Relay Pursuit of an Evader by a Heterogeneous Group of Pursuers using Potential Games

Yoonjae Lee Efstathios Bakolas

***Abstract*— In this paper, we propose a decentralized game-theoretic pursuit policy for a heterogeneous group of pursuers who individually attempt to, without any prescribed cooperative pursuit strategy, capture a single evader who strives to delay or avoid capture if possible. We assume that the pursuers are rational (self-interested) agents who are not necessarily connected via communication network. Our proposed pursuit policy is motivated from the semi-cooperative pursuit policy called *relay pursuit* [1] under which only the pursuer who can capture the evader faster than the others is active while the rest stay put. In contrast to the latter strategy, our proposed method does not rely on geometric tools. It relies instead on reducing the noncooperative pursuit-evasion game into a sequence of maximum weighted bipartite matching problems which seek to find the pursuer-evader assignments which will result in minimum time of capture. To find the optimal assignment in a decentralized manner, the graph matching problem at each time instant is formulated as a static potential game whose pure strategy Nash equilibria correspond to the optimal assignments. Such equilibria are found by iteratively executing a game-theoretic learning algorithm called Joint Strategy Fictitious Play (JSFP) under which every pursuer synchronously takes his best reply strategy (pursue or stay put), depending on the joint actions of other pursuers, until they reach a Nash equilibrium. We illustrate the performance of our method by means of extensive numerical simulations.**

I. Introduction

Pursuit-evasion games (PEGs) with multiple players receive a lot of attention at present due to their relevance to applications involving decision makers with possibly conflicting objectives. The game of pursuit and evasion, a class of multiplayer dynamic game involving two (or more) competitive players, was first formalized with the emergence of differential game theory [2]. Since then, various types of PEGs have been studied in the eyes of differential game theory [3]–[6]. A special type of PEGs where the pursuers have an additional duty to defend a target object or area from the evaders is receiving a lot of attention [7]–[9] at present. Meanwhile, there have also been, albeit fewer, studies on intelligent evasion strategies [10], [11]. For a broader and more thorough review on recent studies on PEGs, one can refer to [12].

The classical differential game methods, however, often face the curse of dimensionality as the number of players increases. Moreover, they may not provide robust solutions in the presence of constraints due to, say, communication or sensing limitations. The authors of [1], the main inspiration of our present work, developed a practical and easy-to-implement pursuit strategy, named *relay pursuit*, that could deal with the aforementioned constraints. Under the latter strategy, at every time instant, only one pursuer is assigned as the active pursuer who is responsible to capture the evader, whereas the other pursuers remain still. In particular, the active pursuer is assigned by a decentralized algorithm based on dynamic Voronoi diagrams.

The Voronoi-based assignment policy, however, is only applicable to PEGs involving a homogeneous group of pursuers (i.e., every pursuer has the same maximum allowable speed). Decentralized workspace partitioning for a heterogeneous group of pursuers is possible but may degrade the merits of relay pursuit strategy due to its computational intractability [13]. In this paper, we aim to overcome the aforementioned issue, not with geometric tools, but with game theory. In particular, we first formulate the active pursuer selection problem associated with the relay pursuit strategy as an optimization problem. We then reformulate the latter problem as a potential game, a subclass of noncooperative games that has been proven as a powerful tool for the game-theoretic control of multi-agent systems [14], [15]. At last, we show that, via proper selection of the pursuers' individual utilities and learning dynamics, a game-theoretic equivalent of relay pursuit strategy, that is still decentralized but applicable to PEGs involving a heterogeneous group of pursuers, emerges as a commonly accepted pursuit policy among the group of pursuers.

The main contributions of this paper are as follows:

1) We present a formulation of the pursuer-target assignment problem that arises in multiple-pursuer single-evader PEGs as a potential game,
2) We show how to design the individual utilities of the pursuers such that the maximization of the latter contributes to the maximization of their global objective,
3) We utilize game-theoretic learning tools and in particular, the Joint Strategy Fictitious Play (JSFP) with inertia algorithm to compute the pure strategy Nash equilibrium of the game in a decentralized manner.

The rest of the paper is organized as follows. In Section II, we formulate our multiple-pursuer single-evader PEG. In Section III, we formulate the active pursuer selection problem that arises in the PEG as a constrained optimization problem and briefly introduce potential games. In Section IV,

This work was supported in part by ARL under award number W911NF2020085 and NSF under award number CMMI-1753687. Y. Lee (graduate student) and E. Bakolas (Associate Professor) are with the Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas 78712-1221, USA, Emails: yol033@utexas.edu; bakolas@austin.utexas.edu

we reformulate the latter optimization problem as a potential game and design the pursuers' local utilities. In Section V, we present the decentralized game-theoretic learning dynamics based on JSFP with inertia. In Section VI, we present and discuss the results of numerical simulations and, lastly in Section VII, we provide concluding remarks and directions for future research.

II. Formulation of The Pursuit-Evasion Game

We consider a class of multiplayer PEG involving N pursuers and a single evader (target), all of which maneuver in an unbounded 2D plane $\mathbb{R}^2$ where there exist no obstacles that hinder the motions of the players. The group of pursuers, whose index set is denoted by $\mathcal{I}_\mathcal{P} = \{1, ..., N\}$, is assumed to be heterogeneous; i.e., the maximum speed of an individual pursuer may differ from each other. For simplicity, let us denote the i^{th} pursuer as $\mathcal{P}_i$ and the single evader as $\mathcal{E}$. In the proposed PEG, $\mathcal{E}$ continuously strives to avoid or delay her capture, whereas every $\mathcal{P}_i$ attempts to capture the latter only if this increases his local utility. Note that no prescribed cooperative strategies are given to the pursuers *a priori*. Suppose that each player (either pursuer or evader) has an infinite sensing range, which allows them to observe the current state (position) and previous action (velocity vector) of every other player throughout the game. Each player's ensuing action, however, is assumed to be unknown to any other player at any instant of time (even among the pursuers due to, say, communication jamming). Yet, since their previous actions are observable, it is fair to assume that all pursuers have complete information about each other's maximum speed.

Let us denote by $\boldsymbol{x}_i \in \mathbb{R}^2$ (resp., $\boldsymbol{x}_i^0 \in \mathbb{R}^2$) the position of $\mathcal{P}_i$ at time $t \geq 0$ (resp., $t = 0$) where $i \in \mathcal{I}_\mathcal{P}$, and by $\boldsymbol{x}_e \in \mathbb{R}^2$ (resp., $\boldsymbol{x}_e^0 \in \mathbb{R}^2$) the position of $\mathcal{E}$ at time $t \geq 0$ (resp., $t = 0$). Capture of $\mathcal{E}$ by $\mathcal{P}_i$ occurs if there is a time instant $t \geq 0$ at which the latter enters the capture zone of the former, that is, $\boldsymbol{x}_i \in \mathcal{B}_\epsilon(\boldsymbol{x}_e)$ for a given capture radius $\epsilon > 0$, where $\mathcal{B}_\epsilon(\boldsymbol{x}_e)$ denotes the closed ball of radius $\epsilon > 0$ centered at $\boldsymbol{x}_e$, that is, $\mathcal{B}_\epsilon(\boldsymbol{x}_e) := \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x} - \boldsymbol{x}_e\| \leq \epsilon\}$, and $\|\cdot\|$ denotes the Euclidean (vector) norm. Furthermore, we denote by t_c the corresponding time of capture, which is equal to the smallest time $t \geq 0$ at which capture occurs. In addition, all the players follow single integrator dynamics, that is,

$$\dot{\boldsymbol{x}}_i(t) = v_i \boldsymbol{u}_i(t), \qquad \boldsymbol{x}_i(0) = \boldsymbol{x}_i^0, \quad i \in \mathcal{I}_\mathcal{P}, \tag{1}$$

$$\dot{\boldsymbol{x}}_e(t) = v_e \boldsymbol{u}_e(t), \qquad \boldsymbol{x}_e(0) = \boldsymbol{x}_e^0, \tag{2}$$

where $\boldsymbol{u}_i \in \mathcal{U}$ (resp., $\boldsymbol{u}_e \in \mathcal{U}$) denotes the control input of $\mathcal{P}_i$ (resp., $\mathcal{E}$) at time t and $\mathcal{U}$ the input value set of the players which is defined as $\mathcal{U} := \mathcal{S}_1 \cup \{0\}$, where $\mathcal{S}_\epsilon$ denotes the sphere of radius $\epsilon > 0$ centered at the origin, that is, $\mathcal{S}_\epsilon := \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\| = \epsilon\}$.

Note that if $\boldsymbol{u}_i \in \mathcal{S}_1$ (resp., $\boldsymbol{u}_e \in \mathcal{S}_1$), the input corresponds to the direction of motion of $\mathcal{P}_i$ (resp., $\mathcal{E}$) along with the maximum speed v_i (resp., v_e), whereas if $\boldsymbol{u}_i = 0$ (resp., $\boldsymbol{u}_e = 0$), then $\mathcal{P}_i$ (resp., $\mathcal{E}$) does not move. Lastly, let $\mathcal{X}$ denote the augmented state of all the players where $\mathcal{X} = [\boldsymbol{x}_1^\intercal, ..., \boldsymbol{x}_N^\intercal, \boldsymbol{x}_e^\intercal]^\intercal$.

In this paper, we assume that the active pursuers chase the evader by adopting a pure pursuit strategy, that is,

$$\boldsymbol{u}_i^{\text{PP}}(\boldsymbol{x}_i, \boldsymbol{x}_e) = \boldsymbol{\xi}_i / \|\boldsymbol{\xi}_i\|, \tag{3}$$

where $\boldsymbol{\xi}_i = \boldsymbol{x}_e - \boldsymbol{x}_i$ denotes the relative position vector of $\mathcal{E}$ with respect to $\mathcal{P}_i$. The evader, on the other hand, tries to delay or avoid capture by any of the pursuers. To this aim, let $\mathcal{E}$ play a pure evasion strategy against $\mathcal{P}_{i^\dagger}$ where $i^\dagger \in \mathcal{I}_\mathcal{P}$ represents the index of the pursuer who is the closest to $\mathcal{E}$, i.e., $i^\dagger = \arg\min_{i \in \mathcal{I}_\mathcal{P}} \|\boldsymbol{\xi}_i\|$. The pure evasion strategy is described as

$$\boldsymbol{u}_e^{\text{PE}}(\mathcal{X}) = \boldsymbol{\xi}_{i^\dagger} / \|\boldsymbol{\xi}_{i^\dagger}\|. \tag{4}$$

Note that the selection of $i^\dagger$ is irrespective of whether this pursuer is really an active pursuer or not especially because the ensuing actions of pursuers at the next time instant is assumed to be unknown to the evader.

As will be explained in detail in the next section, each pursuer is assumed to be a rational (self-interested) decision maker to fit them in the framework of game theory. In other words, our proposed game-based pursuit policy will not allow cooperative capture of $\mathcal{E}$ by more than one pursuer to occur. Therefore, it is justifiable to assume that $\mathcal{P}_i$ for all $i \in \mathcal{I}_\mathcal{P}$ will always play a two-player game only between himself and $\mathcal{E}$. In turn, the latter assumption leads us to decompose the proposed N-player PEG into N two-player games. Then, along with an additional assumption that each active pursuer employs the pure-pursuit strategy given in (3) and $\mathcal{E}$ employs the pure-evasion strategy given in (4), an estimate of the time-to-capture, which is the time it will take for $\mathcal{P}_i$ to capture $\mathcal{E}$, can be computed by solving the following quadratic equation [11]:

$$(v_e^2 - v_i^2)\phi^2 + 2(\langle \boldsymbol{\xi}_i, v_e \boldsymbol{u}_e \rangle - \epsilon v_i)\phi + \|\boldsymbol{\xi}_i\|^2 = \epsilon^2. \tag{5}$$

As long as $\mathcal{P}_i$ is faster than $\mathcal{E}$ (i.e., $v_i > v_e$), (5) admits a non-negative solution for any $\boldsymbol{\xi}_i \in \mathbb{R}^2$ [11]. Recall, however, that the evader's ensuing control input $\boldsymbol{u}_e$ is not assumed to be known to any of the pursuers *a priori*. Thus, the pursuers cannot predict the exact time that will take them to capture $\mathcal{E}$ since they are not aware of the motion of the evader at the next moment. For this reason, each pursuer will instead compute his time-to-capture under the assumption that $\mathcal{E}$ will continue to apply the pure evasion strategy against the pursuer that she was evading at the previous time instant; we denote such (expected) time-to-capture by $\varphi(\boldsymbol{x}_e, \boldsymbol{x}_i)$.

III. The Active Pursuer Selection Problem

In this section, we formulate the active pursuer selection problem, a special class of pursuer-target assignment problem that arises when applying the relay pursuit strategy to multiple-pursuer single-evader PEGs, as a simple optimization problem. The pursuer-target assignment problem is in essence a bipartite matching problem in which the nodes represent the pursuers and targets and the weighted edges

represent, for example, time-to-capture. The Voronoi-based relay pursuit strategy in [1], for instance, matches the pursuer with minimum time-to-capture to the evader in a geometric way. The active pursuer selection problem under the relay pursuit strategy can be described as follows (for any time $t \geq 0$):

$$\begin{aligned} &\max \Big[-\sum\nolimits_{i\in\mathcal{I}_\mathcal{P}} a_i \varphi(\boldsymbol{x}_e, \boldsymbol{x}_i) \Big], \\ &\text{s.t.} \quad \sum\nolimits_{i\in\mathcal{I}_\mathcal{P}} a_i = 1. \end{aligned} \tag{6}$$

where a_i is an assignment of $\mathcal{P}_i$ (i.e., $a_i = 1$ if $\mathcal{P}_i$ is assigned to the task of capturing $\mathcal{E}$ and $a_i = 0$ otherwise). The goal of the pursuers is to maximize the global objective (6) at every time instant t in a decentralized manner. Note that in this problem the pursuers aim to maximize their current payoffs only. In [1], for instance, the active pursuer is selected in a myopic way under the Voronoi-based assignment policy. In game theory, these pursuers are considered myopic agents who take into account neither state transition nor future payoffs when making their current decisions.

In this section, we present a way to find an optimal solution to (6) by formulating the latter as a (static) potential game [16]. To that end, let us consider a finite N-player strategic form game with the set of players $\mathcal{N}$, admissible action sets $\{\mathcal{A}_i\}$, and local utility functions $\{J_i : \mathcal{A} \to \mathbb{R}\}$, where $\mathcal{A}$ denotes the joint set of admissible actions of the pursuers, i.e., $\mathcal{A} := \times_{i=1}^{|\mathcal{N}|} \mathcal{A}_i$, where $|\cdot|$ denotes the cardinality of a set. Let $\boldsymbol{a} \equiv (a_i, a_{-i})$ denote the joint action profile, where a_i denotes the action of player i and a_{-i} the joint admissible action set of all the players but player i, i.e., $a_{-i} = \{a_j\}_{j\in\mathcal{N}\setminus\{i\}}$. The joint admissible action set of all the players but player i is denoted as $\mathcal{A}_{-i} := \times_{j\in\mathcal{N}\setminus\{i\}} \mathcal{A}_j$. The game discussed above is represented as the tuple $\mathcal{G} := (\mathcal{N}, \{\mathcal{A}_i\}_{i\in\mathcal{N}}, \{J_i\}_{i\in\mathcal{N}})$. Next, we introduce two key definitions in game theory.

Definition 1 (Pure Strategy Nash Equilibrium): The joint action profile $\boldsymbol{a}^\star$ is a pure strategy Nash equilibrium of the game $\mathcal{G}$, if there is no unilateral motive for each player to pick a different (pure strategy) action under the assumption that the other pursuers are committed to their current actions, that is,

$$J_i(a_i^\star, a_{-i}^\star) \geq J_i(a_i, a_{-i}^\star),$$

for all $a_i \in \mathcal{A}_i$ and $i \in \mathcal{N}$.

Note that not all games are guaranteed to possess pure strategy Nash equilibria. To solve (6), however, we need to formulate a game with at least one pure strategy Nash equilibrium such that this equilibrium aligns with the optimal solution of the problem. To that end, let us introduce potential games:

Definition 2 (Exact Potential Game): For a finite game $\mathcal{G}$, if there exists a function $\mathscr{P} : \mathcal{A} \to \mathbb{R}$ such that

$$\begin{aligned} J_i(a_i', a_{-i}) - J_i(a_i, a_{-i}) = \\ \mathscr{P}(a_i', a_{-i}) - \mathscr{P}(a_i, a_{-i}), \end{aligned}$$

for all $a_i, a_i' \in \mathcal{A}_i$, $a_{-i} \in \mathcal{A}_{-i}$, and $i \in \mathcal{I}_\mathcal{P}$, then $\mathcal{G}$ is an exact potential game with exact potential $\mathscr{P}$.

The interpretation of Definition 2 is that the difference in the player's utility caused by a change on his own action leads to the same change on the potential function. Most importantly, any ordinal potential game (a more general class that includes exact potential game) is known to have a finite improvement path along which the potential function monotonically increases, and its end point corresponds to a Nash equilibrium. In other words, potential games are guaranteed to have at least one pure strategy Nash equilibrium.

Finally, we introduce the main problem of our present work, that is the formulation of (6) as a potential game.

Problem 1: Let $\mathcal{G}^t = (\mathcal{I}_\mathcal{P}, \{\mathcal{A}_i\}_{i\in\mathcal{I}_\mathcal{P}}, \{J_i^t\}_{i\in\mathcal{I}_\mathcal{P}})$ be a finite game corresponding to the pursuer-target assignment problem for an arbitrary time $t \geq 0$, where $\mathcal{A}_i = \{0, 1\}$ for all $i \in \mathcal{I}_\mathcal{P}$ (0 means stay and 1 means pursue). Define the local utility functions J_i^t and potential function ψ of the game such that the pursuers can find the maximizer of the global objective in (6) in a decentralized manner. Then, find such optimal joint assignment $\boldsymbol{a}^\star$.

IV. Local Utility Design

A potential game-based optimization method consists of two phases: utility design and learning dynamics design [14]. In this section, we discuss the former phase. The fact that a group of pursuers compete for a limited resource (i.e., target) allows us to view (6) as a distributed resource allocation problem for which there exist various local utility designs [17]. Most of these utilities, however, are exclusively designed for unconstrained optimization problems. Thus, in order to leverage the existing utility designs, we must first remove the constraint on the maximum number of pursuers that can be assigned to the evader in (6). One possible method is to add a penalty function to the global objective in (6) such that if there are more than one pursuers assigned to the target then the penalty dominates the reward. This penalty method, however, may no longer assure the optimality of resulting pure strategy Nash equilibria. Instead, let us claim a new objective function as the following:

$$J^t(\boldsymbol{a}; \mathcal{X}) := \begin{cases} -\frac{1}{\gamma} \sum_{i\in\mathcal{I}_\mathcal{P}} a_i \varphi(\boldsymbol{x}_e, \boldsymbol{x}_i), & \text{if } \gamma > 0 \\ -\infty, & \text{if } \gamma = 0 \end{cases} \tag{7}$$

where $\gamma := \sum_{i\in\mathcal{I}_\mathcal{P}} a_i$ denotes the number of active pursuers under assignment $\boldsymbol{a}$. Note that the optimal joint action profile $\boldsymbol{a}^\star$ that maximizes (7) also maximizes (6), and vice versa. This is because, if there exists a pursuer whose expected time-to-capture is the smallest in the group, the average time-to-capture is at its minimum (i.e., J^t is at its maximum) only if that pursuer is active. In other words, J^t is maximized under the joint action profile $\boldsymbol{a} = (a_{i^\star} = 1, a_{-i^\star} = 0)$ where $i^\star = \arg\min_{i\in\mathcal{I}_\mathcal{P}} \varphi(\boldsymbol{x}_e, \boldsymbol{x}_i)$. Thus, as of this point, we will seek to maximize (7) instead of (6).

In our problem, the global objective (7) is selected as a potential function of $\mathcal{G}^t$ such that an increase in the potential

function aligns with an increase in the global objective; i.e., a pure strategy Nash equilibrium of $\mathcal{G}^t$ is the maximizer of (7) and, since both (6) and (7) share the same optimal solutions, the maximizer of (6). Note that (7) has no limit on the number of pursuers assigned to the evader. Hence, we can apply any proper utility design from the literature, that include, to name a few, the equally shared utility (ESU), wonderful life utility (WLU), and Shapley value (SV) [17]. Among these many choices, we will particularly use the WLU [18] as the local utility function of an individual pursuer, which corresponds to the marginal contribution made by the pursuer to the global objective. More precisely,

$$J_i^t((a_i, a_{-i}); \mathcal{X}) := J^t((a_i, a_{-i}); \mathcal{X}) - J^t((\varnothing, a_{-i}); \mathcal{X}). \quad (8)$$

If $a_i = 0$, regardless of the value of γ, substituting (7) into (8) yields $J_i^t = 0$. If $a_i = 1$ and $\gamma = 1$, on the other hand, $J_i^t = \infty$. Lastly, if $a_i = 1$ and $\gamma > 1$, from (8) we obtain

$$\begin{aligned} &J_i^t((a_i, a_{-i}); \mathcal{X}) \\ &= \frac{1}{\gamma - 1} \sum_{j \in \mathcal{I}_\mathcal{P} \setminus \{i\}} a_j \varphi(\boldsymbol{x}_e, \boldsymbol{x}_j) - \frac{1}{\gamma} \sum_{j \in \mathcal{I}_\mathcal{P}} a_j \varphi(\boldsymbol{x}_e, \boldsymbol{x}_j) \\ &= \frac{1}{\gamma(\gamma - 1)} \Bigg[\gamma \sum\nolimits_{j \in \mathcal{I}_\mathcal{P} \setminus \{i\}} a_j \varphi(\boldsymbol{x}_e, \boldsymbol{x}_j) - \\ &\qquad\qquad (\gamma - 1) \sum\nolimits_{j \in \mathcal{I}_\mathcal{P}} a_j \varphi(\boldsymbol{x}_e, \boldsymbol{x}_j) \Bigg] \\ &= \frac{1}{\gamma(\gamma - 1)} \sum\nolimits_{j \in \mathcal{I}_\mathcal{P}} a_j \varphi(\boldsymbol{x}_e, \boldsymbol{x}_j) - \frac{1}{\gamma - 1} \varphi(\boldsymbol{x}_e, \boldsymbol{x}_i). \end{aligned}$$

For brevity, let $\bar{\varphi}(\mathcal{X}; \boldsymbol{a}) := (1/\gamma) \sum_{i \in \mathcal{I}_\mathcal{P}} a_i \varphi(\boldsymbol{x}_e, \boldsymbol{x}_i)$ denote the average time-to-capture of all the active pursuers. Then the results above are collected to define the local utility function of $\mathcal{P}_i$ as follows:

$$J_i^t((a_i, a_{-i}); \mathcal{X}) := \begin{cases} 0, & \text{if } a_i = 0, \\ \infty, & \text{if } (\gamma = 1) \wedge (a_i = 1), \\ \frac{1}{\gamma - 1} \Big(\bar{\varphi}(\mathcal{X}; \boldsymbol{a}) - \varphi(\boldsymbol{x}_e, \boldsymbol{x}_i) \Big), & \text{if } (\gamma > 1) \wedge (a_i = 1). \end{cases} \quad (9)$$

The interpretation of (9) is that $\mathcal{P}_i$ receives no reward if he chooses to remain still, whereas he receives a positive (resp., negative) reward if the average time-to-capture $\varphi(\boldsymbol{x}_e, \boldsymbol{x}_i)$ is greater (resp., less) than $\bar{\varphi}(\mathcal{X}; \boldsymbol{a})$. If $\mathcal{P}_i$ is currently the only active pursuer in assignment $\boldsymbol{a}$, he receives an infinite reward. Therefore, if the action of chasing the target gives him a positive reward (or there is no other active pursuer), $\mathcal{P}_i$ is encouraged to pursue the target in the ensuing time step.

Proposition 1: Let $v_e < v_i$ for all $i \in \mathcal{I}_\mathcal{P}$. Given J^t and J_i^t defined in (7) and (9), respectively, the game $\mathcal{G}^t = (\mathcal{I}_\mathcal{P}, \{\mathcal{A}_i\}_{i \in \mathcal{I}_\mathcal{P}}, \{J_i^t\}_{i \in \mathcal{I}_\mathcal{P}})$ is an exact potential game with exact potential $\mathscr{P} = J$ for any instant of time $t \geq 0$.

Proof: By setting $\mathscr{P} = J^t$, $\mathcal{G}^t$ is classified as an identical interests plus dummy game [19] in which the local utility function J_i^t corresponds to the sum of an identical interest function $\mathcal{F}$, which is a function of the joint action profile of all the pursuers, $\boldsymbol{a}$, and a dummy function $\mathcal{Q}$, which is a function of the joint action profile of all the pursuers but $\mathcal{P}_i$, a_{-i}. More precisely,

$$J_i^t((a_i, a_{-i}); \mathcal{X}) = \mathcal{F}((a_i, a_{-i}); \mathcal{X}, t) + \mathcal{Q}_i((\varnothing, a_{-i}); \mathcal{X}, t), \quad \forall a_i \in \mathcal{A}_i,\ a_{-i} \in \mathcal{A}_{-i},\ i \in \mathcal{I}_\mathcal{P}.$$

Comparing with (8), we see that $\mathcal{F}(\boldsymbol{a}; \mathcal{X}) \equiv J^t(\boldsymbol{a}; \mathcal{X})$ and $\mathcal{Q}_i(a_{-i}; \mathcal{X}) \equiv J^t((\varnothing, a_{-i}); \mathcal{X})$. Since the identical interest function $\mathcal{F}$ is an exact potential of an identical interests plus dummy game [19], we conclude that $\mathcal{G}^t$ is an exact potential game with exact potential $\mathscr{P} = J^t$ for any time $t \geq 0$. ■

V. JSFP-Based Active Pursuer Selection

Given the fact that $\mathcal{G}^t$ is a potential game, there exist a variety of iterative game-theoretic learning algorithms that can discover a pure strategy Nash equilibrium in a decentralized manner. Such learning algorithms include fictitious play (FP), regret matching (RM), spatial adaptive play (SAP) [14], and log-linear learning (LLL) [20]. Because in our problem the group of pursuers must arrive at consensus within a short time interval before the evader runs away, synchronous algorithms (FP and RM) are preferred over asynchronous algorithms (SAP and LLL) as the former type of algorithms generally shows faster convergence time [14]. Under FP, for instance, every pursuer follows best reply dynamics using the history of actions of the entire group. Note that an individual pursuer can obtain information about actions of other pursuers either by communicating with them or observing their actions. The latter algorithm, however, can be computationally intractable since each player must track the history of action profiles. To avoid this issue, we will use JSFP with inertia, an improved FP algorithm suggested in [21] that alleviates the aforementioned computational intractability but still ensures convergence to a pure strategy Nash equilibrium almost surely for any (ordinal) potential game.

The pseudocode of the JSFP-based learning dynamics for active pursuer selection at time t is provided in Algorithm 1. Under this algorithm, at the initial simulation step $k = 0$, every pursuer picks a random action from his admissible action set (Line 3). Thereafter, each one of the pursuers iterates to take his best response action (conditional on the previous joint action profile of the other pursuers that he observed) until there are no more best response actions for the group to take. Once converged, the pursuers execute their actions, move forward to the next time instant, and repeat the algorithm until the evader is captured. The main difference between JSFP and FP is that, at every simulation step, instead of recording the action history, each pursuer updates his expected utility function based on his observation of other pursuers' previous joint action profile. Note that, at each iteration, two different candidate best response actions

are computed (Line 9 and 10) as follows:

$$\beta_i(a_{-i}[k];\mathcal{X},t) = \arg\max_{a_i\in\mathcal{A}_i} J_i^t((a_i,a_{-i}[k]);\mathcal{X}),$$
$$\check{\beta}_i(k;\mathcal{X},t) = \arg\max_{a_i\in\mathcal{A}_i} \bar{J}_i^t(k;\mathcal{X}).$$

The average utility function $\bar{J}_i^t$ is propagated according to the following recursive equation:

$$\bar{J}_i^t(k+1;\mathcal{X}) = \frac{k}{k+1}\bar{J}_i^t(k;\mathcal{X}) + \frac{1}{k+1}J_i^t((a_i,a_{-i}[k]);\mathcal{X}).$$

Then, at every simulation step k, given a time-invariant inertia parameter $\alpha \in (0,1)$, $\check{a}_i[k]$ is selected as the next action $a_i^\star[k+1]$ with probability of α (Line 12), and, if $\check{a}_i[k]$ is not selected (with probability of $1-\alpha$), then $\mathcal{P}_i$ resumes to take the previous action $a_i^\star[k-1]$ (Line 14). The latter process repeats until the terminal condition (Line 17) meets. The converged action profile $\boldsymbol{a}^\star$ at the point of termination is proven to be a pure strategy Nash equilibrium of our potential game [22], thus this $\boldsymbol{a}^\star$ corresponds to a pure strategy Nash equilibrium of $\mathcal{G}^t$. For more details on JSFP, the reader can refer to [21], [22].

Algorithm 1 JSFP-based active pursuer selection at time t

1: $k = 0$
2: **for** $i \in \mathcal{I}_\mathcal{P}$ in parallel **do**
3: $\quad a_i^\star[k] = \texttt{RandSample}(\mathcal{A}_i, 1)$
4: **end for**
5: CONVERGED = FALSE
6: **while** CONVERGED $\neq$ TRUE **do**
7: $\quad k = k+1$
8: $\quad$ **for** $i \in \mathcal{I}_\mathcal{P}$ in parallel **do**
9: $\quad\quad a_i[k] = \beta_i(a_{-i}[k-1];\mathcal{X},t)$
10: $\quad\quad \check{a}_i[k] = \check{\beta}_i(k-1;\mathcal{X},t)$
11: $\quad\quad$ **if** $\texttt{rand}() < \alpha$ **then**
12: $\quad\quad\quad a_i^\star[k] = \check{a}_i[k]$
13: $\quad\quad$ **else**
14: $\quad\quad\quad a_i^\star[k] = a_i^\star[k-1]$
15: $\quad\quad$ **end if**
16: $\quad$ **end for**
17: $\quad$ **if** $\bigwedge_{i=1}^N((a_i[k] \equiv a_i^\star[k]) \wedge (a_i[k] \equiv \check{a}_i[k-1]))$ **then**
18: $\quad\quad$ CONVERGED = TRUE
19: $\quad$ **end if**
20: **end while**

VI. Numerical Simulations

In this section, we present numerical simulation results obtained with the application of the JSFP-based active pursuer selection algorithm (Algorithm 1) for the solution of Problem 1 and implement the latter algorithm in a multiple-pursuer single-evader PEG. We define the sampling time of the PEG as $\Delta t = 0.1$ and the simulation step of the potential game in Algorithm 1 for each instant of time as δt where we assume $\delta t \ll \Delta t$. Since each simulation step is much less than the sampling time, we assume an infinite number of simulation steps can fit in one time interval (between t and $t+\Delta t$ for any $t \geq 0$). Furthermore, we assume that each simulation step is so short that any state transitions that occur during that time interval is negligible; in other words, the states of all the players are fixed to be constant while Algorithm 1 is executed. This assumption is necessary since the game $\mathcal{G}^t$ for all time $t \geq 0$ must be a static game such that the game can be played as a repeated game under the game-theoretic learning dynamics we employ to find the Nash equilibrium.

The parameters chosen for numerical simulations are as follows: $\boldsymbol{x}_i^0 \in [0,10]^2$ and $v_i \in [1,2]$ for all $i \in \mathcal{I}_\mathcal{P}$, $\boldsymbol{x}_e^0 \in [0,10]^2$, $v_e = 0.9$, and $\epsilon = 0.1$. We first test the stability and convergence time of Algorithm 1 by executing the algorithm for 90 different episodes with various values of the following design parameters: N (number of pursuers) and α (inertia). For each episode, a pair of N and α are respectively chosen from the sets $\{1,...,10\}$ and $\{0.1,...,0.9\}$. The average convergence time of each episode is then computed by iterating the episode 1000 times and averaging the results. Figure 1, which illustrates the average convergence times of all the 90 episodes, clearly shows that, 1) Algorithm 1 always converges under the given simulation setup and 2) the convergence time increases as N increases or α decreases. The latter is an expected result because, if the value of α is small, the pursuers are more likely to insist on choosing their previous action over exploring, which slows down the algorithm from converging.

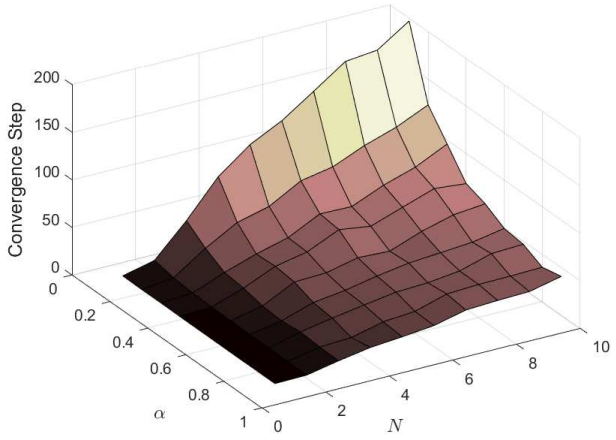
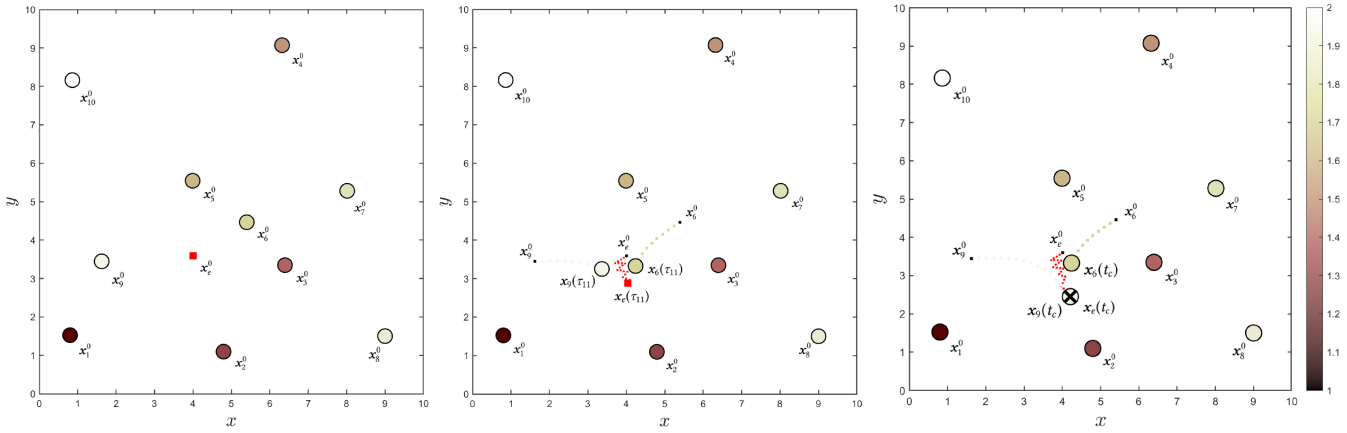


Fig. 1. Convergence of Algorithm 1 (JSFP with inertia)

For the main simulation presented in Figure 2, we choose $N = 10$ and $\alpha = 0.8$ (for fast convergence, we choose large α herein). Figure 2(a) shows the initial positions of all the players, wherein the pursuers are represented as circles and the evader ($\mathcal{E}$) as a red square. Note that the colors of the pursuers indicate their maximum allowable speeds; the lighter the colors are, the faster the pursuers are. Figure 2(b) and 2(c) show the pursuers' positions and trajectories up to time τ_{11} and t_c, respectively, where τ_j for $j \in \{1,\ldots,11\}$, denotes a switching time at which the active pursuer is newly assigned. Throughout the game, there are total 11 switching times between $\mathcal{P}_6$ and $\mathcal{P}_9$. From τ_1 to τ_{11}, either $\mathcal{P}_6$ or $\mathcal{P}_9$ is selected as the active pursuer to pursue (denoted as $\rightsquigarrow$) $\mathcal{E}$. An interesting observation is that $\mathcal{E}$ initially follows a zigzag



(a) $t = 0$ (b) $\boldsymbol{x}_6$ or $\boldsymbol{x}_9 \overset{t}{\rightsquigarrow} \boldsymbol{x}_e,\ \forall t \in [0, \tau_{11})$ (c) $\boldsymbol{x}_9 \overset{t}{\rightsquigarrow} \boldsymbol{x}_e,\ \forall t \in [\tau_{11}, t_c)$

Fig. 2. Relay pursuit using Algorithm 1 ($N = 10, \alpha = 0.8$)

path to delay the capture, whereas after τ_{11}, as $\mathcal{P}_9$ approaches closely, she starts to pure-evade him. For all $t \geq 0$, the rest of the pursuers stay in their initial positions. These simulation results show that, for a class of PEGs involving multiple pursuers and a single evader, individually selfish action by the pursuers guided by Algorithm 1 naturally induces a semi-cooperative team pursuit strategy equivalent to *relay pursuit*.

VII. Conclusions

In this paper, we have presented a decentralized game-theoretic pursuit policy for multiple pursuers to capture the evader. Via this pursuit policy, the pursuers essentially find a solution to the pursuer-target assignment problem associated with the multiplayer pursuit-evasion games involving a single evader and a heterogeneous group of pursuers who are not necessarily cooperative. Our proposed method, which relies on potential games, assume all the pursuers are rational decision makers who selfishly attempt to maximize their individual utilities, whereas such noncooperative behaviors synchronize to maximize the global objective. By setting the local utility of each pursuer equal to his marginal contribution to the global objective and by executing JSFP with inertia, a semi-cooperative group pursuit strategy that resembles *relay pursuit* is obtained among the noncooperative pursuers. In our future work, we will consider an extension of our proposed method to multiple-pursuer multiple-evader PEGs.

References

[1] E. Bakolas and P. Tsiotras, "Relay pursuit of a maneuvering target using dynamic Voronoi diagrams," *Automatica*, vol. 48, no. 9, pp. 2213–2220, 2012.

[2] R. Isaacs, *Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization*. Courier Corporation, 1999.

[3] Y. Ho, A. Bryson, and S. Baron, "Differential games and optimal pursuit-evasion strategies," *IEEE Transactions on Automatic Control*, vol. 10, no. 4, pp. 385–389, 1965.

[4] M. Pachter and P. Wasz, "On a two cutters and fugitive ship differential game," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 913–917, 2019.

[5] A. Von Moll, D. Casbeer, E. Garcia, D. Milutinović, and M. Pachter, "The multi-pursuer single-evader game," *Journal of Intelligent & Robotic Systems*, vol. 96, no. 2, pp. 193–207, 2019.

[6] E. Garcia, D. W. Casbeer, A. Von Moll, and M. Pachter, "Multiple pursuer multiple evader differential games," *IEEE Transactions on Automatic Control*, 2020.

[7] A. Von Moll, E. Garcia, D. Casbeer, M. Suresh, and S. C. Swar, "Multiple-pursuer, single-evader border defense differential game," *Journal of Aerospace Information Systems*, vol. 17, no. 8, pp. 407–416, 2020.

[8] R. Yan, Z. Shi, and Y. Zhong, "Reach-avoid games with two defenders and one attacker: An analytical approach," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1035–1046, 2018.

[9] R. Yan, Z. Shi, and Y. Zhong, "Task assignment for multiplayer reach–avoid games in convex domains via analytical barriers," *IEEE Transactions on Robotics*, vol. 36, no. 1, pp. 107–124, 2019.

[10] J. Selvakumar and E. Bakolas, "Evasion with terminal constraints from a group of pursuers using a matrix game formulation," in *2017 American Control Conference (ACC)*, pp. 1604–1609, IEEE, 2017.

[11] J. Selvakumar and E. Bakolas, "Feedback strategies for a reach-avoid game with a single evader and multiple pursuers," *IEEE Transactions on Cybernetics*, 2019.

[12] I. E. Weintraub, M. Pachter, and E. Garcia, "An introduction to pursuit-evasion differential games," in *2020 American Control Conference (ACC)*, pp. 1049–1066, IEEE, 2020.

[13] E. Bakolas, "Workspace partitioning and topology discovery algorithms for heterogeneous multi-agent networks," *IEEE Transactions on Control of Network Systems*, 2020.

[14] G. Arslan, J. R. Marden, and J. S. Shamma, "Autonomous Vehicle-Target Assignment: A Game-Theoretical Formulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 129, pp. 584–596, 04 2007.

[15] J. R. Marden, G. Arslan, and J. S. Shamma, "Cooperative control and potential games," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 6, pp. 1393–1407, 2009.

[16] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[17] J. R. Marden and A. Wierman, "Distributed welfare games," *Operations Research*, vol. 61, no. 1, pp. 155–168, 2013.

[18] D. H. Wolpert and K. Tumer, "Optimal payoff functions for members of collectives," in *Modeling complexity in economic and social systems*, pp. 355–369, World Scientific, 2002.

[19] J. P. Hespanha, *Noncooperative game theory: An introduction for engineers and computer scientists*. Princeton University Press, 2017.

[20] J. R. Marden and J. S. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," *Games and Economic Behavior*, vol. 75, no. 2, pp. 788–808, 2012.

[21] 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.

[22] H.-L. Choi and S.-J. Lee, "A potential-game approach for information-maximizing cooperative planning of sensor networks," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 6, pp. 2326–2335, 2015.