Learning Consensus in Adversarial Environments

Kyriakos G. Vamvoudakis, Luis R. García Carrillo, João P. Hespanha

Center for Control, Dynamical-systems and Computation (CCDC), University of California, Santa Barbara, CA 93106-9560 USA

ABSTRACT

This work presents a game theory-based consensus problem for leaderless multi-agent systems in the presence of adversarial inputs that are introducing disturbance to the dynamics. Given the presence of enemy components and the possibility of malicious cyber attacks compromising the security of networked teams, a position agreement must be reached by the networked mobile team based on environmental changes. The problem is addressed under a distributed decision making framework that is robust to possible cyber attacks, which has an advantage over centralized decision making in the sense that a decision maker is not required to access information from all the other decision makers. The proposed framework derives three tuning laws for every agent; one associated with the cost, one associated with the controller, and one with the adversarial input.

Keywords: Security, game theory, consensus, approximation, Hamilton-Jacobi equations, optimization

1. INTRODUCTION

Due to the highly uncertain and dynamic nature of military conflict, enabling autonomous agents to gracefully adapt to mission and environmental changes is a very challenging task. These capabilities are necessary in the asymmetric battles waged against insurgencies, where enemy combatants quickly adapt to Army strategies and tactics. The ability to synchronize activities provides an important strategic capability in a wide variety of military missions. The United States Army, Air Force, and Navy have recently shown interest in the cyber security aspect of UAVs and UGVs, which rely heavily on their on-board autopilots and controllers to function. Most of the currently available autopilot systems were built without cyber security considerations, and are thus vulnerable to cyber attacks. This paper provides knowledge to the problem of commanding multiple tactical assets in military and adversarial environments, attending Army's expectations that networked teams will perform in a reliable manner especially when being attacked by advanced and persistent threats.

Machine learning ideas are being used as an essential component to address problems in multi-agent systems with diverse and selfish interests, traditional algorithmic and distributed systems need to be combined with the understanding of game-theoretic and economic issues.[1] A lot of applications require cooperation of separate agents to achieve global objectives and learning is an ideal approach in the cases where classical optimization techniques are infeasible.[2] Meanwhile, multi-agent learning techniques offer important theoretical challenges and especially in construing how each agent learns and adapts when other agents are learning and adapting simultaneously. This introduces game-theoretic issues to the learning process.

A survey of existing cyber-threats in multi-agent systems and models of realistic and rational adversary models are presented in Cardenas et al.[3,4] Consensus in the presence of persistent adversaries has been focused on detecting, identifying and isolating the failing nodes[5] which are computationally expensive and most of the time they use global information and specific graph connectivity. The adversaries can easily drive the system unstable and make the system operate with an undesired behavior. Thus it is better to fight adversaries than guard against them.

Further author information: (Send correspondence to Kyriakos G. Vamvoudakis)
K.G. Vamvoudakis: E-mail: kyriakos@ece.ucsb.edu, Telephone: 1 805 893 7785
L.R. García Carrillo: E-mail: lrgc@ece.ucsb.edu
J.P. Hespanha: E-mail: hespanha@ece.ucsb.edu

Unmanned Systems Technology XV, edited by Robert E. Karlsen, Douglas W. Gage, Charles M. Shoemaker, Grant R. Gerhart, Proc. of SPIE Vol. 8741, 87410K
 · doi: 10.1117/12.2014372

This paper proposes a game-theory based consensus learning algorithm for networked systems that are being attacked by persistent adversaries. As opposed to existing consensus algorithms, our method is not limited to specific graph connectivity and also guarantees robustness of the networked team which can be described by a *minimax* optimization problem for every agent. The dynamics and the performance of every agent depend only on measurable local information.

2. PROBLEM FORMULATION

Consider a multi-agent system consisting of N mobile systems in a set $\Omega \subset \mathbb{R}^m$. Also denote the set $\mathcal{N} = \{1, ..., N\}$. Suppose that the dynamics of every agent are perturbed by an adversary. Thus, the dynamics of each agent are given $\forall t \geq 0$ as,

$$\dot{p}_i = u_i + v_i, \ i \in \mathcal{N} \tag{1}$$

where $p_i \in \Omega \subset \mathbb{R}^m$, is the position, and $u_i \in \mathcal{U}_{iu}$, $v_i \in \mathcal{U}_{iv}$ is the control and adversarial input of agent i respectively. We now assume that the sets $\mathcal{U}_{iu} \subset \mathbb{R}^m$ and $\mathcal{U}_{iv} \subset \mathbb{R}^m$ of admissible control and adversarial policies are compact $\forall i \in \mathcal{N}$.

We model the multi-agent system by considering a fixed topology graph $\mathcal{G} = (\mathcal{V}_\mathcal{G}, \mathcal{E}_\mathcal{G})$ consisting of nonempty finite set of N nodes $\mathcal{V}_\mathcal{G} = \{n_1, ..., n_N\}$ that represent agents and a set of edges or arcs $\mathcal{E}_\mathcal{G} \subseteq \mathcal{V}_\mathcal{G} \times \mathcal{V}_\mathcal{G}$ that represent the inter-agent information exchange links. Denote the connectivity matrix $\mathcal{A}_\mathcal{G} = [a_{ij}]$ of $\mathcal{G}$ as

$$a_{ij} \begin{cases} > 0, \text{ if } (n_j, n_i) \in \mathcal{E}_\mathcal{G} \\ = 0, \text{ otherwise.} \end{cases}$$

We assume the graph is simple, e.g. no repeated edges and $(n_i, n_i) \notin \mathcal{E}_\mathcal{G}, \forall i \in \mathcal{N}$ no self-loops ($a_{ii} = 0$). The set of neighbors of a node n_i is $\mathcal{N}_i = \{n_j : (n_j, n_i) \in \ \mathcal{E}_\mathcal{G}\}$, i.e. the set of nodes with arcs incoming to n_i. Define the *degree* matrix as $\mathcal{D} = diag(d_i)$ with $d_i = \sum_{j \in \mathcal{N}_i} a_{ij}$ the weighted degree of node i (i.e. ith row sum of $\mathcal{A}_\mathcal{G}$). The number of neighbors of agent i is $|\mathcal{N}_i|$ which is equal to d_i. Define the graph Laplacian matrix as $\mathcal{L}_\mathcal{G} = \mathcal{D} - \mathcal{A}_\mathcal{G}$, which has all row sums equal to zero.

The local signals describing the exchange of relative position are given by $s_i \in \mathbb{R}^m$ and are $\forall t \geq 0$ defined as,

$$s_i(t) = \sum_{j \in \mathcal{N}_i} a_{ij}\big(p_i - p_j\big), \ i \in \mathcal{N} \tag{2}$$

where the signals $s_i(t)$ for any agent $i \in \mathcal{N}$ in other words describe the sum of the external output measurements relative to the other vehicles that agent i can measure.

The dynamics of (2) for any agent $i \in \mathcal{N}$ and $\forall t \geq 0$ can be written as,

$$\begin{aligned} \dot{s}_i &= \sum_{j \in \mathcal{N}_i} a_{ij}\big(\dot{p}_i - \dot{p}_j\big) \\ \dot{s}_i &= d_i(u_i + v_i) - \sum_{j \in \mathcal{N}_i} a_{ij}(u_j + v_j) \end{aligned} \tag{3}$$

REMARK 2.1. *The linear dynamical system* (3) *can represent an aircraft rolling dynamics model driven by the control and adversarial inputs of its neighborhood aircrafts.*

We want to achieve consensus while simultaneously optimizing some performance specifications on the agents. To capture this, we will use ideas drawn from game theory because it is more practical to fight adversaries than to guard against them.[7,8]

Write the local performance indices for any agent $i \in \mathcal{N}$ as,

$$J_i(s_i(0), u_i, u_{Ni}, v_i, v_{Ni}) = \frac{1}{2}\int_0^\infty \left(\|s_i\|^2 + \|u_i\|^2 + \sum_{j \in \mathcal{N}_i} \|u_j\|^2 - \gamma_{ii}^2 \|v_i\|^2 - \sum_{j \in \mathcal{N}_i} \gamma_{ij}^2 \|v_j\|^2 \right) dt \tag{4}$$

where $\|\chi\|^2$ given an m-vector denotes the squared Euclidean norm of χ i.e. $\|\chi\|^2 = \chi^T\chi$ (these notations are used interchangeably in the manuscript), also $\gamma_{ii} \geq \gamma_{ii}^* > 0$, $\gamma_{ij} > 0$ where γ_{ii}^* is the smallest γ_{ii} for which the system is stabilized.[9]

2.1 Definition of the Game

In this subsection the game is defined for every agent i and takes into account only distributed information. The following definition is needed.

DEFINITION 2.2. *A pair u_i^* and v_i^*, $\forall i \in \mathcal{N}$ is said to constitute a saddle point $\forall i \in \mathcal{N}$ with u_i^* the minimizer and v_i^* the maximizer and cost J_i for agent i and for policies u_{Ni} and v_{Ni} in the neighborhood, when the following inequality is true,*

$$J_i(u_i^*, u_{Ni}, v_i, v_{Ni}) \leq J_i(u_i^*, u_{Ni}, v_i^*, v_{Ni}) \leq J_i(u_i, u_{Ni}, v_i^*, v_{Ni}),$$

which by assigning $u_{Ni} = u_{Ni}^*$, $v_{Ni} = v_{Ni}^*$ becomes

$$J_i(u_i^*, v_i, u_{Ni}^*, v_{Ni}^*) \leq J_i(u_i^*, v_i^*, u_{Ni}^*, v_{Ni}^*) \leq J_i(u_i, v_i^*, u_{Ni}^*, v_{Ni}^*). \tag{5}$$

It is important to interpret the control inputs u_i, and adversarial inputs v_i, $\forall i \in \mathcal{N}$ as state dependent strategies. Then the value function for every agent $i \in \mathcal{N}$ can be defined as,

$$V_i = \frac{1}{2}\int_t^\infty \left(\|s_i\|^2 + \|u_i\|^2 + \sum_{j\in\mathcal{N}_i} \|u_j\|^2 - \gamma_{ii}^2 \|v_i\|^2 - \sum_{j\in\mathcal{N}_i} \gamma_{ij}^2 \|v_j\|^2 \right) d\tau, \ \forall t \geq 0. \tag{6}$$

Every agent is driven by the control and adversarial inputs of herself and her neighborhood. In the functionals (6) we need to find the best case control policy for the worst case adversarial policy and because of the dependence of every agent $i \in \mathcal{N}$ to her neighborhood, this is called a graphical game.[11] This game is based on the topology of the graph $\mathcal{G} = (\mathcal{V}_\mathcal{G}, \mathcal{E}_\mathcal{G})$.

Then the control/adversarial objective (6) of agent $i \in \mathcal{N}$ in the game is to determine,

$$V_i^* = \min_{u_i} \max_{v_i} \frac{1}{2}\int_t^\infty \left(\|s_i\|^2 + \|u_i\|^2 + \sum_{j\in\mathcal{N}_i} \|u_j\|^2 - \gamma_{ii}^2 \|v_i\|^2 - \sum_{j\in\mathcal{N}_i} \gamma_{ij}^2 \|v_j\|^2 \right) d\tau, \ \forall t \geq 0 \tag{7}$$

subject to dynamical constraints (3). Thus, the control input u_i is the minimizing player and the adversarial input v_i is the maximizing one for every agent.

When V_i, $\forall i \in \mathcal{N}$ is finite, a differential equivalent to (6) is given in terms of the Hamiltonians by using $\rho_i = \frac{\partial V_i}{\partial s_i}$,

$$\begin{aligned} H_i(s_i, \frac{\partial V_i}{\partial s_i}, u_i, u_{Ni}, v_i, v_{Ni}) &\equiv \frac{\partial V_i}{\partial s_i}^T \left(d_i(u_i + v_i) - \sum_{j\in\mathcal{N}_i} a_{ij}(u_j + v_j) \right) \\ &+ \frac{1}{2}\left(\|s_i\|^2 + \|u_i\|^2 + \sum_{j\in\mathcal{N}_i} \|u_j\|^2 - \gamma_{ii}^2 \|v_i\|^2 - \sum_{j\in\mathcal{N}_i} \gamma_{ij}^2 \|v_j\|^2 \right) = 0 \end{aligned} \tag{8}$$

with boundary condition $V_i(0) = 0$.

Now the control and attacker policies for every agent $i \in \mathcal{N}$ are given by,

$$u_i^* = \arg\min_{u_i} H_i(s_i, \frac{\partial V_i}{\partial s_i}, u_i, u_{Ni}, v_i, v_{Ni}) = -d_i \frac{\partial V_i}{\partial s_i} \tag{9}$$

$$v_i^* = \arg\max_{v_i} H_i(s_i, \frac{\partial V_i}{\partial s_i}, u_i, u_{Ni}, v_i, v_{Ni}) = \frac{d_i}{\gamma_{ii}^2} \frac{\partial V_i}{\partial s_i}. \tag{10}$$

2.2 Optimality Equations

By substituting (9) and (10) in (8) then we have the following coupled Hamilton-Jacobi equations $\forall i \in \mathcal{N}$,

$$\frac{\partial V_i}{\partial s_i}^T \left(d_i^2 (\frac{1}{\gamma_{ii}^2}\frac{\partial V_i}{\partial s_i} - \frac{\partial V_i}{\partial s_i}) - \sum_{j\in\mathcal{N}_i} a_{ij}(\frac{1}{\gamma_{jj}^2}\frac{\partial V_j}{\partial s_j} - \frac{\partial V_j}{\partial s_j}) \right)$$
$$+ \frac{1}{2}\left(\|s_i\|^2 + d_i^2 \frac{\partial V_i}{\partial s_i}^T \frac{\partial V_i}{\partial s_i} + \sum_{j\in\mathcal{N}_i} d_j^2 \frac{\partial V_j}{\partial s_j}^T \frac{\partial V_j}{\partial s_j} - \gamma_{ii}^2 \left\| \frac{d_i}{\gamma_{ii}^2}\frac{\partial V_i}{\partial s_i} \right\|^2 - \sum_{j\in\mathcal{N}_i} \gamma_{ij}^2 \left\| \frac{d_j}{\gamma_{jj}^2}\frac{\partial V_j}{\partial s_j} \right\|^2 \right) = 0. \quad (11)$$

The minimum nonnegative definite solution $V_i^*, \forall i \in \mathcal{N}$ of (11) gives the saddle point value for every agent i. Thus the gains γ_{ii} and γ_{ij} should be chosen large enough. Define $u_i^* = u_i(V_i^*)$, $v_i^* = v_i(V_i^*)$ as (9), and (10) respectively given in terms of V_i^*.

It is straightforward to see that we will have N coupled partial differential equations and after solving all of them simultaneously we will converge to the solution of the game. By using classic ideas from control, those equations are impossible to solve. In the next section we will show how to solve those equations in a distributed way by requiring only measurements from the neighborhood and using learning ideas.

3. REACHING CONSENSUS WITH LEARNING IDEAS

In this section a new online learning algorithm is proposed. The agents achieve approximate consensus in the presence of adversarial inputs by simultaneously optimizing some distributed performances.

To solve the Bellman equations given in (8), approximation is required for all the value functions V_i and their gradients (approximation in Sobolev space,[12]). Every cost depends on the agent's state and costate and also the costate of the neighbors because of the control and adversarial inputs. Then we have to use state information from the neighborhood in the following approximation.

We will use an actor-critic framework to approach the problem. Namely we will use critic neural networks to approximate the costs, and action neural networks (actors) to approximate the control and the adversarial inputs respectively. The total number of approximators will be $3N$ for the networked team.

The optimal value function (cost), the optimal control and adversarial input $\forall i \in \mathcal{N}$ can be written by using linearly parametrized approximators (NNs) as,

$$V_i^*(\bar{s}_i) = W_i^T \phi_i(\bar{s}_i) + \epsilon_{ci}(\bar{s}_i)$$

$$u_i^*(\bar{s}_i) = -d_i(\frac{\partial \phi_i}{\partial \bar{s}_i}^T W_i + \frac{\partial \epsilon_{ci}}{\partial \bar{s}_i}) \quad (12)$$

$$v_i^*(\bar{s}_i) = \frac{d_i}{\gamma_{ii}^2}(\frac{\partial \phi_i}{\partial \bar{s}_i}^T W_i + \frac{\partial \epsilon_{ci}}{\partial \bar{s}_i}) \quad (13)$$

where $\bar{s}_i \in \mathbb{R}^{mN}$ is a vector that has zeros in the entries of the agents that are not in the neighborhood and not zeros in positions $i \in \mathcal{N}$ and $\forall j \in \mathcal{N}_i$, also $W_i \in \mathbb{R}^h$ are unknown ideal NN weights, h is the number of neurons, $\phi_i(\bar{s}_i) = [\phi_{i1}(\bar{s}_i)\ \phi_{i2}(\bar{s}_i)\ ...\ \phi_{ih}(\bar{s}_i)] : \mathbb{R}^{mN} \to \mathbb{R}^h$ are smooth basis functions, and $\epsilon_{ci} \in \mathbb{R}$ is the residual error. The basis sets are selected so that as $h \to \infty$, the functions $\phi_i(\bar{s}_i)$ provide a complete independent basis for $V_i^*(\bar{s}_i)$.

3.1 Actor-Critic Framework

Assuming current weight estimates $\hat{W}c_i$, the outputs of the critic NN $\forall i \in \mathcal{N}$ are given by

$$\hat{V}_i = \hat{W}c_i^T \phi_i(\bar{s}_i)$$

where vectors $\phi_i(\bar{s}_i) \in \mathbb{R}^h$ are the basis function vectors and h the number of neurons in the hidden layer.

By using the Stone-Weierstrass higher-order approximation Theorem both $V_i^*(\bar{s}_i)$ and $\frac{\partial V_i^*(\bar{s}_i)}{\partial \bar{s}_i}$ can be approximated by linear in the parameters neural networks and as $h \to \infty$ the residual errors $\epsilon_{ci}(\bar{s}_i) \to 0$, $\frac{\partial \epsilon_{ci}}{\partial \bar{s}_i} \to 0$. We will find the control and attack policy in the form of two action neural networks (actors) which compute (9) and (10) respectively in the following forms,

$$\hat{u}_i = -d_i \frac{\partial \phi_i}{\partial \bar{s}_i}^T \hat{W}u_i \tag{14}$$

and

$$\hat{v}_i = \frac{d_i}{\gamma_{ii}^2} \frac{\partial \phi_i}{\partial \bar{s}_i}^T \hat{W}v_i. \tag{15}$$

Then the Bellman equations (8), $\forall i \in \mathcal{N}$ become,

$$\begin{aligned} H_i(s_i, \hat{W}c_i, \hat{u}_i, \hat{u}_{Ni}, \hat{v}_i, \hat{v}_{Ni}) &= \hat{W}c_i^T \frac{\partial \phi_i}{\partial \bar{s}_i} \left((\hat{u}_i + \hat{v}_i) - \sum_{j \in \mathcal{N}_i} a_{ij}(\hat{u}_j + \hat{v}_j) \right) \\ &+ \frac{1}{2} \left(\|s_i\|^2 + \|\hat{u}_i\|^2 + \sum_{j \in \mathcal{N}_i} \|\hat{u}_j\|^2 - \gamma_{ii}^2 \|\hat{v}_i\|^2 - \sum_{j \in \mathcal{N}_i} \gamma_{ij}^2 \|\hat{v}_j\|^2 \right) \\ &\equiv \hat{W}c_i^T \frac{\partial \phi_i}{\partial \bar{s}_i} \left(d_i(\hat{u}_i + \hat{v}_i) - \sum_{j \in \mathcal{N}_i} a_{ij}(\hat{u}_j + \hat{v}_j) \right) + r_i \equiv e_i. \end{aligned} \tag{16}$$

It is desired to select the critic weights $\hat{W}c_i$ to minimize the following squared error $K_i \in \mathbb{R}^+$,

$$K_i = \frac{1}{2} \int_0^t \|e_i\|^2 \, d\tau. \tag{17}$$

Then $\hat{W}c_i \to W_i$, where W_i are the optimal weights of the approximator. The next subsection will provide the tuning laws for all the approximators just defined.

3.2 Learning Algorithm

Define the critic and actor estimation errors $\forall i \in \mathcal{N}$ as,

$$\tilde{W}c_i = W_i - \hat{W}c_i; \ \tilde{W}u_i = W_i - \hat{W}u_i; \ \tilde{W}v_i = W_i - \hat{W}v_i.$$

The least squares update law for the critic can be found from the integral-squared error (17) $\forall i \in \mathcal{N}$ as,

$$\frac{\partial K_i}{\partial \hat{W}c_i} = \int_0^t e_i(\tau) \frac{\partial e_i(\tau)}{\partial \hat{W}c_i(t)} d\tau = 0.$$

The batch least squares critic estimate can be found $\forall i \in \mathcal{N}$ as,

$$\hat{W}c_i(t) = -\left(\int_0^t \omega_i(\tau)\omega_i(\tau)^T d\tau \right)^{-1} \int_0^t \omega_i(\tau) r_i(\tau) d\tau \tag{18}$$

where $\omega_i = \frac{\partial \phi_i}{\partial \bar{s}_i}\Big(d_i(\hat{u}_i + \hat{v}_i) - \sum_{j\in\mathcal{N}_i} a_{ij}(\hat{u}_j + \hat{v}_j)\Big)$ and provided the inverse exists.

By taking the time derivative of (18) and normalizing we have,

$$\dot{\hat{W}}c_i = -\alpha_i \frac{\Delta_i \omega_i}{(\omega_i^T \Delta_i \omega_i + 1)^2} e_i^T \tag{19}$$

where $\alpha_i \in \mathbb{R}^+$ determines the speed of convergence, and $\Delta_i := \Big(\int_{t_0}^{t} \omega_i \omega_i^T d\tau\Big)^{-1}$ can be found from,

$$\dot{\Delta}_i = -\alpha_i \frac{\Delta_i \omega_i \omega_i^T \Delta_i}{(\omega_i^T \Delta_i \omega_i + 1)^2},\ \Delta_i(t_r^+) = \Delta_i(t_0) = \psi_{1i} I$$

where t_r^+ is the resetting time when $\underline{\lambda}(\Delta_i) \leq \psi_{0i}$, design scalars $\psi_{1i} > \psi_{0i} \in \mathbb{R}^+$ and because $\dot{\Delta}_i \leq 0$ one has,

$$\psi_{0i} I \leq \Delta_i \leq \psi_{1i} I \tag{20}$$

where I is an identity matrix of appropriate dimensions. The resetting ensures that Δ_i is always positive definite and slow adaptation due to lack of persistence of excitation in some directions is avoided. This technique is referred in the literature as covariance resetting modification.[13]

The tuning law for the control input $\forall i \in \mathcal{N}$ is,

$$\dot{\hat{W}}u_i = \mathcal{P}r\Bigg[\alpha_{ui}\Big\{d_i^2 \frac{\partial \phi_i}{\partial \bar{s}_i}\frac{\partial \phi_i}{\partial \bar{s}_i}^T \hat{W}u_i \frac{\bar{\omega}_i^T}{(\omega_i^T \Delta_i \omega_i + 1)}\hat{W}c_i + \sum_{j\in\mathcal{N}_i} d_j^2 \frac{\partial \phi_j}{\partial \bar{s}_j}\frac{\partial \phi_j}{\partial \bar{s}_j}^T \hat{W}u_j \frac{\bar{\omega}_i^T}{(\omega_i^T \Delta_i \omega_i + 1)}\hat{W}c_j$$
$$+ \sigma_{ui}(\hat{W}c_i - \hat{W}u_i)\Big\}\Bigg] \tag{21}$$

and the tuning law for the adversary $\forall i \in \mathcal{N}$ is,

$$\dot{\hat{W}}v_i = \mathcal{P}r\Bigg[\alpha_{vi}\Big\{ -\frac{d_i^2}{\gamma_{ii}^2}\frac{\partial \phi_i}{\partial \bar{s}_i}\frac{\partial \phi_i}{\partial \bar{s}_i}^T \hat{W}v_i \frac{\bar{\omega}_i^T}{(\omega_i^T \Delta_i \omega_i + 1)}\hat{W}c_i - \sum_{j\in\mathcal{N}_i} \frac{d_j^2 \gamma_{ij}}{\gamma_{jj}^4}\frac{\partial \phi_j}{\partial \bar{s}_j}\frac{\partial \phi_j}{\partial \bar{s}_j}^T \hat{W}v_j \frac{\bar{\omega}_i^T}{(\omega_i^T \Delta_i \omega_i + 1)}\hat{W}c_j$$
$$+ \sigma_{vi}(\hat{W}c_i - \hat{W}v_i)\Big\}\Bigg] \tag{22}$$

where $\mathcal{P}r[.]$ is the projection operator[13–15] that will make sure the tuning laws hold all their properties in the absence of projection and moreover guarantees uniform boundedness of $\tilde{W}v_i$ and $\tilde{W}u_i$, $\alpha_{ui}, \alpha_{vi} \in \mathbb{R}^+$ determine the speed of convergence, $\sigma_{ui}, \sigma_{vi} \in \mathbb{R}^+$ are adaptation gains and $\bar{\omega}_i := \frac{\omega_i^T}{\omega_i^T \Delta_i \omega_i + 1}$.

The main Theorem is presented next and provides a Lyapunov stability proof for the proposed game-theory based learning framework.

THEOREM 3.1. *Consider the dynamics given by* (3)*, the control input given by* (14)*, the adversarial input given by* (15) *and the signal* $\bar{\omega}_i$ *be persistently exciting and also* $d_i \neq 0$*. The critic tuning law is given by* (19)*, the control input tuning law is given by* (21) *and the adversarial input tuning law is given by* (22)*. Then, the solution* $\big(s_i, \tilde{W}c_i, \tilde{W}u_i, \tilde{W}v_i\big)$ *is Uniformly Ultimately Bounded (UUB)* $\forall i \in \mathcal{N}$ *and for all* $\big(s_i(0), \tilde{W}c_i(0), \tilde{W}u_i(0), \tilde{W}v_i(0)\big)$ *and* $t \geq 0$ *and thus* $p_i \approx p_j$*,* $\forall i, j \in \mathcal{N}$ *which means that approximate consensus has been achieved provided that the number of basis sets (neurons in the hidden layer) is large enough to guarantee that the residual errors* $\epsilon_{ci} \to 0$*,* $\forall i \in \mathcal{N}$*.*

Proof. Due to space constraints the proof will be provided in a future paper. □

4. SIMULATION RESULTS

In this section we illustrate the theoretical developments for a graph topology of 5 agents networked as shown in Figure 1. The gains in the tuning laws are selected as $\sigma_{ui} = 5, \sigma_{vi} = 5, \alpha_i = 10, \alpha_{ui} = 1, \alpha_{vi} = 1$, $\forall i \in \mathcal{N}$, the covariance matrix is initialized as $\Delta(0) = 4000I$, where I is an identity matrix of appropriate dimensions and all the weights of the tuning laws are initialized randomly inside $[-1, 1]$. To ensure PE, a probing signal $p(t) = \tanh(t)\big(sin(2t)cos(t) + 4sin(t^2) + 10cos(5t)sin(11t)\big)$ is added to the input signals for the first 2 seconds of the simulation. The 5 agents are moving in a bi-dimensional space ($m = 2$) and the initial conditions of position are selected randomly, and all the edge weights are selected equal to 1. By using the framework of Theorem 3.1 and $\gamma_{ii} = 4$, $\gamma_{ij} = 6$, $\forall i \neq j \in \mathcal{N}$ the agents for the given graph topology have reached consensus as shown in Figure 2.

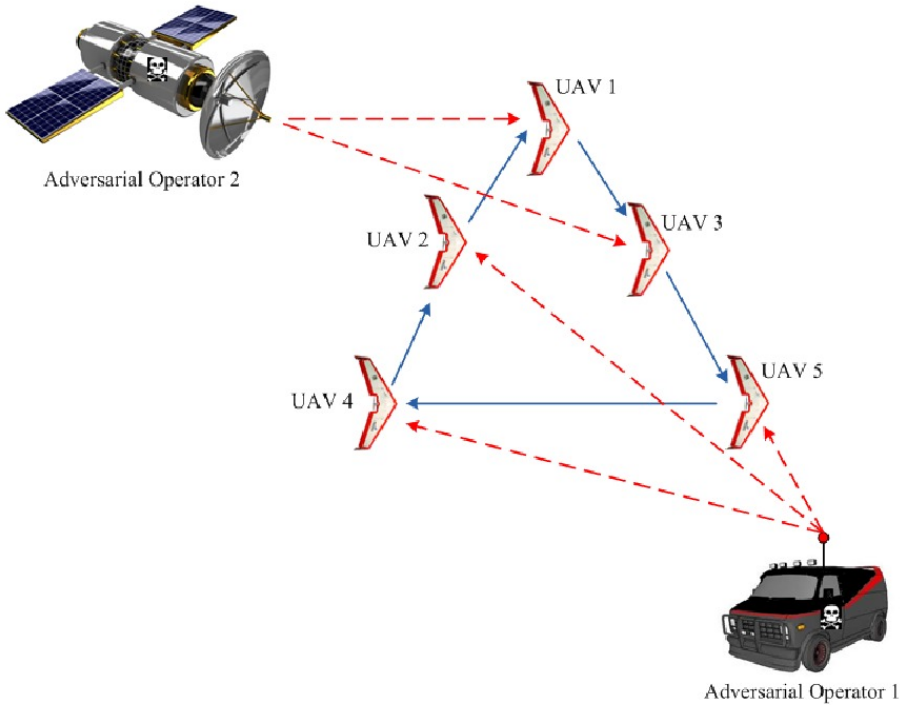

Figure 1. The networked team of UAVs wants to reach a rendez-vous point with minimum energy, while the two adversarial operators are trying to prevent that. The communication between the agents in the team is shown as blue arrows, and the red arrows represent the adversarial inputs.

5. CONCLUSION

In this paper we have derived a game-theory based actor-critic algorithm for reaching approximate consensus of multi agent systems with guaranteed performance and with fixed graph topologies when the dynamics are perturbed by persistent adversaries. Specifically, it is shown that the proposed architecture rejects the adversarial input to achieve consensus and each local controller optimizes its own cost by using only neighborhood information solving in each case a zero-sum game. The optimization is performed online and provides an agreement in position of all the agents in the team. To solve the problem of the curse of dimensionality and the infeasibility of PDEs the algorithm proposed uses three approximators for every agent, one to approximate the cost, one to approximate the control input and one to approximate the adversarial input and thus three tuning laws are being derived for every agent. A Lyapunov stability proof ensures all the signals remain bounded and approximate consensus has been reached. Simulation results show the effectiveness of the proposed approach.

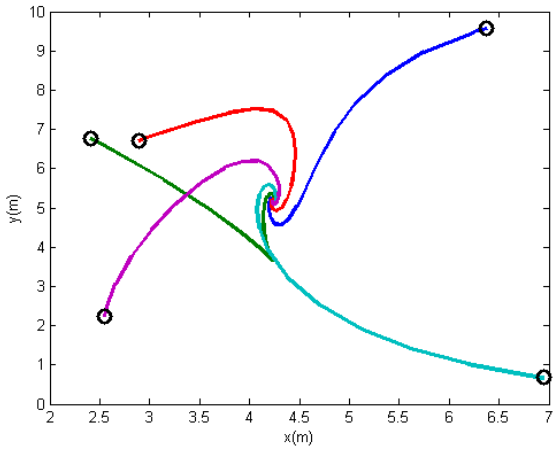

Figure 2. Rendez-vous of the 5 agents

ACKNOWLEDGMENTS

This material is based upon work supported by ARO MURI Grant W911NF0910553 and by ARO Grant W911NF-09-D-0001.

REFERENCES

1. Shoham, Y., and Leyton-Brown, K., *Multiagent systems: algorithmic, gametheoretic, and logical foundations*, Cambridge University Press, 2009
2. Vamvoudakis, K. G., Lewis, F. L., Hudas, G. R., "Multi-Agent Differential Graphical Games: Online Adaptive Learning Solution for Synchronization with Optimality," *Automatica*, vol. 48, no. 8, pp. 1598-1611, 2012
3. Cardenas, A., Amin, S., Sinopoli, B., Giani, A., Perrig, A., and Sastry, S., "Challenges for securing cyber physical systems," *In Workshop on Future Directions in Cyberphysical Systems Security*, 2009
4. Cardenas, A., Amin, S., and Sastry, S., "Secure control: Towards survivable cyber-physical systems," *In First International Workshop on Cyber-Physical Systems (WCPS2008)*, pp. 495500, Beijing, China, 2008
5. Pasqualetti, F., Bicchi, A., and Bullo, F., "Consensus Computation in Unreliable Networks: A System Theoretic Approach," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 90-104, 2012
6. Olfati-Saber, R., Fax, J., Murray, R., "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215-233, 2007
7. Alpcan T., and Basar, T., *Network Security: A Decision and Game Theoretic Approach*, Cambridge University Press, 2010
8. Basar, T., and Olsder, G.J., *Dynamic noncooperative game theory* (2nd ed.), Philadelphia, PA: SIAM, 1999
9. Van der Schaft, A.J., "L2-gain analysis of nonlinear systems and nonlinear state feedback H-∞ control," *IEEE Transactions on Automatic Control*, vol. 37, no. 6, pp. 770-784, 1992
10. Lewis, F. L., and Syrmos, V. L., *Optimal control*, John Wiley, 1995
11. Kearns, M., Littman, M., and Singh, S., "Graphical models for game theory," *In Proc. 17th annual conference on uncertainty in artificial intelligence*, pp. 253-260, 2001
12. Adams, R., and Fournier, J., *Sobolev spaces*, New York: Academic Press, 2003
13. Ioannou, P.A., and Fidan, B., *Adaptive Control Tutorial*, Advances in design and control, SIAM (PA), 2006
14. Krstic, M., and Kanellakopoulos, I., Kokotovic, P.V., *Nonlinear and Adaptive Control Design*, John Wiley and Sons, 1995
15. Pomet J.B., and Praly, L., "Adaptive nonlinear regulation: Estimation from Lyapunov equation," *IEEE Trans. on Autom. Ctrl.*, vol. 37, pp. 729-740, 1992