Decentralized Fictitious Play with Voluntary Communication in Random Communication Networks

Sarper Aydın and Ceyhun Eksin

***Abstract*— We consider autonomous agents communicating over a random communication network that is subject to failures. Each agent aims to maximize its own utility function that depends on the actions of other agents and an unknown state of the environment. Posing this problem as a game, we study a decentralized fictitious play algorithm with a voluntary communication protocol (DFP-V) for Nash equilibrium (NE) computation. In the voluntary communication protocol, each agent locally manages whom to exchange information with by assessing the novelty of its information and the potential effect of its information on others' assessments of their utility functions. We show convergence of the algorithm to a pure NE in finite time for the class of weakly acyclic games. Numerical experiments demonstrate that the voluntary communication protocol reduces number of communication attempts significantly without hampering performance.**

I. Introduction

From engineering to economics, applications of multi-agents systems arise in many different fields including, e.g., robotics, energy systems and cybersecurity. In these systems, autonomous agents are given a common objective. Autonomous agents have to rely on information exchanged over a wireless communication network in order to reach their objectives. In addition to potential failures and packet drops, communication costs energy, and uses limited resources, e.g., bandwidth and time, in wireless networks. When decentralized operations assume perpetual communication between agents to guarantee optimal performance, they may incur unnecessary costs and use of limited resources. The premise of this paper is that communication needs to managed as an integral part of the decentralized operation.

Multi-agent systems with common objectives can be modeled as each agent in the system has its own objective whose value depends on its own action, actions of other agents, and an unknown state of the environment. Agents want to maximize their objectives given their beliefs on the state of the environment. We pose this problem as a game. Joint actions constitute a Nash equilibrium (NE) action profile when each agent maximizes its objective given the maximizing actions of other agents. We consider Nash equilibria as the optimal operating condition of the multi-agent system given common beliefs about the state of the environment. Given the premise above, we design a decentralized game-theoretic learning algorithm that converges to a NE action profile while effectively managing communication attempts.

S. Aydin and C. Eksin are with the Industrial and Systems Engineering Department, Texas A&M University, College Station, TX 77843. E-mail: sarper.aydin@tamu.edu; eksinc@tamu.edu. This work was supported by NSF CCF-2008855.

The decentralized game-theoretic algorithm proposed here is based on fictitious play (FP) [1]–[3]. In FP, each agent takes an action that maximizes its expected utility (best responds) assuming other agents select their actions randomly from a stationary distribution. Agents assume this stationary distribution is given by the past empirical frequency of past actions. Recent works [4]–[7] consider a decentralized form of the fictitious play (DFP) algorithm, in which agents form estimates on empirical frequencies of other agents' actions by averaging the estimates received from their neighbors in a communication network. These algorithms are shown to converge to a NE in weakly acyclic games. However, they rely on perpetual communication between agents in the communication network.

In this paper, we design a decentralized communication protocol for the DFP that allows agents to determine whom to communicate with and when to cease communication (Section II). The communication protocol is based on the idea that agents do not need to send their selections to other agents unless they carry new information. In the context of DFP, we realize this idea by linking novelty of information and potential effect on others' evaluations to two metrics computed respectively as the change in the empirical frequency caused by the current action, and the error others make in estimating the agent's empirical frequency. Our main result (Theorem 1) shows that the DFP-V algorithm with the threshold-based communication protocol converges to a pure NE action profile in finite time given small enough thresholds, if agents' beliefs about the state of the environment weakly converges to a common belief. Numerical experiments demonstrate that the communication protocol can lower the communication attempts by half while showing similar convergence properties as the standard DFP algorithm (Section IV).

The communication-censoring protocols, similar to the communication protocol studied here, are common in distributed optimization algorithms based on, e.g., gradient descent [8], [9], and ADMM [10]. The key contributions of the paper are to propose a novel voluntary communication protocol for the DFP algorithm, a best response type game-theoretic learning algorithm, and to show its convergence for random communication networks.

II. Multi-Agent Systems in Time-Varying Random Networks with Incomplete Information

A multi-agent system consists of a set of agents denoted with $\mathcal{N} = \{1, 2, \cdots, N\}$. Each agent $i \in \mathcal{N}$ selects an action a_i to maximize its utility function $u_i(a_i, a_{-i}, \theta)$ where index

notation $-i$ to represent the set of agents other than $i \in \mathcal{N}$, and $a_{-i} = \{a_j : j \in -i\}$, and θ is the unknown state of the environment. We assume each agent $i \in \mathcal{N}$ chooses among the same set of K actions, i.e., $a_i \in \mathcal{A}_i := \mathcal{A} = \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$ where $\mathbf{e}_k$ is a unit vector whose k^{th} element is 1 and other indices are 0. $a_i = \mathbf{e}_k$ indicates that agent i selects action k. The set of possible states of the environment is given by Θ, and the associated Borel $\sigma-$algebra is $\mathcal{B}(\Theta)$. Given these definitions, the utility function is defined as $u_i : \mathcal{A}^N \times \Theta \to \mathbb{R}$ where $\mathcal{A}^N := \prod_{i\in\mathcal{N}} \mathcal{A}$ is the set of possible action profiles.

We assume agent $i \in \mathcal{N}$ has a belief μ_i on the state of the environment, that is defined as a probability measure $\mu_i(\theta) \in [0, 1]$, for all $\theta \subseteq \Theta$ and $\theta \in \mathcal{B}(\Theta)$. The expected utility of u_i with respect to belief μ_i, for all $i \in \mathcal{N}$, is provided below,

$$u_i(a_i, a_{-i}, \mu_i) = \int_{\theta\in\Theta} u_i(a_i, a_{-i}) d\mu_i(\theta). \tag{1}$$

Assuming the beliefs converge to the same common belief μ over the state of the environment, (see Assumption 1), we can model the multi-agent system as a game defined using the tuple $\Gamma := (\mathcal{N}, \{\mathcal{A}, u_i\}_{i\in\mathcal{N}})$. A Nash equilibrium (NE) of the game Γ is a joint (mixed) action profile $\sigma \in \Delta^N(\mathcal{A})$ such that no agent can have an increment in its utility function u_i with mixed action σ_i, given others' actions σ_{-i}.

Definition 1 (Nash Equilibrium) *The joint (mixed) action profile $\sigma^* = (\sigma_i^*, \sigma_{-i}^*) \in \Delta^N(\mathcal{A})$ is a Nash equilibrium of the game Γ if and only if for all $i \in \mathcal{N}$*

$$u_i(\sigma_i^*, \sigma_{-i}^*, \mu) \geq u_i(\sigma_i, \sigma_{-i}^*, \mu), \quad \text{for all } \sigma_i \in \Delta(\mathcal{A}). \tag{2}$$

A pure NE strategy profile σ^* is a NE $\sigma^* = (\sigma_i^*, \sigma_{-i}^*) \in \Delta^N(\mathcal{A})$, that puts weight 1 on an action profile $a = (a_i, a_{-i}) \in \mathcal{A}^N$.

As mentioned above, agents can determine the equilibrium action profiles in (2), if they have access to actions of other agents and have the same belief on the state of the environment. However, these assumptions are not realistic in a multi-agent networked system, when communication is random and agents' beliefs about the state of the environment is different and evolving.

In the following, we first describe the random communication network model and then the decentralized algorithm.

Communication network.

Given communication attempts are subject to failures, we model the probability of existence of a communication link $c_{ij}(t)$ between agent $i \in \mathcal{N}$ and agent $j \in \mathcal{N} \setminus \{i\}$ at time $t \in \mathbb{N}^+$ as a conditioned Bernoulli random variable,

$$c_{ij}(t) \sim \text{Bernoulli}(p_{ij}(t)), \tag{3}$$

where $0 < \epsilon_{com} \leq p_{ij}(t) < 1$ is the probability of successful communication. This probability is time-varying, and can differ among pairs of agents since several factors such as interference and distance may affect the success of each point-to-point communication.

Given the communication network, agents repeatedly take actions, attempt to send information to agents of their choice, and update beliefs on the state of the environment and on information received from other agents.

A. Decentralized Fictitious Play with Voluntary Communication

In (centralized) FP algorithm, it is assumed that agents repeatedly select actions according to a stationary distribution that is determined by the histogram of their past actions. The histogram, i.e., the empirical frequency $f_i(t)$, is as follows,

$$f_i(t) = (1 - \rho) f_i(t - 1) + \rho a_i(t), \tag{4}$$

where $a_i(t) \in \mathcal{A}$ is the selection of agent i at time $t \in \mathbb{N}^+$ and $\rho \in (0, 1)$ is a fading memory constant assessing the importance of current actions. FP algorithm is a best-response algorithm, where agents maximize their expected utilities with respect to the given empirical frequencies. However, in the decentralized setting, agents do not have immediate access to others' histograms. Rather, each agent i must form its own estimates using information received.

Let the estimate of agent i on agent j's empirical frequency in (4) be denoted as $f_j^i(t)$. The estimate $f_j^i(t)$ belongs to the space of probability distributions on $\mathcal{A}$ denoted as $\Delta(\mathcal{A})$. Then, the expected utility of agent i from taking action $a_i \in \mathcal{A}$ with respect to its estimates $f_{-i}^i(t) := \{f_j^i(t)\}_{j\in\mathcal{N}\setminus\{i\}}$ is given as

$$u_i(a_i, f_{-i}^i(t), \mu_i(t)) = \sum_{a_{-i}\in\mathcal{A}^{N-1}} u_i(a_i, a_{-i}, \mu_i(t)) f_{-i}^i(t)(a_{-i}). \tag{5}$$

We assume agents take actions according to best-response dynamics with inertia,

$$a_i(t) = \begin{cases} \text{argmax}_{a_i\in\mathcal{A}}\, u_i(a_i, f_{-i}^i(t), \mu_i(t)) & \text{w.pr. } 1 - \epsilon, \\ a_i(t - 1) & \text{w.pr. } \epsilon. \end{cases} \tag{6}$$

where $\epsilon \in (0, 1)$. In best-response dynamics with inertia, agents repeat the action taken at previous time step $a_i(t-1)$ with probability ϵ, or take the action take maximizes their expected utility with probability $1-\epsilon$. We note that agent i's action selection depends on its estimate of others' empirical frequencies $f_{-i}^i(t)$ but not on $f_i(t)$. Next, we describe how each agent updates estimates of others' histograms.

Information exchange and estimate updates. At each time step t, agents update their individual empirical frequency $f_i(t)$ in accordance with (4). We let $f_i^i(t) = f^i(t)$ to define agent i's estimate of its own empirical frequency. After updating their own empirical frequencies, agents seek to exchange with each other. Agent i's update for agent $j \in \mathcal{N} \setminus \{i\}$ is as follows,

$$f_j^i(t) = \begin{cases} f_j^j(t), \text{ if } c_{ji}(t) = 1, \\ f_j^i(t - 1), \text{ otherwise}, \end{cases} \tag{7}$$

where $c_{ji}(t)$ is defined in (3). As per (7), each agent can only update its estimate on agent j's selection if agent j is able to transmit its information. We state the voluntary communication protocol next.

Voluntary communication. Agent i decides whether to communicate or not with agent j based on two metrics: (a) novelty of its information $H_{ii}(t) := ||f_i^i(t) - a_i(t)||$; (b) discrepancy between agent j and agent i in estimating i's empirical frequency $f_i^i(t)$, $H_{ij}(t) := ||f_i^i(t) - f_i^j(t)||$. $H_{ii}(t)$ represents the change in the empirical frequency of agent i. $H_{ii}(t)$ becomes small when agent i repeatedly takes the same action as per (4). $H_{ij}(t)$ is the error that agent j is making about agent i's empirical frequency. This error is zero when agent i successfully transmits its empirical frequency as per the update in (7).

If both novelty and discrepancy conditions are respectively below predetermined threshold constants $\eta_1 \in (0,1)$, and $\eta_2 \in (0,1)$, i.e.,

$$H_{ii}(t) < \eta_1 \text{ and } H_{ij}(t) < \eta_2, \tag{8}$$

we let $\mathbb{P}(c_{ij}(t) = 1) = 0$. Condition (8) implies that agent i does not attempt to communicate with agent j when its empirical frequency is changing slowly, and agent j has an accurate estimate of it.

Since, $f_i^j(t)$ is not available to agent i, agent i cannot locally compute $H_{ij}(t)$.We propose an acknowledgement protocol that allows agent i to keep track of $f_i^j(t)$. We assume after each successful communication, the receiver agent j successfully sends an acknowledgement signal (ACK) immediately back to the sender agent i.

B. DFP-V Algorithm

The algorithm described in the previous section is a decentralized fictitious play (DFP) algorithm with voluntary communication, abbreviated as DFP-V. Algorithm 1 outlines the steps of DFP-V.

Algorithm 1 DFP-V for Agent i

1: **Input:** The parameters $\rho, \epsilon, \eta_1, \eta_2$.
2: **Given:** $f_{-i}^i(0)$, $f_i^{-i}(0)$, $\mu_i(0)$, and $a(0)$ for all $i \in \mathcal{N}$.
3: **for** $t = 1, 2, \cdots$ **do**
4: Select an action $a_i(t)$ using (6).
5: Update $f_i^i(t)$ with the selected action via (4).
6: Compute $H_{ii}(t)$ and $H_{ij}(t)$ for all $j \in \mathcal{N} \setminus \{i\}$.
7: Transmit empirical frequency $f_i^i(t)$ to j if $H_{ii}(t) \geq \eta_1$ and $H_{ij}(t) \geq \eta_2$ for all $j \in \mathcal{N} \setminus i$, and then, receives ACK from agent j if transmission is successful, $c_{ij}(t) = 1$, and update $f_i^j(t) = f_i^i(t)$ (7).
8: Update $\{f_j^i(t)\}_{j \in \mathcal{N}}$ using (7).
9: Update $\mu_i(t)$.
10: **end for**

At each time t, agents take the best-response action with inertia (Step 4). Then, each agent i updates its estimate (Step 5). Agent i decides whether to communicate with agent j or not based on the communication metrics ($H_{ii}(t)$ and $H_{ij}(t)$), and condition in (8) (Step 6). Agent i sends the empirical frequencies, and updates on $f_i^j(t)$ using the ACKs (Step 7). Next, agent i updates $f_j^i(t)$ for all $j \in \mathcal{N} \setminus i$ (Step 8), and its belief about the environment (Step 9). Here we do not specify the updates on $\mu_i(t)$ but we will make an assumption about the convergence of agents' beliefs to guarantee convergence of DFP-V in our analysis.

DFP-V algorithm generalizes the DFP algorithm proposed in [11] by including a voluntary communication protocol. In DFP, agents utilize a deterministic communication network structure, and assume repeated communication without failures. The voluntary communication protocol in DFP-V is based on the premise that agents do not need to transmit their information if they have no new information for the receiver. In order to assess whether an agent needs the information available at sender agent i, agent i needs to keep track of the estimates at the potential receivers. Here, this assessment is made possible by an acknowledgement procedure. Assessment of information needs of others distinguishes the voluntary communication protocol from recent communication censoring based protocols used in distributed optimization [8], [9].

III. Convergence Analysis

Here we show convergence of the action profiles under DFP-V to a pure NE of the game in finite time for particular class of games, called weakly acyclic games [12]. Below, we define weakly acyclic games.

Definition 2 (Weakly acyclic Games) *A game Γ is weakly acyclic if from any joint action profile $a = (a_i, a_{-i})$, there exists a best-response path to a pure NE $a^* = (a_i^*, a_{-i}^*)$.*

The existence of a best-response path implies that a (finite) sequence of best-response updates will converge to a pure NE. In our analysis, we make use of the following set of assumptions.

Assumption 1 *There exists a probability measure $\bar{\mu}$ such that agent i's measure on the environment $\mu_i(t)$ converges weakly to $\bar{\mu}$ for all $i \in \mathcal{N}$, i.e., $\mu_i(t) \xrightarrow{w} \bar{\mu}$.*

Note that in Algorithm 1, we were agnostic to the specifics of the update mechanisms for $\mu_i(t)$. This assumption requires that agents need to eventually agree on their estimates about the environment.

Assumption 2 *Agent $j \in \mathcal{N} \setminus \{i\}$ can acknowledge if the information is successfully transmitted from the sender agent i, whenever $c_{ij}(t) = 1$.*

The above assumption makes sure that acknowledgements are received by the sender agent. This is a critical assumption for agents to keep track of others' estimates about their empirical frequency. Given the assumption agents can compute the communication metric $H_{ij}(t)$. Further, usage of acknowledgement procedure does not burden agents' limited

resources compared to the situation without voluntary communication scheme. Since, each agent sends their empirical frequencies with the complexity at least $O(|\mathcal{A}_i|)$, while the acknowledgement signal is just $O(1)$. In addition, because of the 1-bit acknowledgement signal is cheap, it is not demanding to make sure the ACK signal is sent without failure.

Next assumption states that the estimates are measurable with respect to the observations of the agents.

Assumption 3 *Let $\mu(t) = (\mu_1(t), \mu_2(t), \cdots, \mu_N(t))$ be a vector of measures by agents at time t. Then, $\{\mathcal{F}_t\}_{t\geq 0}$ is defined as a filtration with $\mathcal{F}_t := \sigma(\{a(s)\}_{s=1}^t, \{f(s)\}_{s=1}^t, \{\mu(s)\}_{s=1}^t)$. The estimate $f_j^i(t)$ of agent i for agent j's strategy is measurable with respect to $\mathcal{F}_t$.*

Next two assumptions impose certain restrictions on the utility functions.

Assumption 4 *For any pure NE action profile $a^* \in \mathcal{A}^N$ of the game Γ, it holds that,*

$$\{a_i^*\} = \underset{a_i \in \mathcal{A}}{\operatorname{argmax}}\, u_i(a_i, a_{-i}^*, \bar{\mu}). \tag{9}$$

This assumption assures that an agent cannot be indifferent between any two actions if other agents take actions in accordance with a pure NE action profile.

Assumption 5 *The utility functions u_i for all $i \in \mathcal{N}$ are equicontinuous.*

The equicontinuity of the utility function guarantees that if the estimates $f_{-i}^i(t) \in \Delta^{N-1}(\mathcal{A})$ converge to pure strategies $a_{-i} \in \mathcal{A}^{N-1}$ and measures $\mu(t)$ converge to $\bar{\mu}$, the gap between values of utility functions $|u_i(a_i, f_{-i}^i(t), \mu_i(t)) - u_i(a_i, a_{-i}(t), \bar{\mu})|$ goes to 0.

The main convergence result (Theorem 1) relies on two key lemmas: Lemma 5 and Lemma 6. Lemma 5 makes sure that the action profile under DFP-V stays at a pure NE when it is reached with positive probability. Lemma 6 shows that there exists a positive probability of transitioning to a pure NE before agents cease communication. Next we state technical Lemmas 1-4 that are used to prove Lemma 5.

Lemma 1 ***(positive probability of repetition and communication)*** *Suppose Assumption 3 holds and condition in (8) is not true for all $(i,j) \in \mathcal{N} \times \mathcal{N} \setminus \{j\}$. Let E_1 be the event is defined follows ,*

$$E_1(t) = \{a(s) = a, c_{ij}(t+T) = 1, \\ \text{for all } s \in \{t, t+1, \cdots, t+T\}\}, \tag{10}$$

where $a(s)$ is a joint action profile at time s and $c_{ij}(t)$ is the realization of Bernoulli random variable determining communication link between i and j. Then, the probability of the event $E_1(t)$ conditioned on $\mathcal{F}(t)$ is bounded below by a positive constant $\epsilon_1(T)$,

$$\mathbb{P}(E_1(t)|\mathcal{F}_t, \beta_{ij}(t+T) = 1) \geq \epsilon_1(T). \tag{11}$$

Proof: The proof follows from the fact that due to inertia, the probability of repetition of actions in finite T time is always positive. Further, there is a positive probability of communication (at least $p_{ij}(t) > \epsilon_{com}$) before condition (8) is satisfied. Thus the probability of action repetitions and communication is at least $\epsilon_1 = \epsilon^{NT}\epsilon_{com}^{N(N-1)}$. ■

Lemma 2 *Let the empirical frequencies $\{f_i(t)\}_{t\geq 0}$ and estimates of empirical frequencies $\{f_j^i(t)\}_{t\geq 0}$ follow the update rule in DFP-V. Suppose Assumption 2 holds and the event $E_1(t)$ defined in (10) happened. Then, for any $\xi_1 > 0$, there exists a $T > T_1 \in \mathbb{N}_+$, it holds $\|f_i^i(t+T) - \mathbf{e}_k\| < \xi_1$ for all $i \in \mathcal{N}$ and $\|f_i^j(t+T) - \mathbf{e}_k\| < \xi_1$ for all $j \in \mathcal{N} \setminus \{i\}$.*

Proof: Proof is given in the proof to Lemma 6 in [13]. Note that Lemma 6 in [13] is a more general statement because the update (7) is a specific case of the estimate updates used in [13]. ■

Lemma 3 *Suppose Assumptions 1-5 hold, and the event $E_1(t)$ defined in (10) happened. Then, for $t \geq \bar{t}$, the utility function u_i for all $i \in \mathcal{N}$ defined by (1) satisfies the inequality below,*

$$|u_i(a_i, f_{-i}^i(t+T), \mu_i(t)) - u_i(a_i, a_{-i}, \mu)| < \xi_2 \tag{12}$$

Proof : The proof follows by combining the results from Lemma 9 [11] and Lemma 1, and by using the equicontinuity property. ■

Lemma 4 *Suppose Assumptions 1-5 hold, and the event $E_1(t)$ defined in (10) happened. Then for $t \geq \bar{t}$, there exists constants $\xi_1 > 0$ and $\xi_2 > 0$ such that after $T > T_1$ consecutive stages, it holds $\operatorname{argmax}_{a_i \in \mathcal{A}} u_i(a_i, f_{-i}^i(t+T), \mu_i(t+T)) \subseteq \operatorname{argmax}_{a_i \in \mathcal{A}} u_i(a_i, a_{-i}, \mu)$ for all $i \in \mathcal{N}$.*

Proof: The proof follows by using the same steps used in proving Lemma 3 in [13]. ■

Lemma 4 establishes that as the estimates $f^i(t)$ converges to actions by repetition and communication, DFP-V mimics behaviours of a centralized best response updates. Next we state that agents remain at a pure NE, once reached and they are able to communicate their actions to each other.

Lemma 5 ***(absorption property)*** *Suppose Assumption 1-5 hold. Let $a^* \in \mathcal{A}^N$ be a pure NE action profile. Further suppose starting from time $t \geq \bar{t}$, the event $E_1(t)$ defined in (10) happened, with $a(s) = a^*$, for all $s \in \{t, t+1, \cdots, t+T\}$, i.e. a pure NE action profile is repeated T steps and then all pairs of agents communicate at time $t+T$. Then, $a(s) = a^* = (a_1^*, a_2^*, \cdots, a_N^*)$ holds, for all $s \geq t$.*

Proof: By Lemma 4, after repeated actions and successful communication, it holds that $\operatorname{argmax}_{a_i \in \mathcal{A}} u_i(a_i, f_{-i}^i(t+T), \mu_i(t+T)) \subseteq \operatorname{argmax}_{a_i \in \mathcal{A}} u_i(a_i, a{-}i, \mu)$. Since, $a(s) = a^*$ and by Assumption 4, the set of optimal actions given others' actions a_{-i}^* reduces to $\operatorname{argmax}_{a_i \in \mathcal{A}} u_i(a_i, f_{-i}^i(t+$

$T), \mu_i(t+T)) = \text{argmax}_{a_i \in \mathcal{A}} u_i(a_i, a_{-i}, \mu) = \{a_i^*\}$, which is a singleton. Thus, by definition of pure NE (Definition 1), joint action profile stays at the pure NE, i.e., $a(s) = a^*$, for all $s \geq t+T$. ■

Lemma 5 proves that if a NE action profile is repeated consecutively sufficiently long enough, agents will not transition to another joint action profile. The next result indicates that there is a positive probability to reach a NE action profile with small enough communication threshold constants.

Lemma 6 (positive probability of absorption) *Suppose Assumptions 1-5 hold. Let $a(t)$ be the joint action profile at time t and $f^i(t)$ be agent i's estimate on all agents at time t. At time $t > \bar{t}$, we define the following event for all $(i,j) \in \mathcal{N} \times \mathcal{N} \setminus \{j\}$,*

$$\begin{aligned} E_2(t) = &\{a(s) = a^*, c_{ij}(s+T) = 1, \\ &\textit{for all } s \in \{\bar{s}, \bar{s}+1, \cdots, \bar{s}+T\} \\ &\textit{for some } \bar{s} \in \{t, t+1, \cdots, t+K(T_1+T_2)\}\} \end{aligned}$$

where a^* is a pure NE and $c_{ij}(t)$ is the realization of Bernoulli random variable determining communication link between i and j. There exists $\eta_1 > 0$ and $\eta_2 > 0$ small enough such that the transition probability $\mathbb{P}(E_2(t)|\mathcal{F}(t)) \geq \bar{\epsilon}(T_1)$, is bounded below by $\bar{\epsilon}(T_1) > 0$ and always positive for all $t \in \mathcal{T}$.

Proof: The case $a(t) = a^*$, is trivially satisfied by inertia in best response (6). For the case $a(t) \neq a^*$, see that as long as the thresholds $H_{ii}(t)$ and $H_{ij}(t)$ are not satisfied, communication between agents continues with at least probability $\epsilon_{com} > 0$. Then, suppose that the event $E_1(t)$ happened. The probability of this repetition and communication is at least $\epsilon_1(T)$ by Lemma 1.

Since $a(t) \neq a^*$, there is at least one agent that can improve because of the existence of a best-response improvement path. Suppose $a(t)$ is repeated $T > T_1$ stages where T_1 is defined in Lemma 4. By Lemma 4, with sufficiently small η_1 and η_2, it holds that $\text{argmax}_{a_i \in \mathcal{A}} u_i(a_i, f^i_{-i}(t+T), \mu_i(t+T)) \subseteq \text{argmax}_{a_i \in \mathcal{A}} u_i(a_i, a_{-i}, \mu)$. That is, agent i can select a best-response action. The probability that only one agent improves its action and the others stay at the same action is given by $\epsilon_2 := (1-\epsilon)\epsilon^{N-1}$. Once agent i^* takes the best response action at time $t+T$, either $a(t+T) = a^*$ or $a(t+T) \neq a^*$. If $a(t+T) = a^*$, then the probability of absorption is satisfied by Lemma 1. Indeed after the final improvement, the probability of absorption is given by $\epsilon_4 = \epsilon^T \epsilon_{com}^{N-1}$. This is the probability that the action profile is repeated again for $T > T_1$ steps and agent i^* communicated its empirical frequency to all agents at time $t+2T$.

If $a(t+T) \neq a^*$, we need to continue the path of improvement for at most finite $K \in \mathbb{N}^+$ number of steps as per the definition of weakly acyclic games. In order for the next improvement to happen, all agents need to repeat their actions for the next T steps, and agent i^* communicates successfully with all agents at time $t+2T$. The probability of this event happening is $\epsilon_3 = \epsilon^{N(T-1)} \epsilon_{com}^{N-1}$. Thus, we have $\mathbb{P}(E_2(t)|\mathcal{F}(t))$ bounded below by $\bar{\epsilon}(T_1, T_2) = \epsilon_1(\epsilon_2\epsilon_3)^K \epsilon_4$. ■

Lemma 6 states that DFP-V can follow finite-best response path with positive probability. To show this, Lemma 6 utilizes property of inertia and the communication protocol defined by (8). Finally, we state our main convergence theorem.

Theorem 1 *Let $\{a(t) = (a_1(t), (a_2(t), \cdots, a_N(t))\}_{t \geq 1}$ and $\{f^i(t) = (f^i_1(t), f^i_2(t), \cdots, f^i_i(t), \cdots, f^i_N(t))\}_{t \geq 1}$ be a sequence of actions and estimates of each agent $i \in \mathcal{N}$ generated by Algorithm DFP-V. If the assumptions in Lemma 6 hold, then the action sequence $\{a(t)\}_{t \geq 1}$ converges to a pure NE a^* of the game Γ, almost surely.*

Proof: By Lemma 5, when the joint action profile $a(t)$ converges to a pure NE a^*, it stays at a^* forever, i.e., the action profile is absorbed. Hence, the game is played until agents agree on pure NE and absorbed in finite time τ due to existence of positive probability by Lemma 6. ■

IV. Numerical Experiments

We assess the effectiveness of DFP-V on a channel interference game.

A. Channel Interference Game

Channel interference problem arises in communication systems, due to using the same channel. It is common to model the problem as a game with channels representing actions of entities (agents) [14], [15].

Let $\mathcal{A}$ be the set of channels. Agents can only select one channel at each time step t. We use Kronecker delta function, defined as

$$\delta_{ij}(t) = \begin{cases} 1, a_i(t) = a_j(t) \\ 0, a_i(t) \neq a_j(t). \end{cases} \tag{13}$$

to indicate interference at time t caused by the actions of $(i,j) \in \mathcal{N} \times \mathcal{N} \setminus \{j\}$. Agent i aims to select a channel that minimizes the total channel interference it experiences given others' selections. We express this goal with the following utility function,

$$u_i(a_i(t), a_{-i}(t)) = -\sum_{j \in \mathcal{N} \setminus \{i\}} \theta_{ij} \delta_{ij}(t) \tag{14}$$

where $\theta_{ij} \in \mathbb{R}+$ denotes the mutual interference constants between two agents determined by the transmission powers. If these constants are symmetric, i.e., $\theta_{ij} = \theta_{ji}$, the channel interference game is a potential game [16]. In the following, we assume $\{\theta_{ij}\}_{i \in \mathcal{N}, j \in \mathcal{N} \setminus i}$ are symmetric but unknown.

B. Numerical Setup

We consider the channel interference game with $N = 5$ communication nodes and $K = 5$ channels. We assume $\theta_{ij} = M$ where M is uniformly selected from integers between 10 and 15 for each pair $i \in \mathcal{N}$ and $j \in \mathcal{N} \setminus \{i\}$. Agent i learns θ_{ij} by continuously receiving signals drawn randomly from a normal distribution with mean θ_{ij} and standard variance 1 for $j \in \mathcal{N} \setminus \{i\}$. The connection

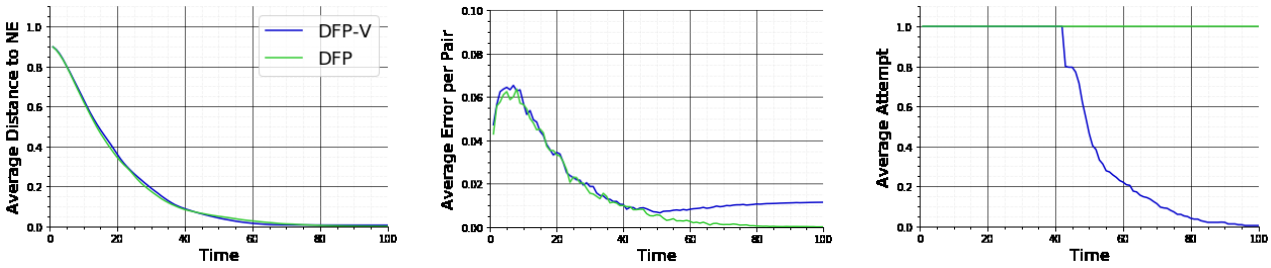

Fig. 1. Convergence results over 50 replications. (Left) Convergence of empirical frequencies to pure NE $\frac{1}{N}\sum_{i\in\mathcal{N}}||f_i^i(t)-a_i^*||$ on average. (Middle) Convergence of estimates $\frac{1}{N(N-1)}\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}\setminus\{i\}}||f_i^i(t)-f_i^j(t)||$. (Right) Average attempt per communication link over time. Agents cease communication starting from $t=40$ on average.

probability $p_{ij}(t)$ is uniformly sampled between 0.4 and 0.6 at each time step for each pair i and j. The parameters ρ and ϵ are selected as 0.3 and 0.1. The voluntary communication parameters are η_1 and η_2 are 0.01 and 0.02.

We compare DFP-V with DFP as a benchmark algorithm with $\eta_1 = 0$, $\eta_2 = 0$. We set the final time step as $T_f = 100$. Fig. 1(Left) shows average convergence rates to a pure NE until time $T_f = 100$. We can conclude that DFP-V achieves the same convergence rate as DFP despite the voluntary communication protocol. That is, the voluntary communication does not adversely affect the convergence to a pure NE. Fig. 1(Middle) shows the total estimation error of agents due to failures in communication attempts. As the initial actions are taken, we observe an increase in the total error as agents begin to select best response actions and communication attempts fail half the time on average. After a peak around $t = 10$, the error gradually drops for both algorithms, due to agents taking the same action more often. The error goes to 0 in DFP algorithm. In contrast, after time around $T_f/2 = 50$, we see a gradual increase in DFP-V due to voluntary communication in accordance with Fig. 1(Right). Fig. 1(Right) shows that agents begin to cease communication starting at time around $T_f/2 = 50$. However, this increase does not lead to a problem in convergence to a pure NE, since, agents are taking the best response actions to the actual actions of each other given their current estimates. The gradual increase is due to the updates in empirical frequencies as per (4). Fig. 1(Right) also shows that agents completely cease communication by the final time T_f while converging to a pure NE in all the runs. The total communication attempts is almost halved by the voluntary communication protocol in comparison to DFP.

V. Conclusion

We considered a decentralized game-theoretic learning algorithm, DFP-V, in which agents actively managed whom to exchange information with. Agents communicated over a random communication network with links prone to failures. The communication protocol was a threshold-based rule that depends on two metrics: novelty of information and potential effect of local information on other's expectation of individual utility function. We considered a communication acknowledgement protocol so that agents are able to compute both metrics locally. We showed that DFP-V converges in finite time to an optimal action profile (pure NE) for the class of weakly acyclic games, which includes the class of potential games. Numerical results demonstrated advantages of the voluntary communication protocol in reducing communication attempts, while retaining convergence properties.

References

[1] G. W. Brown, "Iterative solution of games by fictitious play," *Activity analysis of production and allocation*, vol. 13, no. 1, pp. 374–376, 1951.

[2] H. P. Young, *Strategic learning and its limits*. OUP Oxford, 2004.

[3] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 208–220, 2009.

[4] B. Swenson, S. Kar, and J. Xavier, "Empirical centroid fictitious play: An approach for distributed learning in multi-agent games," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3888 – 3901, 2015.

[5] C. Eksin and A. Ribeiro, "Distributed fictitious play for multiagent systems in uncertain environments," *IEEE Transactions on Automatic Control*, vol. 63, no. 4, pp. 1177–1184, 2017.

[6] B. Swenson, C. Eksin, S. Kar, and A. Ribeiro, "Distributed inertial best-response dynamics," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4294–4300, 2018.

[7] S. Arefizadeh and C. Eksin, "Distributed fictitious play in potential games with time-varying communication networks," *arXiv preprint arXiv:1912.03592*, 2019.

[8] Y. Chen, B. M. Sadler, and R. S. Blum, "Ordered transmission for efficient wireless autonomy," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 1299–1303.

[9] T. Chen, G. Giannakis, T. Sun, and W. Yin, "Lag: Lazily aggregated gradient for communication-efficient distributed learning," in *Advances in Neural Information Processing Systems*, 2018, pp. 5050–5060.

[10] Y. Liu, W. Xu, G. Wu, Z. Tian, and Q. Ling, "Coca: Communication-censored admm for decentralized consensus optimization," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 33–37.

[11] B. Swenson, C. Eksin, S. Kar, and A. Ribeiro, "Fictitious play with inertia learns pure equilibria in distributed games with incomplete information," *Available at ArXiv: http://arxiv. org/pdf/1605.00601 v1. pdf*, 2016.

[12] H. P. Young, "The evolution of conventions," *Econometrica: Journal of the Econometric Society*, pp. 57–84, 1993.

[13] S. Aydin and C. Eksin, "Decentralized learning-aware communication and communication-aware mobility control for the target assignment problem," *Available at ArXiv:https://arxiv.org/abs/2003.03225v2. pdf*, 2020.

[14] B. Babadi and V. Tarokh, "Gadia: A greedy asynchronous distributed interference avoidance algorithm," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6228–6252, 2010.

[15] Q. Wu, Y. Xu, J. Wang, L. Shen, J. Zheng, and A. Anpalagan, "Distributed channel selection in time-varying radio environment: Interference mitigation game with uncoupled stochastic learning," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4524–4538, 2013.

[16] Q. D. Lã, Y. H. Chew, and B.-H. Soong, *Potential Game Theory*. Springer, 2016.