}) - \hat{V}_i^l(E_{it})\right\| \le \delta$ end, where δ is a small constant.

8. Simulation Study

In this part, the applicability of the developed NN-based adaptive optimal event-triggered algorithm is demonstrated by the simulation results. Consider the graph structure of Figure 1.

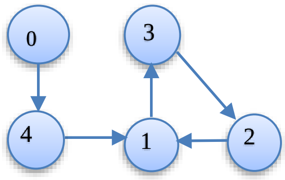

Figure 1. Topology of interactions among four agents.

The drift, input and disturbance matrices for every agent are given as

$$f_i(s_i) = \begin{bmatrix} s_{i2} + s_{i1} \\ -s_{i1} + 0.5\left(1 - s_{i1}^2\right)s_{i2} \end{bmatrix} \quad , i = 1,2,3,4$$

$$g_i(s_i) = \begin{bmatrix} 1.5 \\ -s_{i1}^2 \end{bmatrix} \quad , i = 1,2,3,4$$

$$h_1(s_1) = \begin{bmatrix} 1 \\ -0.8 \end{bmatrix}, h_2(s_2) = \begin{bmatrix} 1 \\ -0.1 \end{bmatrix}$$

$$h_3(s_3) = \begin{bmatrix} 0.9 \\ -s_{31}^2 \end{bmatrix}, h_4(s_4) = \begin{bmatrix} 1 \\ -s_{41}^2 \end{bmatrix}$$

The drift matrix of the leader is as

$$f_0(s_0) = \begin{bmatrix} s_{01} + s_{02} \\ -s_{01} + 0.5\left(1 - s_{01}^2\right)s_{02} \end{bmatrix}$$

The pinning gains are $a_1 = a_2 = a_3 = 0, a_4 = 1$, and the edge weights is considered as $c_{12} = 0.7, c_{23} = 0.5, c_{31} = 0.7, c_{14} = 0.4$. The chosen learning rates are $\hat{\mu}_{is} = 0.3$, $\hat{\mu}_{ic} = 0.1, \hat{\mu}_{ia} = 0.3$ and $\hat{\mu}_{id} = 0.3$. The disturbance attenuation is given by $\gamma = 1.5$ and bound for actuators is considered as $\overline{Y} = 1$.

The following is a list of the matrices in the performance indices

$O_{11} = O_{22} = O_{33} = O_{44} = I_{2\times 2}$

$T_{11} = T_{22} = T_{33} = T_{44} = 1$

$T_{12} = T_{14} = T_{23} = T_{31} = 1$

Figures 2, 3, 4 and 5 respectively depict the convergence of the control actor, disturbance actor, critic and identifier for agent 1. In these figures, it is shown that all weights converge. The update for all neural networks is done only at the event trigger and remains constant at other times. This reduces the calculation and execution time. Figures 6 and 7 depict the local error and the estimated states of all agents, respectively, where the local error dynamics converge to zero and the synchronization of the states of all agents with the leader state is achieved while maintaining their optimality.

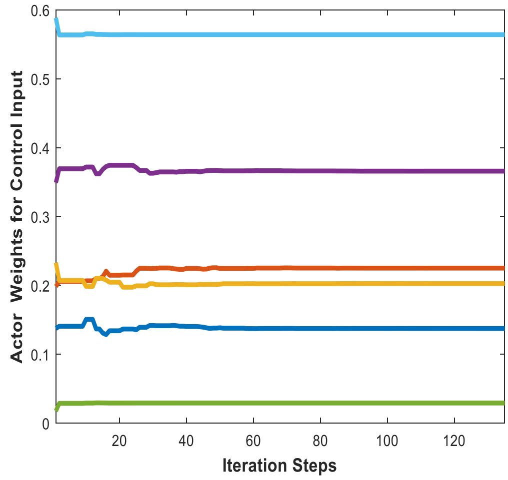

Figure 2. Control input actor weights update for agent (1).

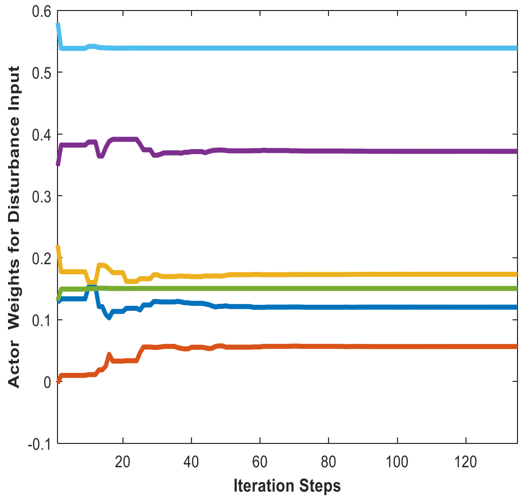

Figure 3. Disturbance actor weights update for agent (1).

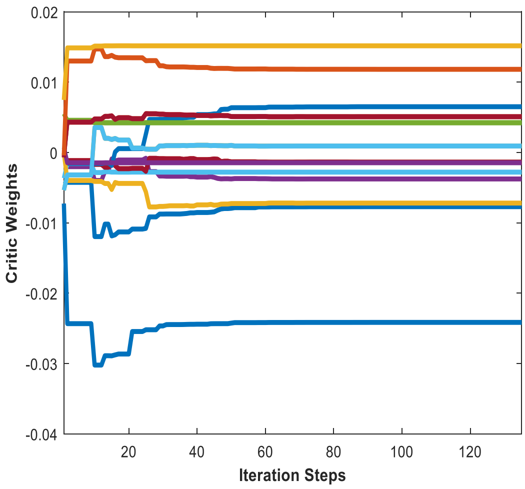

Figure 4. Critic weights update for agent (1).

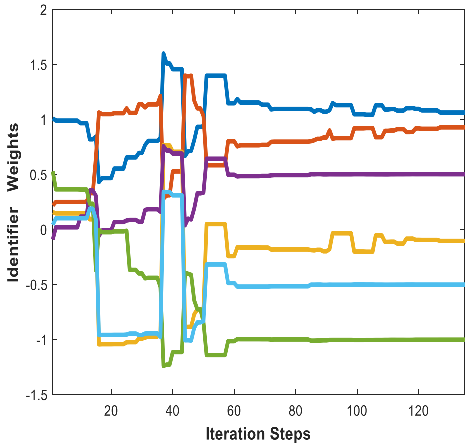

Figure 5. Identifier weights update for agent (1).

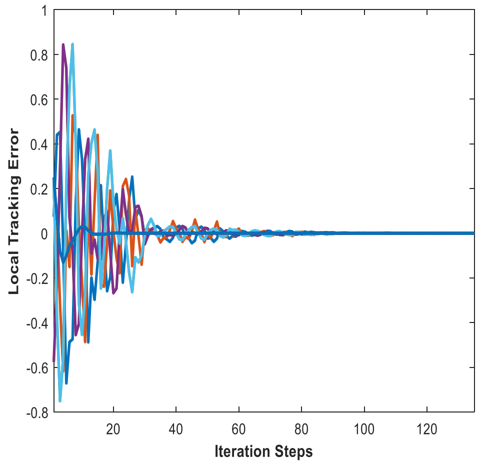

Figure 6. Tracking error versus iteration steps.

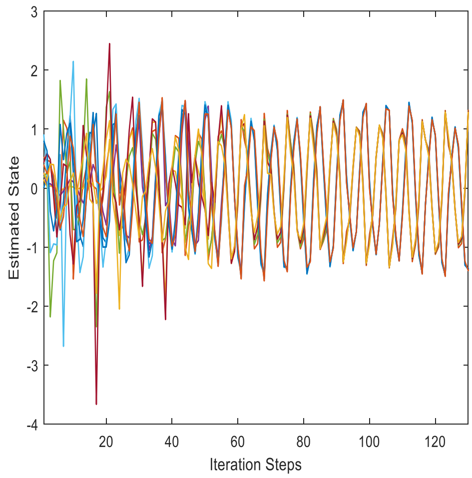

Figure 7. Synchronization of follower agents to leader agent.

9. Conclusion

An event-triggered control scheme and a distributed adaptive optimal algorithm are proposed in this paper to solve the N-player leader-follower synchronization problem for unknown nonlinear systems with disturbances and input constraints. The proposed method can reduce the excessive consumption of communication and computing resources since updates and calculations are performed only when the event happened and remain constant at other times. Besides, the actor-critic structure and NN-based system identification techniques are used to approximate the optimal event-triggered value function, optimal control, worst-case disturbance policies and unknown dynamics of the players. The closed-loop stability according to Lyapunov and the convergence of Nash game equilibrium are also shown. Finally, the proposed algorithm's efficiency in synchronizing with the leader is demonstrated.

References

[1] D. L. Fernandes, A. L. M. Leopoldino, J. de Santiago, C. Verginis, A. A. Ferreira, and J. G. de Oliveira, “Distributed control on a multi-agent environment co-simulation for DC bus voltage control,” *Electric Power Systems Research*, vol. 232, pp. 110408, 2024.

[2] K. Sun, H. Yu, and X. Xia, “Distributed control of nonlinear stochastic multi-agent systems with external disturbance and time-delay via event-triggered strategy,” *Neurocomputing*, vol. 452, pp. 275-283, 2021.

[3] M. H. Rezaei and M. B. Menhaj, “Adaptive output stationary average consensus for heterogeneous unknown linear multi-agent systems,” *IET Control Theory & Applications*, vol. 12, no. 7, pp. 847-856, 2018.

[4] W. Meng, Q. Yang, J. Sarangapani, and Y. Sun, “Distributed control of nonlinear multiagent systems with asymptotic consensus,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 5, pp. 749–757, May 2017.

[5] C. J. Li and G. P. Liu, “Data-driven consensus for non-linear networked multi-agent systems with switching topology and time-varying delays,” *IET Control Theory & Applications*, vol. 12, no. 12, pp. 1773–1779, Aug. 2018.

[6] H. Li, Q. Liu, G. Feng, and X. Zhang, “Leader–follower consensus of nonlinear time-delay multiagent systems: A time-varying gain approach,” *Automatica*, vol. 126, pp. 109444, Apr. 2021.

[7] J. Kandasamy, R. Ramachandran, V. Veerasamy, and A. X. R. Irudayaraj, “Distributed leader-follower based adaptive consensus control for networked microgrids,” *Applied Energy*, vol. 353, pp. 122083, 2024.

[8] F. Tatari, and M. B. Naghibi-Sistani, “Optimal adaptive leader-follower consensus of linear multi-agent systems: Known and unknown dynamics,” *Journal of Artificial Intelligence and Data Mining*, vol. 3, no. 1, pp. 101-111, 2015.

[9] J. Long, W. Wang, J. Huang, J. Lu, and K. Liu, “Adaptive leaderless consensus for uncertain high-order nonlinear multiagent systems with event-triggered communication,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 11, pp. 7101–7111, Nov. 2022.

[10] M. Rehan, M. Tufail, and S. Ahmed, “Leaderless consensus control of nonlinear multi-agent systems under directed topologies subject to input saturation using adaptive event-triggered mechanism,” *Journal of the Franklin Institute*, vol. 358, no. 12, pp. 6217–6239, Aug. 2021.

[11] X. Liu, J. Sun, L. Dou, and J. Chen, “Leader-following consensus for discrete-time multi-agent systems with parameter uncertainties based on the event-triggered strategy,” *Journal of Systems Science and Complexity*, vol. 30, no. 1, pp. 30–45, Feb. 2017.

[12] Q. Yuan, G. Chen, Y. Tian, Y. Yuan, Q. Zhang, X. Wang, and J. Liu, “Leader-following consensus of discrete-time nonlinear multi-agent systems with asymmetric saturation impulsive control,” *Mathematics*, vol. 12, no. 3, pp. 469, 2024.

[13] T. Basar and G. J. Olsder, “Dynamic noncooperative game theory,” *Society for Industrial and Applied Mathematics*, 2nd ed, Jan. 1998.

[14] K. G. Vamvoudakis, F. L. Lewis, and G. R Hudas, “Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality,” *Automatica*, vol. 48, no. 8, pp. 1598–1611, Aug. 2012.

[15] Y. Guo, Q. Sun, Y. Wang, and Q. Pan, “Differential graphical game-based multi-agent tracking control using integral reinforcement learning,” *IET Control Theory & Applications*, 2024.

[16] F. Tatari, M. B. Naghibi-Sistani, and K. G. Vamvoudakis, “Distributed optimal synchronization control of linear networked systems under unknown dynamics,” *American Control Conference(ACC)*, pp. 668-673, IEEE , May 2017.

[17] F. Tatari, M. B. Naghibi-Sistani, and K. G. Vamvoudakis, “Distributed learning algorithm for non-linear differential graphical games,” *Transactions of the Institute of Measurement and Control*, vol. 39, no. 2, pp. 173–182, Jul. 2017.

[18] Q. Jiao, H. Modares, S. Xu, F.L. Lewis, and K.G. Vamvoudakis, “Multi-agent zero-sum differential graphical games for disturbance rejection in distributed control,” *Automatica*, vol. 69, pp. 24-34, 2016.

[19] F. Tatari, K.G. Vamvoudakis, and M. Mazouchi, “Optimal distributed learning for disturbance rejection in networked nonlinear games under unknown dynamics,” *IET Control Theory and Applications*, vol. 13, no. 17, pp.2838-2848, 2019.

[20] D. Liu, H. Li, and D. Wang, “Neural-network-based zero-sum game for discrete-time nonlinear systems via iterative adaptive dynamic programming algorithm,” *Neurocomputing*, vol. 110, pp. 92–100, 2013.

[21] W. Wang, X. Chen, and J. Du, “Model-free finite-horizon optimal control of discrete-time two-player zero-sum games,” *International Journal of Systems Science*, vol. 54, no. 1, pp. 167-179, 2023.

[22] R. Song and L. Zhu, “Stable value iteration for two-player zero-sum game of discrete-time nonlinear systems based on adaptive dynamic programming,” *Neurocomputing*, vol. 340, pp. 180–195, May 2019.

[23] M. Abouheaf, F. L. Lewis, K. G. Vamvoudakis, Sofie Haesaert, and R. Babuska, “Multi-agent discrete-time graphical games and reinforcement learning solutions,” *Automatica*, vol. 50, no. 12, pp. 3038–3053, 2014.

[24] M. I. Abouheaf, F. L. Lewis, M. S. Mahmoud, and D. G. Mikulski, “Discrete-time dynamic graphical games: model-free reinforcement learning solution,” *Control Theory and Technology*, vol. 13, no. 1, pp. 55–69, Feb. 2015.

[25] R. S. Sutton, and A. G. Barto, “Reinforcement Learning: An Introduction,” *Robotica*, vol. 17, no. 2, pp. 229-235, 1999.

[26] M. Taghian, A. Asadi, and R. Safabakhsh, “A reinforcement learning-based encoder-decoder framework for learning stock trading rules,” *Journal of Artificial Intelligence and Data Mining*, vol. 11, no. 1, pp. 103-118, 2023.

[27] B. Kiumarsi, H. Modares, and F. Lewis, “Reinforcement learning for distributed control and multi-player games,” *In Handbook of Reinforcement Learning and Control, Springer*, Cham: Springer International Publishing, pp. 7-27, 2021.

[28] H. Jiang, H. Zhang, G. Xiao, and X. Cui, “Data-based approximate optimal control for nonzero-sum games of multi-player systems using adaptive dynamic programming” *Neurocomputing*, vol. 275, pp. 192-199, 2018.

[29] H. Jiang, H. Zhang, Y. Luo, and X. Cui, “H-infinity control with constrained input for completely unknown nonlinear systems using data-driven reinforcement learning method,” *Neurocomputing*, vol. 237, pp. 226-234, 2017.

[30] P. Liu, H. Zhang, C. Liu, and H. Su, “Online dual-network-based adaptive dynamic programming for solving partially unknown multi-player non-zero-sum games with control constraints,” *IEEE Access*, vol. 8, pp. 182295–182306, 2020.

[31] B. Luo, D. Liu, and H. N. Wu, “Adaptive constrained optimal control design for data-based nonlinear discrete-time systems with critic-only structure,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2099–2111, 2017.

[32] X. i, Y. Tang, and H. R. Karimi, “Consensus of multi-agent systems via fully distributed event-triggered control,” *Automatica*, vol. 116, pp. 108898, 2020.

[33] Y. Y. Qian, L. Liu, and G. Feng, “Output consensus of heterogeneous linear multi-agent systems with adaptive event-triggered control,” *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2606–2613, Jun. 2018.

[34] X. Li, Z. Sun, Y. Tang, and H. R. Karimi, “Adaptive event-triggered consensus of multiagent systems on directed graphs,” *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1670–1685, 2020.

[35] S. Hu, D. Yue, X. Yin, X. Xie, and Y. Ma, “Adaptive event-triggered control for nonlinear discrete-time systems,” *International Journal of Robust and Nonlinear Control*, vol. 26, no. 18, pp. 4104–4125, 2016.

[36] L. Dong, X. Zhong, C. Sun, and H. He, “Adaptive event-triggered control based on heuristic dynamic programming for nonlinear discrete-time systems,” *IEEE transactions on neural networks and learning systems*, vol. 28, no. 7, pp. 1594–1605, Jul. 2017.

[37] Z. Wang, Q. Wei, D. Liu, and Y. Luo “Event-triggered adaptive control for discrete-time zero-sum games,” *In 2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-7. IEEE, 2019.

[38] S. Khoo, L. Xie, and Z. Man, “Robust finite-time consensus tracking algorithm for multirobot systems,” *IEEE/ASME transactions on mechatronics*, vol. 14, no. 2, pp. 219–228, 2009.

[39] M. Abu-Khalaf and F. L. Lewis, “Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach,” *Automatica*, vol. 41, no. 5, pp. 779–791, May 2005.

کنترل اجماع رهبر-پیرو بهینه تطبیقی مبتنی بر رویداد برای سیستم‌های غیرخطی زمان گسسته ورودی محدود ناشناخته

زهرا جهان، عباس دیدبان* و فرزانه تاتاری

دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران.

ارسال ۲۰۲۴/۱۱/۱۱؛ بازنگری ۲۰۲۴/۰۴/۰۸؛ پذیرش ۲۰۲۴/۰۶/۰۵

چکیده:

در این مقاله، یک الگوریتم توزیع‌شده بهینه تطبیقی برمبنای کنترل مبتنی بر رویداد برای حل بازی‌های گرافی مجموع صفر زمان گسسته چندعاملی با سیستم‌های ورودی محدود غیرخطی ناشناخته و همراه با اغتشاشات خارجی معرفی می‌شود. الگوریتم پیشنهادی معادلات همیلتون-جاکوبی-ایزاکس همراه با رویداد را با فرض دینامیک ناشناخته و بر اساس برنامه‌ریزی پویای تطبیقی تکرار ارزش حل می‌کند تا کنترل کننده‌های بهینه توزیع شده را برای اجماع رهبر-پیرو عواملی که روی یک گراف ارتباطی تعامل دارند، توسعه دهد. الگوریتم با استفاده از شبکه‌های عصبی نقاد-عملگر پیاده سازی می شود و دینامیک‌های ناشناخته سیستم با شبکه عصبی شناساگر تقریب زده می‌شوند. همچنین، ورودی کنترل، اغتشاش خارجی و وزن‌های شبکه عصبی به‌صورت دوره‌ای و فقط در لحظه‌های رویداد به‌روزرسانی می‌شوند تا فرآیند محاسباتی را ساده‌تر کنند. معرفی بازی‌های گرافی غیرخطی مجموع صفر زمان گسسته با دینامیک ناشناخته، محدودیت ورودی کنترل و اغتشاشات خارجی و حل آنها با روش‌های مبتنی بر رویداد، این مقاله را از آثار منتشر شده قبلی متمایز می‌کند. همچنین پایداری سیستم حلقه بسته و همگرایی به تعادل نش اثبات شده است. در انتها، نتایج شبیه سازی برای توصیف کارایی روش پیشنهادی ارائه شده‌است.

کلمات کلیدی: کنترل بهینه توزیع شده، روش مبتنی بر رویداد، اجماع رهبر-پیرو بهینه، یادگیری تقویتی، شبکه‌های عصبی.