Autonomous Vehicle-Target Assignment: A Game-Theoretical Formulation

Gürdal Arslan
Department of Electrical Engineering,
University of Hawaii,
Manoa, Honolulu, HI 96822
e-mail: gurdal@hawaii.edu

Jason R. Marden
e-mail: marden@ucla.edu

Jeff S. Shamma
e-mail: shamma@ucla.edu

Department of Mechanical and Aerospace Engineering,
University of California, Los Angeles,
Los Angeles, CA 90095

We consider an autonomous vehicle-target assignment problem where a group of vehicles are expected to optimally assign themselves to a set of targets. We introduce a game-theoretical formulation of the problem in which the vehicles are viewed as self-interested decision makers. Thus, we seek the optimization of a global utility function through autonomous vehicles that are capable of making individually rational decisions to optimize their own utility functions. The first important aspect of the problem is to choose the utility functions of the vehicles in such a way that the objectives of the vehicles are localized to each vehicle yet aligned with a global utility function. The second important aspect of the problem is to equip the vehicles with an appropriate negotiation mechanism by which each vehicle pursues the optimization of its own utility function. We present several design procedures and accompanying caveats for vehicle utility design. We present two new negotiation mechanisms, namely, "generalized regret monitoring with fading memory and inertia" and "selective spatial adaptive play," and provide accompanying proofs of their convergence. Finally, we present simulations that illustrate how vehicle negotiations can consistently lead to near-optimal assignments provided that the utilities of the vehicles are designed appropriately. [DOI: 10.1115/1.2766722]

1 Introduction

Designing autonomous vehicles with intelligent and coordinated action capabilities to achieve an overall objective is a major part of the recent theme of "cooperative control," which has received significant attention in recent years. Whereas much of the work in this area focuses on "kinetic" coordination, e.g., multivehicle trajectory generation (e.g., [1], and references therein), the focus here is on *strategic* coordination. In particular, we consider an autonomous vehicle-target assignment problem (illustrated in Fig. 1), where a group of vehicles are expected to assign themselves to a set of targets to optimize a global utility function. When viewed as a combinatorial optimization problem, the vehicle-target assignment problem considered in this paper is a generalization of the well-known weapon-target assignment problem [2] to the case where the global utility is a general function of the vehicle-target assignments. In its full generality, the weapon-target assignment problem is known to be nondeterministic-polynomial-time-complete [2], and the existing literature on the weapon-target assignment problem is concentrated on heuristic methods to quickly obtain near optimal assignments in relatively large instances of the problem—very often with no guarantees on the degree of suboptimality (cf., [3], and references therein). Therefore, from an optimization viewpoint, the vehicle-target assignment problem considered in this paper is, in general, a hard problem, even though optimal assignments can be obtained quite efficiently in very special cases.

Our viewpoint in this paper deviates from that of direct optimization. Rather, we emphasize the design of vehicles that are individually capable of making coordination decisions to optimize their own utilities, which then indirectly translates to the optimization of a global utility function. The main potential benefit of this approach is to enable autonomous vehicles that are individually capable of operating in uncertain and adversarial environments, with limited information, communication, and computation, to autonomously optimize a global utility. The optimization methods available in the literature are not suitable for our purposes because even a distributed implementation of such optimization algorithms need not induce "individually rational" behavior, which is the key to realize the expected benefits of our approach. Furthermore, an optimization approach would typically require constant dissemination of global information throughout the network of the vehicles as well as increased communication and computation.

Accordingly, in this paper we formulate our autonomous vehicle-target assignment problem as a multiplayer game [4,5], where each vehicle is interested in optimizing its own utility. We use the notion of pure Nash equilibrium to represent the assignments that are agreeable to the rational vehicles, i.e., the assignments at which there is no incentive for any vehicle to unilaterally deviate. We use algorithms for multiplayer learning in games as negotiation mechanisms by which the vehicles seek to optimize their utilities. The problem of optimizing a global utility function by the autonomous vehicles then reduces to the proper design of (i) the vehicle utilities and (ii) the negotiation mechanisms.

Designing vehicle utilities is essential to obtaining desirable collective behavior through self-interested vehicles (cf., [6]). An important consideration in designing the vehicle utilities is that the vehicle utility functions should be "aligned" with the global utility function in the sense that agreeable assignments (i.e., Nash equilibria) should lead to high, ideally maximal, global utility. There are multiple ways that such alignment can be achieved. An obvious instance is to set the vehicle utilities equal the global utility. This choice is not desirable in the case of a large number of interaction vehicles, because another consideration in designing the vehicle utilities is that the vehicle utilities should be "localized," i.e., a vehicle's utility should depend only on the local information available to the vehicle. For example, in a large vehicle-target assignment problem, the vehicles may have range restrictions and a vehicle may not even be aware of the targets and/or the vehicles outside its range. In such a case, a vehicle whose utility is set to the global utility would not have sufficient information to compute its own utility. Therefore, a vehicle's utility should be localized to its range while maintaining the align-

Contributed by the Dynamic Systems, Measurement, and Control Division of ASME for publication in the JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT, AND CONTROL. Manuscript received March 31, 2006; final manuscript received April 1, 2007. Review conducted by Tal Shima.

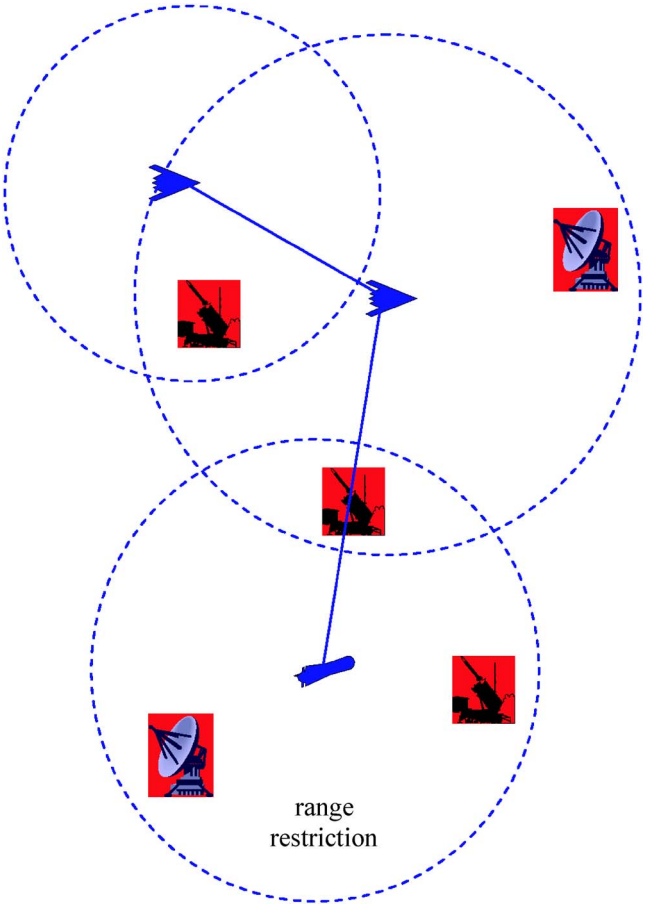

Fig. 1 Illustration of vehicle target assignment

ment with the global utility. More generally, we will discuss the properties of being aligned and localized for several utility design procedures in Sec. 3.

Obtaining optimal assignments using the approach presented in this paper also requires that the vehicles use a negotiation mechanism that is convergent in the multiplayer game induced by the vehicle utilities. We will show that when vehicle utilities are aligned with the global utility, they always lead to a class of games known as "ordinal potential games" [7]. The significance of this connection is that certain multiplayer learning algorithms, such as fictitious play (FP) [8], are known to converge in potential games, and hence can be used as vehicle negotiation mechanisms. However, FP has an intensive informational requirement. Spatial adaptive play (SAP) [9] is another such algorithm, which leads to an optimizer of the potential function in potential games with arbitrarily high probability. Although SAP reduces the information requirement, there can be a high implementation cost when vehicles have a large number of possible actions.

This paper goes beyond existing work in the area through the introduction of new negotiating mechanisms that alleviate the informational and implementation requirement, namely, "generalized regret monitoring with fading memory and inertia" and "selective spatial adaptive play." We establish new convergence results for both algorithms and simulate their performance on an illustrative weapon-target assignment problem.

The remainder of this paper is organized as follows. Section 2 sets up an autonomous vehicle-target assignment problem as a multiplayer game. Section 3 discusses the issue of designing the utility functions of the vehicles that are localized to each vehicle yet aligned with a given global utility function. Section 4 reviews selected learning algorithms available in the literature and presents two new algorithms, along with convergence results, that offer some advantages over existing algorithms. Section 5 present some simulation results to illustrate the possibility of obtaining near optimal assignments through vehicle negotiations. Finally, Section 6 contains some concluding remarks.

2 Game-Theoretical Formulation of an Autonomous Vehicle-Target Assignment Problem

We begin by considering an optimal assignment problem where n_v vehicles are to be assigned to n_t targets. Each entity, whether a vehicle or a target, may have different characteristics. The vehicles are labeled as $\mathcal{V}_1,\ldots,\mathcal{V}_{n_v}$, and the targets are labeled as $\mathcal{T}_0,\mathcal{T}_1,\ldots,\mathcal{T}_{n_t}$, where a fictitious target $\mathcal{T}_0$ represents the "null target" or "no target." Let $\mathcal{V}:=\{\mathcal{V}_1,\ldots,\mathcal{V}_{n_v}\}$ and $\mathcal{T}:=\{\mathcal{T}_0,\mathcal{T}_1,\ldots,\mathcal{T}_{n_t}\}$. A vehicle can be assigned to any target in its range, denoted by $\mathcal{A}_i\subset\mathcal{T}$ for vehicle $\mathcal{V}_i\in\mathcal{V}$. The null target always satisfies $\mathcal{T}_0\in\mathcal{A}_i$. Let $\mathcal{A}:=\mathcal{A}_1\times\cdots\times\mathcal{A}_{n_v}$. The assignment of vehicle $\mathcal{V}_i$ is denoted by $a_i\in\mathcal{A}_i$, and the collection of vehicle assignments $(a_1,\ldots,a_{n_v})$, called the assignment profile, is denoted by a. Each assignment profile, $a\in\mathcal{A}$, corresponds to a global utility, $U_g(a)$, that can be interpreted as the objective of a global planner.

We view the vehicles as "autonomous" decision makers, and accordingly, each vehicle, e.g., vehicle $\mathcal{V}_i\in\mathcal{V}$, is assumed to select its own target assignment, $a_i\in\mathcal{A}_i$, to maximize its own utility function, $U_{\mathcal{V}_i}(a)$. In general, vehicle utility functions may be different and each of them may depend on the whole assignment profile a. Hence, the vehicles do not necessarily face an optimization problem, but rather, they face a (finite) multiplayer game. In such a setting, the vehicles are to negotiate an assignment profile that is mutually agreeable. The autonomous target assignment problem is to design the utilities, $U_{\mathcal{V}_i}(a)$, as well as appropriate negotiation procedures so that the vehicles can negotiate a mutually agreeable target assignment that yields maximal global utility, $U_g(a)$.

To be able to deal with the intricacies of our autonomous target assignment problem, we adopt some concepts and methods from the theory of games [4,5]. We start with the concept of equilibrium to characterize the target assignments that are agreeable to the vehicles. A well-known equilibrium concept for multiplayer games is the notion of Nash equilibrium. In the context of an autonomous target assignment problem, a Nash equilibrium is an assignment profile $a^*=(a_1^*,\ldots,a_{n_v}^*)$ such that no vehicle could improve its utility by unilaterally deviating from a^*. Before introducing the notion of Nash equilibrium in more precise terms, we will introduce some notation. Let a_{-i} denote the collection of the target assignments of the vehicles *other than* vehicle $\mathcal{V}_i$, i.e.,

$$a_{-i}=(a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_{n_v})$$

and let

$$\mathcal{A}_{-i}:=\mathcal{A}_1\times\ldots\times\mathcal{A}_{i-1}\times\mathcal{A}_{i+1}\times\ldots\times\mathcal{A}_{n_v}$$

With this notation, we will sometimes write an assignment profile a as (a_i,a_{-i}). Similarly, we may write $U_{\mathcal{V}_i}(a)$ as $U_{\mathcal{V}_i}(a_i,a_{-i})$. Using the above notation, an assignment profile a^* is called a *pure Nash equilibrium* if, for all vehicles $\mathcal{V}_i\in\mathcal{V}$,

$$U_{\mathcal{V}_i}(a_i^*,a_{-i}^*)=\max_{a_i\in\mathcal{A}_i}U_{\mathcal{V}_i}(a_i,a_{-i}^*) \tag{1}$$

In this paper, we will represent the agreeable target assignment profiles by the set of pure Nash equilibria even though in the literature some non-Nash solution concepts for multiplayer games are also available. We will introduce one such concept called *efficiency* for future reference. An assignment profile is called efficient if there is no other assignment that yields higher utilities to all vehicles. For given vehicle utilities, a Nash equilibrium assignment may or may not be efficient. Our justification of a pure Nash equilibrium as an agreeable assignment is based on the autono-

mous and self-interested nature of the vehicles. Clearly, an efficient pure Nash equilibrium should be more appealing to the vehicles than an inefficient pure Nash equilibrium.

In general, a pure Nash equilibrium may not exist for an arbitrary set of vehicle utilities. However, as will be seen in Sec. 3, any reasonable set of vehicle utilities tailored to the autonomous vehicle-target problem would have at least one pure Nash equilibrium.

We conclude this section with the definition of potential games and ordinal potential games [7]. These games form an important class of games because of their relevance to autonomous vehicle-target assignment as well as their desirable convergence properties mentioned earlier.

DEFINITION 2.1 ([ORDINAL] POTENTIAL GAMES). A potential game *consists of vehicle utilities*, $U_{\mathcal{V}_i}(a)$, $\mathcal{V}_i \in \mathcal{V}$, *and a potential function*, $\phi(a): \mathcal{A} \mapsto \mathcal{R}$, *such that*, *for every vehicle*, $\mathcal{V}_i \in \mathcal{V}$, *for every* $a_{-i} \in \mathcal{A}_{-i}$, *and for every* $a_i', a_i'' \in \mathcal{A}_i$,

$$U_{\mathcal{V}_i}(a_i', a_{-i}) - U_{\mathcal{V}_i}(a_i'', a_{-i}) = \phi(a_i', a_{-i}) - \phi(a_i'', a_{-i})$$

An ordinal potential game *consists of vehicle utilities* $U_{\mathcal{V}_i}(a)$, $\mathcal{V}_i \in \mathcal{V}$, *and a potential function* $\phi(a): \mathcal{A} \mapsto \mathcal{R}$ *such that*, *for every vehicle* $\mathcal{V}_i \in \mathcal{V}$, *for every* $a_{-i} \in \mathcal{A}_{-i}$, *and for every* $a_i', a_i'' \in \mathcal{A}_i$,

$$U_{\mathcal{V}_i}(a_i', a_{-i}) - U_{\mathcal{V}_i}(a_i'', a_{-i}) > 0 \Leftrightarrow \phi(a_i', a_{-i}) - \phi(a_i'', a_{-i}) > 0$$

In a potential game, the difference in utility received by any one vehicle for its two different target choices, when the assignments of other vehicles are fixed, can be measured by a potential function that only depends on the assignment profile and not on the label of any vehicle.

In an ordinal potential game, an improvement in utility received by any one vehicle for its two different target choices, when the assignments of other vehicles are fixed, always results in an improvement of a potential function that, again, only depends on the assignment profile and not on the label of any vehicle. Clearly, ordinal potential games form a broader class than potential games.

3 Utility Design

In this section, we discuss various important aspects of designing the vehicle utilities to achieve a high global utility. We cite [7,10] as the key references for this section, since we freely use some of the terminology and the ideas presented in them. To make the discussion more concrete and relevant, we assume a certain structure for the global utility, even though it is possible to present the ideas at a more abstract level. We assume that all vehicles that assign themselves to a particular target form a team and engage their common target in a coordinated manner. An engagement with target $\mathcal{T}_j \in \mathcal{T}$ generates some utility denoted by $U_{\mathcal{T}_j}(a)$; $U_{\mathcal{T}_0}(a)=0$ for any a.

It is important to distinguish between a target utility, $U_{\mathcal{T}_j}(a)$, and a vehicle utility, $U_{\mathcal{V}_i}(a)$. The realized target utility represents the overall value for engaging target $\mathcal{T}_j$, whereas a vehicle utility partly reflects vehicle $\mathcal{V}_i$'s share of that value. Furthermore, it may be that vehicle $\mathcal{V}_i$ shares this reward even if it did not engage target $\mathcal{T}_i$. This will depend on the final specification of vehicle utilities.

We will assume that the utility generated by an engagement with target $\mathcal{T}_j$ depends only on the characteristics of target $\mathcal{T}_j$ and the vehicles engaging target $\mathcal{T}_j$. This is stated more precisely in the following assumption.

ASSUMPTION 3.1. *Let* a *and* $\tilde{a}$ *be two action profiles in* $\mathcal{A}$, *and for any target*, $\mathcal{T}_j \in \mathcal{T}$, *define the sets*

$$S_j = \{\mathcal{V}_i \in \mathcal{V} | a_i = \mathcal{T}_j\} \quad \text{and} \quad \tilde{S}_j = \{\mathcal{V}_i \in \mathcal{V} | \tilde{a}_i = \mathcal{T}_j\}$$

Then,

$$S_j = \tilde{S}_j \Rightarrow U_{\mathcal{T}_j}(a) = U_{\mathcal{T}_j}(\tilde{a})$$

We now define the global utility to be the total sum of the utilities generated by all engagements, i.e.,

$$U_g(a) = \sum_{\mathcal{T}_j \in \mathcal{T}} U_{\mathcal{T}_j}(a) \tag{2}$$

This summation is only one approach to aggregate the target utility functions. See [11] for a more general discussion from the perspective of multiobjective optimization.

It will be convenient to model an engagement with a target as a random event that is assumed to be independent of the other target engagements. At the end of an engagement, the target and some of the engaging vehicles are destroyed with certain probability. The statistics of the outcome of an engagement depend on the characteristics of the target as well as the composition of the engaging vehicles. As an example, it may be the case that only a particular team of vehicles may destroy a particular target with reasonable probability. In this case, the utility generated by an engagement is taken to be the expected difference between the value of a destroyed target and the total value of the destroyed vehicles. These issues are discussed further for the well-known weapon-target assignment problem in Sec. 5.

An important consideration in specifying the vehicle utilities, $U_{\mathcal{V}_i}(a)$, $i=1,\ldots,n_v$, is to make them "aligned" with the global utility, $U_g(a)$. Ideally, this means that the vehicles can only agree on an *optimal* assignment profile, i.e., an assignment profile that maximizes the global utility. Because it is not always straightforward to achieve the alignment of the vehicle utilities with the global utility in this ideal sense (without first calculating an optimal assignment), we adopt a more relaxed notion of alignment from [10]. That is, a vehicle can improve its own utility by unilateral action if and only if the same unilateral action also improves the global utility.

DEFINITION 3.1 (ALIGNMENT). *We will say that a set of vehicle utilities* $U_{\mathcal{V}_i}(a)$, $\mathcal{V}_i \in \mathcal{V}$, *is* aligned[1] *with the global utility* $U_g(a)$ *when the following condition is satisfied. For every vehicle*, $\mathcal{V}_i \in \mathcal{V}$, *for every* $a_{-i} \in \mathcal{A}_{-i}$, *and for every* $a_i', a_i'' \in \mathcal{A}_i$,

$$U_{\mathcal{V}_i}(a_i', a_{-i}) - U_{\mathcal{V}_i}(a_i'', a_{-i}) > 0 \Leftrightarrow U_g(a_i', a_{-i}) - U_g(a_i'', a_{-i}) > 0 \tag{3}$$

We see that the notion of alignment coincides with the notion of ordinal potential games in Definition 2.1.

It turns out that alignment does not rule out pure Nash equilibria that may be suboptimal from the global utility perspective. Moreover, such suboptimal pure Nash equilibria may even yield the highest utilities to all vehicles and hence may be efficient. Nevertheless, alignment also guarantees that the optimal assignment profiles are always included in the set of pure Nash equilibria; hence, they are agreeable to the vehicles even though they may be inefficient.

The above discussion on alignment is summarized by the following proposition, whose proof is straightforward.

PROPOSITION 3.1. *Let* a_{opt} *denote an optimal assignment profile*, *i.e.*,

$$a_{\text{opt}} \in \arg\max_{a \in \mathcal{A}} U_g(a)$$

Under the alignment condition (3), *the resulting game is an ordinal potential game that has* a_{opt} *as a* (*possibly nonunique*) *pure Nash equilibrium.*

3.1 Identical Interest Utility (IIU). One obvious, but ultimately ineffective, way of making the vehicle utilities aligned with the global utility is to set all vehicle utilities to the global utility. In game-theory terminology, setting

$$U_{\mathcal{V}_i}(a) = U_g(a), \quad \text{for all vehicles} \quad \mathcal{V}_i \in \mathcal{V} \tag{4}$$

[1]The notion of alignment we adopt here is called *factoredness* in [10].

results in an *identical interest game*. Obviously, an identical interest game with $U_{\mathcal{V}_i}(a)=U_g(a)$, for all vehicles $\mathcal{V}_i \in \mathcal{V}$, is also a potential game with the potential $U_g(a)$, and hence, the vehicle utilities (4) are aligned with the global utility. In fact, optimal assignments in this case yield the highest vehicle utilities and therefore are efficient. However, suboptimal Nash equilibria may still exist.

As will be seen later, the vehicles negotiate by proposing targets and responding to the previous target assignment proposals that are exchanged among the vehicles. Each vehicle whose utility is set to the global utility needs to know (i) the proposals made by all other vehicles as well as (ii) the characteristics of all the vehicles and the targets to be able to generate a new proposal. The reason for this is that vehicle $\mathcal{V}_i$'s utility would depend on all engagements with all targets, including those that are not in $\mathcal{A}_i$. Therefore, when the vehicle utilities are set to the global utility, continuous dissemination of global information is required among the vehicles.

3.2 Range-Restricted Utility (RRU). A possible way of making the vehicle utilities more localized than IIU would be to set the utility of vehicle $\mathcal{V}_i$ equal to the sum of the utilities generated by the engagements with the targets that belong to vehicle $\mathcal{V}_i$'s target set $\mathcal{A}_i$, i.e.,

$$U_{\mathcal{V}_i}(a) = \sum_{\mathcal{T}_j \in \mathcal{A}_i} U_{\mathcal{T}_j}(a), \quad \text{for all vehicles} \quad \mathcal{V}_i \in \mathcal{V} \tag{5}$$

Note that in this case the global information requirement on the vehicles is alleviated. Moreover, the vehicle utilities (5) are still aligned with the global utility. This guarantees that the optimal assignments are agreeable to the vehicles, but they may be inefficient; see Example 3.3. In fact, the vehicle utilities lead to a *potential game*; see [7]. The following proposition is an immediate consequence of Assumption 3.1.

PROPOSITION 3.2. *Vehicle utilities that satisfy* (5) *form a potential game with the global utility* $U_g(a)$ *serving as a potential function*.

Note that when all vehicles have the same set of available targets, i.e., $\mathcal{A}_1=\cdots=\mathcal{A}_{n_v}$, then (5) leads to an identical interest game.

A concern regarding vehicle utilities (4) (and possibly (5)) stems from the so-called learnability issue introduced in [10]. That is, a vehicle may not be able to influence its own utility in a significant way when a large number of vehicles can assign themselves to the same large set of targets. In this case, since the utility of a vehicle is the total sum of the utilities generated by a large number of engagements involving a large number of targets and vehicles, the proposals made by an individual vehicle may not have any significant effect on its own utility. Hence, a negotiating vehicle may find itself approximately indifferent to the available target choices if the negotiation mechanism employed is utility based, i.e., the vehicle proposes targets in response to the actual utilities corresponding to its past proposals, as in reinforcement learning.

3.3 Equally Shared Utility (ESU). One way to limit the influence of other vehicles on vehicle $\mathcal{V}_i$'s utility is to set

$$U_{\mathcal{V}_i}(a) = \frac{U_{\mathcal{T}_j}(a)}{n_{\mathcal{T}_j}(a)}, \quad \text{if} \quad a_i = \mathcal{T}_j \tag{6}$$

where $n_{\mathcal{T}_j}(a)$ is the total number of vehicles engaging target $\mathcal{T}_j$. The rationale behind (6) is to distribute the utility generated by an engagement equally among the engaging vehicles. Note that in this case vehicle $\mathcal{V}_i$'s utility is independent of the engagements to which vehicle $\mathcal{V}_i$ does not participate.

Even though the total sum of vehicle utilities (6) equals the global utility, it turns out that (6) need *not* be exactly aligned with the global utility.

$\mathcal{V}_1$: \ $\mathcal{V}_2$:	$\mathcal{T}_0$	$\mathcal{T}_1$	$\mathcal{T}_2$
$\mathcal{T}_0$	0, 0	0, 2	0, 10
$\mathcal{T}_1$	2, 0	1, 1	2, 10
$\mathcal{T}_2$	10, 0	10, 2	5, 5

Fig. 2 Misaligned vehicle utilities

Example 3.1. Consider two targets $\mathcal{T}_1$ and $\mathcal{T}_2$ with values 2 and 10, respectively, and two anonymous vehicles $\mathcal{V}_1$ and $\mathcal{V}_2$, i.e., $\mathcal{V}_1$ and $\mathcal{V}_2$ have identical characteristics. Assume that each vehicle is individually capable of destroying any one of the targets with probability 1, while the targets in no case have any chance of destroying any of the vehicles. The vehicle utilities in this example can be represented in the matrix form, shown in Fig. 2, where if vehicle $\mathcal{V}_i \in \{\mathcal{V}_1, \mathcal{V}_2\}$ chooses target $a_i \in \{\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2\}$ then the first number (respectively the second number) in the entry (a_1, a_2) represents the utility to the first vehicle (respectively to the second vehicle). The global planner would of course prefer each vehicle to engage a different target, since this would yield a maximal global utility 12. However, such an optimal assignment profile might leave the vehicle engaging the low-value target unsatisfied with a utility 2, and this unsatisfied vehicle might be able to improve its utility to 5 by unilaterally switching to the high-value target at the expense of lowering the global utility to 10. Because of the misalignment of (6) with the global utility in this example, an optimal assignment profile may not be agreeable by all vehicles, whereas the vehicles may find the suboptimal Nash equilibrium assignment $(a_1, a_2)=(\mathcal{T}_2, \mathcal{T}_2)$ agreeable.

However, in the case of *anonymous* vehicles, (6) does lead to a potential game.

DEFINITION 3.2 (ANONYMITY). *Vehicles are* anonymous *if for any permutation*

$$\sigma: \{1,2,\ldots,n_v\} \to \{1,2,\ldots,n_v\}$$

and for any two assignments, a and $\tilde{a}$, *related by*

$$\tilde{a}_i = a_{\sigma(i)}, \quad \forall\, i \in \{1,2,\ldots,n_v\}$$

the equality

$$U_{\mathcal{T}_j}(a) = U_{\mathcal{T}_j}(\tilde{a})$$

holds for any target $\mathcal{T}_j$.

As the terminology implies, the utility generated by an engagement with a target does not depend on the identities of the vehicles engaging the target, but only the number of vehicles engaging the target.

PROPOSITION 3.3. *Anonymous vehicles with utilities that satisfy* (6) *form a potential game with potential function*

$$\phi(a) = \sum_{\mathcal{T}_j \in \mathcal{T}} \sum_{\ell=1}^{n_{\mathcal{T}_j}(a)} \frac{U_{\mathcal{T}_j}(\ell)}{\ell}$$

where $n_{\mathcal{T}_j}(a)$ *is the total number of vehicles assigned to target* $\mathcal{T}_j$ *and* $U_{\mathcal{T}_j}(\ell)$ *is the utility generated by an engagement of* ℓ *anonymous vehicles with target* $\mathcal{T}_j$.

Hence, in the case of anonymous vehicles, (6) is aligned with the above potential function, which is the same potential function introduced in [12] in the context of so-called congestion games, but different from the global utility function $U_g(a)$. The significance of this observation is that the existence of a potential function associated with the vehicle utilities guarantees the existence of agreeable (possibly suboptimal) assignment profiles in the form

$\mathcal{V}_1$:	$\mathcal{V}_2$: $\mathcal{T}_0$	$\mathcal{T}_1$	$\mathcal{T}_2$
$\mathcal{T}_0$	0,0	0,−2	0,−2
$\mathcal{T}_1$	10,0	4,4	10,−2
$\mathcal{T}_2$	10,0	10,−2	4,4

Fig. 3 Misaligned vehicle utilities with no pure Nash equilibrium

of pure Nash equilibria. Furthermore, there exist learning algorithms that are known to converge in potential games and these convergent learning algorithms can be used by the vehicles as negotiation mechanisms always leading to a settlement on an assignment profile. If the vehicles are not anonymous, then the misalignment of the vehicle utilities (6) with the global utility can be even more severe.

Example 3.2. Consider two targets $\mathcal{T}_1$ and $\mathcal{T}_2$ with values 10 each, and two distinguishable vehicles, $\mathcal{V}_1$ and $\mathcal{V}_2$, with values 2 each. Assume that vehicle $\mathcal{V}_1$ is individually capable of destroying any one of the targets with probability one, and not one of the targets is ever capable of destroying $\mathcal{V}_1$. Assume further that vehicle $\mathcal{V}_2$ is never capable of destroying any of the targets, and any one of the targets can destroy vehicle $\mathcal{V}_2$ with probability one. This setup leads to the vehicle utilities shown in Fig. 3. In this example, the two vehicles may not be able to agree on any assignment profile, optimal or suboptimal, because while vehicle $\mathcal{V}_1$ would be better off by engaging a target alone, vehicle $\mathcal{V}_2$ would be better off by engaging a target together with vehicle $\mathcal{V}_1$. Yet, the global planner would prefer vehicle $\mathcal{V}_1$ engaging one of the targets and vehicle $\mathcal{V}_2$ not engaging any target. If these two vehicles were to use a negotiation mechanism that allows settlement only on a pure Nash equilibrium, then they would not be able to agree on any assignment because a pure Nash equilibrium does not exist in this example. A mixed, but not pure, Nash equilibrium is still guaranteed to exist, but would not lead to an agreement on a particular assignment. Therefore, in the distinguishable vehicles case, the vehicle utilities (6) might lead to a situation where the vehicles are not only in conflict with the global planner but also in conflict among themselves.

3.4 Wonderful Life Utility (WLU). A solution to the problem of designing individual utility functions that are more learnable than (4) or (5) and still aligned with the global utility is offered in [10] in the form of a family of utility structures called the *wonderful life utility*. In our context, a particular WLU structure would be obtained by setting the utility of a vehicle to the marginal contribution made by the vehicle to the global utility, i.e.,

$$U_{\mathcal{V}_i}(a_i,a_{-i}) = U_g(a_i,a_{-i}) - U_g(\mathcal{T}_0,a_{-i}), \quad \text{for all vehicles} \quad \mathcal{V}_i \in \mathcal{V} \tag{7}$$

From the definition of the global utility (2), the WLU (7) can be written as

$$U_{\mathcal{V}_i}(a_i,a_{-i}) = U_{\mathcal{T}_j}(a_i,a_{-i}) - U_{\mathcal{T}_j}(\mathcal{T}_0,a_{-i}), \quad \text{if} \quad a_i = \mathcal{T}_j$$

for all vehicles $\mathcal{V}_i \in \mathcal{V}$, which means that the utility of a vehicle is its marginal contribution to the utility generated by the engagement that the vehicle participates. WLU is expected to make each vehicle's utility more learnable by removing the unnecessary dependencies on other vehicles' assignment decisions, while still keeping the vehicle utilities aligned with the global utility. It turns out that WLU (7) also leads to a potential game with the global utility being the potential function.

PROPOSITION 3.4. *Vehicle utilities that satisfy* (7) *form a potential game with the global utility* $U_g(a)$ *serving as a potential function*.

Another interpretation of the WLU is that a vehicle is rewarded with a side payment equal to the externality it may create by not assigning itself to any target, which is the idea behind "internalizing the externalities" in economics [13].

3.5 Comparisons. Each of the vehicle utilities IIU (4), RRU (5), and WLU (7) lead to a potential game with the globally utility function being the potential function, and hence, they are aligned with the global utility. This guarantees that the optimal assignments are in each case included in the set of pure Nash equilibria. However, in each case, there may also be suboptimal Nash equilibria that may be pure and/or mixed. There is ample evidence in the literature that a mixed equilibrium cannot emerge as a stable outcome of vehicle negotiations, particularly in potential games (e.g., [14]). However, a suboptimal pure Nash equilibrium can emerge as a stable outcome, depending on the negotiation mechanism used by the vehicles.

Example 3.3. Consider $N>2$ vehicles, $\mathcal{V}_1,\ldots,\mathcal{V}_N$, and $N+1$ targets, $\mathcal{T}_1,\ldots,\mathcal{T}_{N+1}$, where $\mathcal{A}_i=\{\mathcal{T}_i,\mathcal{T}_{N+1}\}$. Assume that any vehicle $\mathcal{V}_i$ engaging target $\mathcal{T}_i$ generates 1 unit of utility. Assume also that an engagement with target $\mathcal{T}_{N+1}$ generates 0 utility unless all vehicles engage $\mathcal{T}_{N+1}$ in which case they generate 2 units of utility. Clearly, the optimal assignment is given by $a^*=(\mathcal{T}_1,\mathcal{T}_2,\ldots,\mathcal{T}_N)$. The optimal assignment profile a^* is a pure Nash equilibrium when the vehicle utilities are given by any of (4) and (5), or (7). However, there is another pure Nash equilibrium $a^{**}=(\mathcal{T}_{N+1},\mathcal{T}_{N+1},\ldots,\mathcal{T}_{N+1})$ for any of vehicle utilities (4) and (5), or (7) which is suboptimal with respect to the global utility. The global utility and the vehicle utilities corresponding to a^* and a^{**} are summarized as follows:

$$U_g(a^*) = N \; U_g(a^{**}) = 2$$

$$U_{\mathcal{V}_i}(a^*) = N \; U_{\mathcal{V}_i}(a^{**}) = 2 \text{ if vehicles utilities are given by (4)}$$

$$U_{\mathcal{V}_i}(a^*) = 1 \; U_{\mathcal{V}_i}(a^{**}) = 2 \text{ if vehicles utilities are given by (5) or (7)}$$

Note that the optimality gap $N-2$ between a^* and a^{**} can be arbitrarily large for large N. Note also that if the vehicle utilities are given by RRU (5) or WLU (7) the suboptimal Nash equilibrium a^{**} yields higher utilities to all vehicles than the optimal Nash equilibrium a^*.

In the case of RRU or WLU, if the negotiation mechanism employed by the vehicles were to eliminate the inefficient assignment profiles, the vehicles would never be able to agree on the optimal assignment a^*. This example illustrates the fact that the vehicle utilities cannot be designed independently of the negotiation mechanism employed by the vehicles.

4 Negotiation Mechanisms

The issue of which Nash equilibrium will emerge as a stable outcome of vehicle negotiations is studied under the topic of *equilibrium selection* in game theory. In this section, we will discuss equilibrium selection and other important properties of some negotiation mechanisms. In particular, we will present a negotiation mechanism from the literature that leads to an optimal Nash equilibrium in potential games with arbitrarily high probability.

We will adopt various learning algorithms available in the literature for multiplayer games as vehicle negotiation mechanisms to make use of the theoretical and computational tools provided by game theory. The negotiation mechanisms that will be presented in this section will provide the vehicles with strategic decision-making capabilities. In particular, each vehicle will negotiate with other vehicles without any knowledge about the utilities of the other vehicles. One of the reasons for such a requirement is that the vehicles may not have the same information regarding their environment. For example, a vehicle may not know all the targets and/or the potential collaborating vehicles

available to another vehicle and, moreover, it may not be possible to pass on such information due to limited communication bandwidth. Another reason for the private utilities requirement is to make the vehicles truly autonomous in the sense that each vehicle is individually capable of making robust strategic decisions in uncertain and adversarial environments. In this case, any individual vehicle is cooperative with the other vehicles only to the extent that cooperation helps the vehicle to maximize its own utility, which is, of course, carefully designed by the global planner.

Accordingly, we will consider some negotiation mechanisms that require each vehicle to know, at most, its own utility function, the proposals made by the vehicle itself, and the proposals made by those other vehicles that can influence the utility of the vehicle. We will review these negotiation mechanisms in terms of convergence, equilibrium selection, and computational efficiency. We will present our review primarily in the context of potential games, since many of the vehicle utility structures considered in Sec. 3 fall into this category. In some cases, we will point to existing results in the literature, while in some other cases we will point to open problems.

4.1 Review: Selected Recursive Averaging Algorithms

4.1.1 Action-Based Fictitious Play. Action-based fictitious play, or simply FP, was originally introduced as a computational method to calculate the Nash equilibria in zero-sum games [15], but later proposed as a learning mechanism in multi-player games (cf., [8]).

One can also think of FP as a negotiation mechanism employed by the vehicles to select their targets. At each negotiation step, $k=1,2,\ldots$, vehicles simultaneously propose targets

$$a(k) := [a_1(k), \ldots, a_{n_v}(k)]$$

where $a_i(k) \in \mathcal{A}_i$ is the label of the target proposed by vehicle $\mathcal{V}_i$. The objective is to construct a negotiation mechanism so that the proposed assignments, $a(k)$, ultimately converge for large k. FP is one such mechanism that is guaranteed to converge for potential games.

In FP, the target assignment proposals at stage k are functions of past proposed assignments over the interval $[1,k-1]$ as follows. First, enumerate the targets available to vehicle $\mathcal{V}_i$ as $\mathcal{A}_i=\{A_i^1,\ldots,A_i^{|\mathcal{A}_i|}\}$. For any target index $j \in [1,|\mathcal{A}_i|]$, let $n_j(k;\mathcal{V}_i)$ denote the total number of times vehicle $\mathcal{V}_i$ proposed target A_i^j up to stage k. Now define the *empirical frequency* vector, $q_i(k) \in \mathbb{R}^{|\mathcal{A}_i|}$, of vehicle $\mathcal{V}_i$ as follows:

$$q_i(k) = \left(\frac{n_1(k-1;\mathcal{V}_i)}{k-1}, \quad \frac{n_2(k-1;\mathcal{V}_i)}{k-1}, \ldots, \frac{n_{|\mathcal{A}_i|}(k-1;\mathcal{V}_i)}{k-1} \right)$$

In words, $q_i(k)$ reflects the histogram of proposed target assignments by vehicle $\mathcal{V}_i$ over the interval $[1,k-1]$. Note that the elements of the empirical frequency vector are all positive and sum to unity. Therefore, $q_i(k)$ can be identified with a probability vector on the probability simplex of dimension $|\mathcal{A}_i|$.

We are now set to define the FP process. At stage k, vehicle $\mathcal{V}_i$ selects its proposed assignment $a_i(k) \in \mathcal{A}_i$ in accordance with maximizing its expected utility *as though* all other vehicles make a simultaneous and independent random selection of their actions, a_{-i}, based on the product distribution defined by empirical frequencies, $q_1(k),\ldots,q_{i-1}(k),q_{i+1}(k),\ldots,q_{n_v}(k)$, i.e.,

$$a_i(k) \in \arg\max_{\alpha \in \mathcal{A}_i} \mathbf{E}_{a_{-i}}[U_{\mathcal{V}_i}(\alpha, a_{-i})]$$

In case the maximizer is not unique, then any maximizer will do.

One appealing property of FP is that the empirical frequencies generated by FP converge to the set of Nash equilibria in potential games [7,16]. Although the empirical frequencies may converge to a mixed Nash equilibrium while the proposals are cycling (see the related churning issue in [17]), it is generally believed that convergence of empirical frequencies to a mixed (but not pure) Nash equilibrium happens rarely when vehicle utilities are not equivalent to a zero sum game [18,19]. Thus, if the vehicles negotiate using FP and their utilities constitute a potential game, then in most cases we can expect them to asymptotically reach an agreement on an assignment profile. We should also mention numerous stochastic versions of FP with similar convergence properties [20].

The main disadvantage of FP for the purposes of this paper is its computational burden on each vehicle. The most computationally intensive operation is the optimization of the utilities during the negotiations, which effectively requires an enumeration of all possible combined assignments by other vehicles [21,22]. This makes FP computationally prohibitive when there are large numbers of vehicles with large target sets. To make FP truly scalable, it is clear that the vehicles need to evaluate their utilities more directly without using the empirical frequencies.

4.1.2 Utility-Based FP. The distinction between action-based and utility-based FP, see [23,24], is that the vehicles predict their utilities during the negotiations based on the actual utilities corresponding to the previous proposals. Utility Based FP is in essence a multi-agent Reinforcement Learning algorithm [25,26]. The difference is that in reinforcement learning, the utility evaluation is based on experience, whereas in utility based FP, it is based on a call to a simulated utility function evaluator.

The main advantage of utility-based FP is its very low computational burden on each vehicle. In particular, the vehicles do not need to compute the empirical frequencies of the past proposals made by any vehicle and do not need to compute their expected utilities based on the empirical frequencies. It only requires an individual vehicle to process a (state) vector whose dimension is its number of targets and to select a (randomized) maximizer. This significantly alleviates the computational bottleneck of FP. However, the convergence of utility-based FP for potential games is still an *open issue*.

There are also other utility-based learning algorithms that are proven to converge in partnership games [27–29]. These algorithms are similar to multiagent reinforcement learning algorithms and have comparable computational burden to that of utility based FP. However, convergence requires fine tuning of various parameters, such as the learning rates of each agent. Moreover, utility-based learning algorithms are prone to the issue learnability and may exhibit a slower convergence than action-based FP.

4.1.3 Regret Matching. The discussion on FP in Sec. 4.1.2 motivates a learning algorithm that is computationally feasible as well as convergent in potential games, both theoretically and practically. Accordingly, we introduce regret matching, from [30], whose main distinction is that the vehicles propose targets based on their *regret* for not proposing particular targets in the past negotiation steps.

As before, let us enumerate the targets available to vehicle $\mathcal{V}_i$ as $\mathcal{A}_i=\{A_i^1,\ldots,A_i^{|\mathcal{A}_i|}\}$. Vehicle $\mathcal{V}_i$ selects its proposed target, $a_i(k)$, according to a probability distribution, $p_i(k) \in \Delta(|\mathcal{A}_i|)$, that will be specified shortly. The ℓth component, $p_i^\ell(k)$, of $p_i(k)$ is the probability that vehicle $\mathcal{V}_i$ selects the ℓth target in $\mathcal{A}_i$ at the negotiation step k, i.e., $p_i^\ell(k)=\text{Prob}\{a_i(k)=A_i^\ell\}$. Vehicle $\mathcal{V}_i$ does not know the utility $U_{\mathcal{V}_i}[a(k)]$ before proposing its own target $a_i(k)$. Accordingly, before selecting $a_i(k)$, $k>1$, vehicle $\mathcal{V}_i$ computes its average regret

$$R_{\mathcal{V}_i}^\ell(k) := \frac{1}{k-1}\sum_{m=1}^{k-1} \left\{ U_{\mathcal{V}_i}(A_i^\ell, a_{-i}(m)) - U_{\mathcal{V}_i}(a(m)) \right\}$$

for not proposing A_i^ℓ in all past negotiation steps, assuming that the proposed targets of all other vehicles remain unaltered. Clearly, vehicle $\mathcal{V}_i$ can compute $R_{\mathcal{V}_i}^\ell(k)$ using the recursion

$$R^{\ell}_{\mathcal{V}_i}(k+1) = \frac{k-1}{k} R^{\ell}_{\mathcal{V}_i}(k) + \frac{1}{k}\Big\{U_{\mathcal{V}_i}(A_i^{\ell}, a_{-i}(k)) - U_{\mathcal{V}_i}(a(k))\Big\}, \quad k > 1$$

We note that, at any step $k>1$, vehicle $\mathcal{V}_i$ updates all entries in its average regret vector $R_{\mathcal{V}_i}(k) := [R^1_{\mathcal{V}_i}(k), \ldots, R^{|\mathcal{A}_i|}_{\mathcal{V}_i}(k)]^{\mathrm{T}}$, whose dimension is $|\mathcal{A}_i|$. In particular, the vehicles do not need to compute the empirical frequencies of the past proposals made by any vehicle and do not need to compute their expected utilities based on the empirical frequencies. We also note that it is sufficient for vehicle $\mathcal{V}_i$, at step $k>1$, to have access to $a_i(k-1)$ and $U_{\mathcal{V}_i}(A_i^{\ell}, a_{-i}(k-1))$ for all $\ell \in \{1, \ldots, |\mathcal{A}_i|\}$. In other words, it is sufficient for vehicle $\mathcal{V}_i$'s to have access to its proposal at step $k-1$ and its actual utility $U_{\mathcal{V}_i}(a(k-1))$ received at step $k-1$ as well as its hypothetical utilities $U_{\mathcal{V}_i}(A_i^{\ell}, a_{-i}(k-1))$, which would have been received if it had proposed target A_i^{ℓ} [instead of $a_i(k-1)$] and all other vehicle proposals $a_{-i}(k-1)$ had remained unchanged at step $k-1$.

Once vehicle $\mathcal{V}_i$ computes its average regret vector, $R_{\mathcal{V}_i}(k)$, it proposes a target $a_i(k)$, $k>1$, according to the probability distribution

$$p_i(k) = \frac{[R_{\mathcal{V}_i}(k)]^+}{\mathbf{1}^{\mathrm{T}}[R_{\mathcal{V}_i}(k)]^+}$$

provided that the denominator above is positive; otherwise, $p_i(k)$ is the uniform distribution over $\mathcal{A}_i$ [$p_i(1) \in \Delta(|\mathcal{A}_i|)$ is always arbitrary]. Roughly speaking, a vehicle using regret matching proposes a particular target at any step with probability proportional to the average regret for not playing that particular target in the past negotiation steps. It turns out that the average regret of a vehicle using regret matching would asymptotically vanish (similar results hold for different regret based adaptive dynamics); see [30–32]. Although this result characterizes the long-term behavior of regret matching in general games, it *need not* imply that the negotiations of vehicles using regret matching will converge to a pure equilibrium assignment profile when vehicle utilities constitute a potential game, an objective which we will pursue in Sec. 4.2.

4.2 Generalized Regret Monitoring With Fading Memory and Inertia. To enable convergence to a pure equilibrium in potential games, we will modify regret matching in two ways. First, we will assume that each vehicle has a fading memory; that is, each vehicle exponentially discounts the influence of its past regret in the computation of its average regret vector. More precisely, each vehicle computes a discounted average regret vector according to the recursion

$$\widetilde{R}^{\ell}_{\mathcal{V}_i}(k+1) = (1-\rho)\widetilde{R}^{\ell}_{\mathcal{V}_i}(k) + \rho\{U_{\mathcal{V}_i}(A_i^{\ell}, a_{-i}(k)) - U_{\mathcal{V}_i}(a(k))\}, \quad \text{for all} \quad \ell \in \{1, \ldots, |\mathcal{A}_i|\}$$

where $\rho \in (0,1]$ is a parameter with $1-\rho$ being the discount factor, and $\widetilde{R}^{\ell}_{\mathcal{V}_i}(1) = 0$.

Second, we will assume that each vehicle proposes a target based on its discounted average regret using some inertia. Therefore, each vehicle $\mathcal{V}_i$ proposes a target $a_i(k)$, at step $k>1$, according to the probability distribution

$$\alpha_i(k)\mathrm{RM}_i(\widetilde{R}_{\mathcal{V}_i}(k)) + [1-\alpha_i(k)]\mathbf{v}^{a_i(k-1)}$$

where $\alpha_i(k)$ is a parameter representing vehicle $\mathcal{V}_i$'s willingness to optimize at time k, $\mathbf{v}^{a_i(k-1)}$ is the vertex of $\Delta(|\mathcal{A}_i|)$ corresponding to the target $a_i(k-1)$ proposed by vehicle $\mathcal{V}_i$ at step $k-1$, and $\mathrm{RM}_i : \mathbb{R}^{|\mathcal{A}_i|} \to \Delta(|\mathcal{A}_i|)$ is any continuous function satisfying

$$x^{\ell} > 0 \Leftrightarrow \mathrm{RM}_i^{\ell}(x) > 0 \quad \text{and} \quad \mathbf{1}^{\mathrm{T}}[x]^+ = 0 \Rightarrow \mathrm{RM}_i(x) = \frac{1}{|\mathcal{A}_i|}\mathbf{1} \tag{8}$$

where x^{ℓ} and $\mathrm{RM}_i^{\ell}(x)$ are the ℓth components of x and $\mathrm{RM}_i(x)$, respectively.

We will call the above dynamics generalized regret monitoring (RM) with fading memory and inertia. The reason behind the term "monitoring" is that the algorithm leaves as unspecified how an agent reacts to regrets through the function $\mathrm{RM}_i(\cdot)$. One particular choice for the function RM_i is

$$\mathrm{RM}_i(x) = \frac{[x]^+}{\mathbf{1}^{\mathrm{T}}[x]^+} \quad (\text{when} \quad \mathbf{1}^{\mathrm{T}}[x]^+ > 0)$$

which leads to regret matching with fading memory and inertia. Another particular choice is

$$\mathrm{RM}_i^{\ell}(x) = \frac{e^{\frac{1}{\tau}x^{\ell}}}{\sum_{x^m>0} e^{\frac{1}{\tau}x^m}} I\{x^{\ell} > 0\} \quad (\text{when} \quad \mathbf{1}^{\mathrm{T}}[x]^+ > 0) \tag{9}$$

where $\tau > 0$ is a parameter. Note that, for small values of τ, vehicle $\mathcal{V}_i$ would choose, with high probability, the target corresponding to the maximum regret. This choice leads to a stochastic variant of an algorithm called joint strategy fictitious play (with fading memory and inertia); see [22]. Also, note that, for large values of τ, $\mathcal{V}_i$ would choose any target having positive regret with equal probability.

According to these rules, vehicle $\mathcal{V}_i$ will stay with its previous proposal $a_i(k-1)$ with probability $1-\alpha_i(k)$ regardless of its regret. We make the following standing assumption on the vehicles' willingness to optimize.

ASSUMPTION 4.1. *There exist constants* $\underline{\varepsilon}$ *and* $\bar{\varepsilon}$ *such that*

$$0 < \underline{\varepsilon} < \alpha_i(k) < \bar{\varepsilon} < 1$$

for all time $k>1$ *and for all* $i \in \{1, \ldots, n_v\}$.

This assumption implies that vehicles are always willing to optimize with some nonzero inertia.[2] The following theorem establishes the convergence of generalized regret monitoring with fading memory and inertia to a pure equilibrium.

THEOREM 4.1. *Assume that vehicle utilities constitute an ordinal potential game*[3] *and no vehicle is indifferent between distinct strategies, i.e.,*

$$U_{\mathcal{V}_i}(a_i^1, a_{-i}) \neq U_{\mathcal{V}_i}(a_i^2, a_{-i}), \quad \forall\, a_i^1, \quad a_i^2 \in \mathcal{A}_i, \quad a_i^1 \neq a_i^2, \quad \forall\, a_{-i} \in \mathcal{A}_{-i}, \quad \forall\, i \in \{1, \ldots, n_v\}$$

Then, the target proposals $a(t)$ *generated by generalized regret monitoring with fading memory and inertia satisfying Assumption* 4.1 *converge to a pure Nash equilibrium almost surely.*

Proof. We will state and prove a series of claims. The first claim states that if a vehicle proposes a target with positive (discounted average) regret, then all subsequent target proposals will also have positive regret.

CLAIM 4.1. *Fix any* $k_0 > 1$. *Then,* $\widetilde{R}^{a_i(k_0)}_{\mathcal{V}_i}(k_0) > 0 \Rightarrow \widetilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$ *for all* $k > k_0$.

Proof. Suppose $\widetilde{R}^{a_i(k_0)}_{\mathcal{V}_i}(k_0) > 0$. If $a_i(k_0+1) = a_i(k_0)$, then

$$\widetilde{R}^{a_i(k_0+1)}_{\mathcal{V}_i}(k_0+1) = (1-\rho)\widetilde{R}^{a_i(k_0)}_{\mathcal{V}_i}(k_0) > 0$$

If $a_i(k_0+1) \neq a_i(k_0)$, then

$$\widetilde{R}^{a_i(k_0+1)}_{\mathcal{V}_i}(k_0+1) > 0$$

The argument can be repeated to show that $\widetilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$, for all $k > k_0$. □

Define

[2] This assumption can be relaxed to holding for sufficiently large k, as opposed to all k.

[3] This theorem also holds in the more general weakly acyclic games, see [33].

$$M_u := \max\{U_{\mathcal{V}_i}(a): a \in \mathcal{A}, \mathcal{V}_i \in \mathcal{V}\}$$

$$m_u := \min\{U_{\mathcal{V}_i}(a): a \in \mathcal{A}, \mathcal{V}_i \in \mathcal{V}\}$$

$$\delta := \min\{|U_{\mathcal{V}_i}(a^1) - U_{\mathcal{V}_i}(a^2)|: a^1, a^2 \in \mathcal{A}, a^1_{-i} = a^2_{-i}, |U_{\mathcal{V}_i}(a^1) - U_{\mathcal{V}_i}(a^2)| > 0, \mathcal{V}_i \in \mathcal{V}\}$$

$$N := \min\left\{ n \in \{1,2,\dots\}: (1-(1-\rho)^n)\delta - (1-\rho)^n(M_u - m_u) > \frac{\delta}{2} \right\}$$

$$f := \min\left\{ \mathrm{RM}_i^m(x): |x^\ell| \le M_u - m_u, \ \forall\ \ell, x^m \ge \frac{\delta}{2}, \quad \text{for one } m, \quad \forall\ \mathcal{V}_i \in \mathcal{V} \right\}$$

Note that $\delta, f > 0$, and $|\tilde{R}^{a_i}_{\mathcal{V}_i}(k)| \le M_u - m_u$, for all $\mathcal{V}_i \in \mathcal{V}$, $a_i \in \mathcal{A}_i$, $k > 1$.

The second claim states that if the current proposal is a strict Nash equilibrium and if the proposal is repeated a sufficient number of times, then all subsequent proposals will also be that Nash equilibrium.

CLAIM 4.2. *Fix* $k_0 > 1$. *Assume*

1. $a(k_0)$ *is a strict Nash equilibrium, and*
2. $\tilde{R}^{a_i(k_0)}_{\mathcal{V}_i}(k_0) > 0$ *for all* $\mathcal{V}_i \in \mathcal{V}$, *and*
3. $a(k_0) = a(k_0+1) = \cdots = a(k_0+N-1)$.

Then, $a(k) = a(k_0)$, *for all* $k \ge k_0$.

Proof. For any $\mathcal{V}_i \in \mathcal{V}$ and any $a_i \in \mathcal{A}_i$, we have

$$\tilde{R}^{a_i}_{\mathcal{V}_i}(k_0+N) = (1-\rho)^N \tilde{R}^{a_i}_{\mathcal{V}_i}(k_0) + [1-(1-\rho)^N]\{U_{\mathcal{V}_i}(a_i, a_{-i}(k_0)) - U_{\mathcal{V}_i}(a_i(k_0), a_{-i}(k_0))\}$$

Since $a(k_0)$ is a strict Nash equilibrium, for any $\mathcal{V}_i \in \mathcal{V}$ and any $a_i \in \mathcal{A}_i$, $a_i \ne a_i(k)$, we have

$$U_{\mathcal{V}_i}(a_i, a_{-i}(k_0)) - U_{\mathcal{V}_i}(a_i(k_0), a_{-i}(k_0)) \le -\delta$$

Therefore,

$$\tilde{R}^{a_i}_{\mathcal{V}_i}(k_0+N) \le (1-\rho)^N(M_u - m_u) - [1-(1-\rho)^N]\delta < -\frac{\delta}{2} < 0$$

We also know that, for all $\mathcal{V}_i \in \mathcal{V}$,

$$\tilde{R}^{a_i(k_0)}_{\mathcal{V}_i}(k_0+N) = (1-\rho)^N \tilde{R}^{a_i(k_0)}_{\mathcal{V}_i}(k_0) > 0$$

This proves the claim. □

The third claim states that if the current proposal is not a Nash equilibrium and if the proposal is repeated a sufficient number of times, then the subsequent assignment proposal will have a higher global utility with at least a fixed probability.

CLAIM 4.3. *Fix* $k_0 > 1$. *Assume*

1. $a(k_0)$ *is not a Nash equilibrium, and*
2. $a(k_0) = a(k_0+1) = \cdots = a(k_0+N-1)$

Let $a^* = (a_i^*, a_{-i}(k_0))$ *be such that*

$$U_{\mathcal{V}_i}(a_i^*, a_{-i}(k_0)) > U_{\mathcal{V}_i}(a_i(k_0), a_{-i}(k_0))$$

for some $\mathcal{V}_i \in \mathcal{V}$ *and some* $a_i^* \in \mathcal{A}_i$. *Then,* $\tilde{R}^{a_i^*}_{\mathcal{V}_i}(k_0+N) > \delta/2$, *and* a^* *will be proposed at step* k_0+N *with at least probability* $\gamma := (1-\bar{\epsilon})^{n_v-1}\underline{\epsilon} f$.

Proof. We have

$$\tilde{R}^{a_i^*}_{\mathcal{V}_i}(k_0+N) \ge -(1-\rho)^N(M_u - m_u) + [1-(1-\rho)^N]\delta > \frac{\delta}{2}$$

Therefore, the probability of vehicle $\mathcal{V}_i$ proposing a_i^* at step k_0+N is at least $\underline{\epsilon} f$. Because of players' inertia, the probability that all vehicles will propose action a^* at step k_0+N is at least $(1-\bar{\epsilon})^{n_v-1}\underline{\epsilon} f$. □

The fourth claim specifies an event and associated probability that guarantees that all vehicles will only propose targets with positive regret.

CLAIM 4.4. *Fix* $k_0 > 1$. *We have* $\tilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$ *for all* $k \ge k_0 + 2Nn_v$ *and for all* $\mathcal{V}_i \in \mathcal{V}$ *with probability at least*

$$\prod_{i=1}^{n_v} \frac{1}{|\mathcal{A}_i|} \gamma (1-\bar{\epsilon})^{2Nn_v}$$

Proof. Let $a^0 := a(k_0)$. Suppose $\tilde{R}^{a_i^0}_{\mathcal{V}_i}(k_0) \le 0$. Furthermore, suppose that a^0 is repeated N consecutive times, i.e., $a(k_0) = \cdots = a(k_0+N-1) = a^0$, which occurs with at least probability at least $(1-\bar{\epsilon})^{n_v(N-1)}$.

If there exists an $a^* = (a_i^*, a_{-i}^0)$ such that $U_{\mathcal{V}_i}(a^*) > U_i(a^0)$, then, by Claim 4.3, $\tilde{R}^{a_i^*}_{\mathcal{V}_i}(k_0+N) > \delta/2$ and a^* will be proposed at step k_0+N with at least probability γ. Conditioned on this, we know from Claim 4.1 that $\tilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$ for all $k \ge k_0+N$.

If there does not exist such an action a^*, then $\tilde{R}^{a_i^0}_{\mathcal{V}_i}(k_0+N) < 0$ for all $a_i \in \mathcal{A}_i$. A proposal profile (a_i^w, a_{-i}^0) with $U_{\mathcal{V}_i}(a_i^w, a_{-i}^0) < U_i(a^0)$ will be proposed at step k_0+N with at least probability $(1/|\mathcal{A}_i|) \times (1-\bar{\epsilon})^{n_v-1}$. If $a(k_0+N) = (a_i^w, a_{-i}^0)$, and if, furthermore, (a_i^w, a_{-i}^0) is repeated N consecutive times, i.e., $a(k+N) = \cdots = a(k+2N-1)$, which happens with probability at least $(1-\bar{\epsilon})^{n_v(N-1)}$, then, by Claim 4.3, $\tilde{R}^{a_i^0}_{\mathcal{V}_i}(k_0+2N) > \delta/2$ and the joint target a^0 will be proposed at step (k_0+2N) with at least probability γ. Conditioned on this, we know from Claim 4.1 that $\tilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$ for all $k \ge k_0 + 2N$.

In summary, $\tilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$ for all $k \ge k_0+2N$ with at least probability

$$\frac{1}{|\mathcal{A}_i|} \gamma (1-\bar{\epsilon})^{2Nn_v}$$

We can repeat this argument for each vehicle to show that $\tilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$ for all times $k \ge k_0 + 2Nn_v$ and for all $\mathcal{V}_i \in \mathcal{V}$ with probability, at least

$$\prod_{i=1}^{n_v} \frac{1}{|\mathcal{A}_i|} \gamma (1-\bar{\epsilon})^{2Nn_v}$$

□

Final Step: Establishing Convergence to a Pure Nash Equilibrium. Fix $k_0 > 1$. Let $k_1 := k_0 + 2Nn_v$. Suppose $\tilde{R}^{a_i(k)}_{\mathcal{V}_i}(k) > 0$ for all $k \ge k_1$ and for all $\mathcal{V}_i \in \mathcal{V}$, which, by Claim 4.4, occurs with probability, at least

$$\prod_{i=1}^{n_v} \frac{1}{|\mathcal{A}_i|} \gamma (1-\bar{\epsilon})^{2Nn_v}$$

Suppose further that $a(k_1) = \cdots = a(k_1+N-1)$ which occurs with at least probability $(1-\bar{\epsilon})^{n_v(N-1)}$. If $a(k)$ is a Nash equilibrium, then by Claim 4.1, we are done. Otherwise, according to Claim 4.3, a proposal profile $a' = (a_i', a_{-i}(k_1))$ with $U_{\mathcal{V}_i}(a') > U_{\mathcal{V}_i}(a(k_1))$ for some $\mathcal{V}_i \in \mathcal{V}$ will be played at step k_1+N with at least probability γ. Note that this would imply $U_g(a(k_1+N)) > U_g(a(k_1))$. Suppose now $a(k_1+N) = \cdots = a(k_1+2N-1)$, which occurs with at least

probability $(1-\bar{\epsilon})^{n_v(N-1)}$. If a' is a Nash equilibrium, then, by Claim 4.2, we are done. Otherwise, according to Claim 4.3, a proposal profile $a''=(a_i'',a_{-i}')$ with $U_{\mathcal{V}_i}(a'')>U_{\mathcal{V}_i}(a(k_1+N))$ for some $\mathcal{V}_i\in\mathcal{V}$ will be played at step k_1+2N with at least probability γ. Note that this would imply $U_g(a(k_1+2N))>U_g(a(k_1+N))$. Note that this procedure can only be repeated a finite number of times because the global utility is strictly increasing each time.

We can repeat the above arguments until we reach a pure Nash equilibrium a^* and stay at a^* for N consecutive steps. This means that there exists constants $\tilde{\epsilon}>0$ and $\tilde{T}>0$, both of which are independent of k_0, such that the following event happens with at least probability $\tilde{\epsilon}$: $a(k)=a^*$ for all $k\geq k_0+\tilde{T}$. This proves Theorem 4.1. □

Note that an agreed assignment that emerges from generalized RM with fading memory and inertia can be suboptimal. Characterizing the equilibrium selection properties in potential games still remains as an open problem. As in FP, regret-based dynamics introduced above would require communication of proposed target assignments as part of a negotiation process. FP is guaranteed to converge for potential games but requires an individual vehicle to process the empirical frequencies of all other vehicles that affect its utility and to use these empirical frequencies to compute the maximizer of its expected utility. Generalized RM with fading memory and inertia is guaranteed to converge to a pure equilibrium in almost all (ordinal) potential games; however, its computational requirements are significantly lower. It only requires an individual vehicle to process an average regret vector whose dimension is its number of targets and to select a (randomized) target based on the positive part of its average regret vector.

4.3 Review: One-Step Memory Spatial Adaptive Play. The previous negotiation mechanisms were called recursive averaging algorithms since they maintained a running average (or fading memory average) of certain variables, e.g., averaged actions of other players (FP) or averaged regret measures (RM). These algorithms have "infinite memory" in that the long-term effect of a measured variable may diminish but is never completely eliminated.

In this section, we will consider an opposite extreme, namely, a specific one-step memory algorithm called spatial adaptive play. (SAP) spatial adaptive play was introduced in [9] (Chap. 6) (which also reviews other multistep memory algorithms) as a learning process for games played on graphs. SAP can be a very effective negotiation mechanism in our autonomous vehicle-target assignment problem because it would have low computational burden on each vehicle and it would lead to an optimal solution in potential games with arbitrarily high probability.

Unlike the other negotiation mechanisms we considered thus far, at any step of SAP negotiations, one vehicle is randomly chosen, where each vehicle is equally likely to be chosen, and only this chosen vehicle is given the chance to update its proposed target.[4] Let $a(k-1)$ denote the profile of proposed targets at step $k-1$. At step k, the vehicle that is given the chance to update its proposed target, say vehicle $\mathcal{V}_i$, proposes a target according to a probability distribution $p_i(k)\in\Delta(|A_i|)$ that maximizes

$$p_i^{\mathrm{T}}(k)\begin{bmatrix} U_{\mathcal{V}_i}(A_i^1,a_{-i}(k-1)) \\ \vdots \\ U_{\mathcal{V}_i}(A_i^{|\mathcal{A}_i|},a_{-i}(k-1)) \end{bmatrix}+\tau\mathcal{H}[p_i(k)]$$

where $\mathcal{H}(\cdot)$ is the entropy function that rewards randomization (see Nomenclature) and $\tau>0$ is a parameter that controls the level of randomization. For any $\tau>0$, the maximizing probability $p_i(k)$ is uniquely given by

$$p_i(k)=\sigma\left(\frac{1}{\tau}\begin{bmatrix} U_{\mathcal{V}_i}[A_i^1,a_{-i}(k-1)] \\ \vdots \\ U_{\mathcal{V}_i}[A_i^{|\mathcal{A}_i|},a_{-i}(k-1)] \end{bmatrix}\right)$$

where $\sigma(\cdot)$ is the logit or soft-max function (see Nomenclature). For any $\tau>0$, $p_i(k)$ assigns positive probability to all targets in $\mathcal{A}_i$. We are interested in small values of $\tau>0$ because then $p_i(k)$ approximately maximizes vehicle $\mathcal{V}_i$'s (unperturbed) utility based on other vehicles' proposals at the previous step. For other interpretations of the entropy term, see [35,36]; and for different ways of randomization, see [20].

The computational burden of SAP on each updating vehicle is comparable to that of RM on each vehicle. Each vehicle needs to observe and maintain the proposal profile $a(k)$ (actually, only the relevant part of $a(k)$). If given the chance to update its proposal, vehicle $\mathcal{V}_i$ needs to call its utility function evaluator only $|\mathcal{A}_i|$ times. Because only one vehicle updates its proposal at a given negotiation step, the convergence of negotiations may be slow when there are large number of vehicles.[5] However, if the vehicles have a relatively small number of common targets in their target sets, then multiple vehicles can be allowed to update their proposals at a given step as long as they do not have common targets. Allowing such multiple updates may potentially speed up the negotiations substantially. In our simulations summarized in Sec. 5, typically SAP provided convergence to a near-optimal assignment faster than the most other negotiations mechanisms.

4.4 Selective Spatial Adaptive Play. We will now introduce "selective spatial adaptive play" (sSAP) for the cases where a vehicle has a large number of targets in its target set or calling its utility function evaluator is computationally expensive. We will parameterize sSAP with $n=(n_1,\ldots,n_{n_v})$ where $1\leq n_i\leq|\mathcal{A}_i|-1$ represents the number of times that vehicle $\mathcal{V}_i$ calls its utility function evaluator when it is given the chance to update its proposal. Let us say that vehicle $\mathcal{V}_i$, using sSAP, is given the chance to update its proposal at step k. First, vehicle $\mathcal{V}_i$ sequentially selects n_i targets from $\mathcal{A}_i\backslash\{a_i(k-1)\}$ without replacement where each target is selected independently and with uniform probability over the remaining targets. Call these selected targets $\mathcal{A}_i^{\ell_1}(k),\ldots,\mathcal{A}_i^{\ell_{n_i}}(k)$, and let $\mathcal{A}_i^{\ell_0}(k):=a_i(k-1)$ be appended to these set of selected targets. Then, at step k,vehicle $\mathcal{V}_i$ proposes a target according to the probability distribution

$$p_i(k)=\sigma\left(\frac{1}{\tau}\begin{bmatrix} U_{\mathcal{V}_i}[\mathcal{A}_i^{\ell_0}(k),a_{-i}(k-1)] \\ \vdots \\ U_{\mathcal{V}_i}[\mathcal{A}_i^{\ell_{n_i}}(k),a_{-i}(k-1)] \end{bmatrix}\right)$$

for some $\tau>0$. In other words, at step k, vehicle $\mathcal{V}_i$ proposes a target to approximately maximize its own utility based on the selected targets $\mathcal{A}_i^{\ell_0}(k),\ldots,\mathcal{A}_i^{\ell_{n_i}}(k)$ and other vehicles' proposals at the previous step. Thus, to compute $p_i(k)$, vehicle $\mathcal{V}_i$, needs to call its utility function evaluator only n_i times where $n_i\geq 1$ could be much smaller than $|\mathcal{A}_i|$. It turns out that we can characterize the

[4]We will not deal with the issue of how the autonomous vehicles can randomly choose exactly one vehicle (or multiple vehicles with no common targets) to update its proposal without centralized coordination. In actuality, such asynchronous updating may be easier to implement than implementing the aforementioned negotiation mechanisms that require synchronous updating. One possible implementation of asynchronous updating would be similar to the implementation of well known Aloha protocol in multiaccess communication, where multiple transmitting nodes attempt to access a single communication channel without colliding with each other [34].

[5]If SAP is used as a centralized optimization tool, then the computational burden at each step will be very small because only one entry in $a(k)$ will be updated at each step.

long-term behavior of sSAP quite precisely following along similar lines of proof of Theorem 6.1 in [9].

THEOREM 4.2. *Assume that the vehicle utilities constitute a potential game where the global utility U_g is a potential function. Then, the target proposals $a(k)$ generated by sSAP satisfy*

$$\lim_{\tau\downarrow 0}\lim_{k\to\infty}\mathbf{Prob}\{a(k) \quad \textit{is an optimal target assignment profile}\} = 1$$

Proof. sSAP induces an irreducible Markov process where the state space is $\mathcal{A}$ and the state at step k is the profile $a(k)$ of proposed targets. The empirical frequencies of the visited states converge to the unique stationary distribution of this induced Markov process. As in Theorem 6.1 in [9], we show that, this stationary distribution, denoted by μ^τ, is given as

$$\mu^\tau(a)=\frac{e^{(1/\tau)U_g(a)}}{\sum_{\bar{a}\in\mathcal{A}}e^{(1/\tau)U_g(\bar{a})}},\ \forall\, a\in\mathcal{A}$$

by verifying the detailed balance equations

$$\mu^\tau(a)\mathbf{Prob}\{a\to b\}=\mu^\tau(b)\mathbf{Prob}\{b\to a\},\ \forall\, a,b\in\mathcal{A}$$

The only nontrivial case that requires the verification of the above equations is when a and b differ in exactly one position. Fix a and b such that $a_i\neq b_i$ and $a_{-i}=b_{-i}$. Then, we have

$$\mathbf{Prob}\{a\to b\} = \frac{1}{n_v}\sum_{(a^0,\ldots,a^{n_i})\in S(a,b)}\frac{1}{(|\mathcal{A}_i|-1)\cdots(|\mathcal{A}_i|-n_i)}\frac{e^{(1/\tau)U_{\mathcal{V}_i}(b)}}{\sum_{j=0}^{n_i}e^{(1/\tau)U_{\mathcal{V}_i}(a^j)}}$$

where

$$S(a,b)=\{(a^0,\ldots,a^{n_i})\in\mathcal{A}^{n_i+1}:(a^j_{-i}=a_{-i},\quad\forall\, j)\quad(a^0=a)\quad(a^j=b,\quad\text{for one } j),\quad(a^j\neq a^m,\quad\forall\, j,m)\}$$

It is now straightforward to see that

$$\frac{\mathbf{Prob}\{a\to b\}}{\mathbf{Prob}\{b\to a\}}=e^{(1/\tau)[U_{\mathcal{V}_i}(b)-U_{\mathcal{V}_i}(a)]}=e^{(1/\tau)[U_g(b)-U_g(a)]}=\frac{\mu^\tau(b)}{\mu^\tau(a)}$$

Therefore, μ^τ is indeed as given above, and it can be written, in the alternative vector form, as

$$\mu^\tau=\sigma\left(\frac{1}{\tau}U_g\right)$$

where, by an abuse of notation, U_g is also used to represent a vector whose "ath entry" equals $U_g(a)$. Finally, the fact that the Markov process induced by sSAP with $\tau>0$ being irreducible and aperiodic readily leads to the desired result. □

Thus, in the setup above, μ^τ assigns arbitrarily high probability to those assignment profiles that maximize a potential function for the game as $\tau\downarrow 0$. Clearly, this result indicates that in the case of vehicle utilities IIU (4), RRU (5), or WLU (7), sSAP negotiations would lead to an optimal target assignment with arbitrarily high probability provided that $\tau>0$ is chosen sufficiently small. Of course, one can gradually decrease τ to allow initial exploration. We believe that one can obtain convergence, in probability, of proposals $a(k)$ to an optimal assignment if τ is decreased sufficiently slowly as in simulated annealing [37,38]. In our simulations, choosing τ inversely proportional to k^2 during the negotiations typically resulted in fast convergence of the proposals to a near optimal assignment.

5 Simulation Results

In this section, we present some numerical results to illustrate that when the individual utility functions and the negotiation mechanisms are properly selected the autonomous vehicles can agree on a target assignment profile that yield near-optimal global utility. We consider two scenarios. In the first scenario, we illustrate the near optimality of our approach by simulating a special case of the well-known weapon target assignment model where an optimal assignment can be obtained for large number of weapons and targets in a short period of time [2]. In the second scenario, we simulate a general instance of the problem and compare various negotiation algorithms in terms of their performance and speed of convergence.

Scenario 1. Here, the vehicles are identical and have zero values, whereas the targets are different and have positive values. Each vehicle can be assigned to any of the targets.[6] Let V_j be the value of target $\mathcal{T}_j$ and p_j be the probability that target $\mathcal{T}_j$ gets eliminated when only a single vehicle engages target $\mathcal{T}_j$. When multiple vehicles are assigned to target $\mathcal{T}_j$, each of the vehicles is assumed to engage target $\mathcal{T}_j$, independently. Hence, if the number of vehicles engaging target $\mathcal{T}_j$ is x_j, then $\mathcal{T}_j$ will be eliminated with probability $1-(1-p_j)^{x_j}$. Therefore, as a function of the assignment profile a, the utility generated by the engagement with target $\mathcal{T}_j$ is given by

$$U_{\mathcal{T}_j}(a)=V_j[1-(1-p_j)^{\sum_{i=1}^{n_v}I\{a_i=\mathcal{T}_j\}}]$$

which leads to the following global utility function:

$$U_g(a)=\sum_{j=1}^{n_t}V_j[1-(1-p_j)^{\sum_{i=1}^{n_v}I\{a_i=\mathcal{T}_j\}}]$$

Given the parameters n_v, n_t, $V_1,\ldots,V_{n_t}$, and $p_1,\ldots,p_{n_t}$, an optimal vehicle-target assignment that maximizes the global utility function given above can be quickly obtained using an iterative procedure called minimum marginal return algorithm [2].

To test the effectiveness of our approach, we simulated the vehicle negotiations using the above model with 200 vehicles and 200 targets in MATLAB on a single personal computer with 1.4 GHz Pentium(R) M processor and 1.1 GB of RAM. Each of the target values, $V_1,\ldots,V_{200}$, and each of the elimination probabilities, $p_1,\ldots,p_{200}$, are once independently chosen according to uniform probability distribution on $[0,1]$ and thereafter kept constant throughout the simulations. We first conducted 100 runs of generalized RM negotiations (RM_i function is as in (9), $\rho=0.1$, $\alpha=0.5$) with WLU utilities (7), where each negotiation consisted of 100 steps. We then repeated this with 100 runs of SAP negotiations with WLU utilities (7) where each run consisted of 1000 steps. We also conducted 100 runs of utility based FP negotiations with WLU utilities (7), where each negotiation consisted of 1000 steps. In all cases, the randomization level τ is decreased as $10/k^2$, where k is the negotiation step. Evolution of global utility during typical runs of generalized RM, SAP, and utility-based FP negotiations is shown in Fig. 4. Also, the global utility corresponding to the assignment profile at the end of each run of negotiations and the CPU time required for each run were recorded. A summary of these numerical results is provided in Table 1.

All negotiations consistently yielded near-optimal assignments. Global utility generated by SAP negotiations were almost always monotonically increasing, whereas global utility generated by generalized RM and utility-based FP negotiations exhibited fluctuations as seen in Fig. 4.

In any SAP negotiation step, only one vehicle calls its utility function evaluator 200 times; whereas in any generalized RM negotiation step, all vehicles call their utility function evaluators (200 times for each vehicle). As a result, although a typical generalized RM negotiation converged in 100 steps as opposed to 1000 steps in the case of SAP, a typical 100 step generalized RM negotiation took 593 s CPU time, on average, whereas a typical 1000-step SAP negotiation took 49 s CPU time, on average. However, it is important to note that *these numbers reflect sequential*

[6] Note that there is no reason to consider a null target $\mathcal{T}_0$ here.

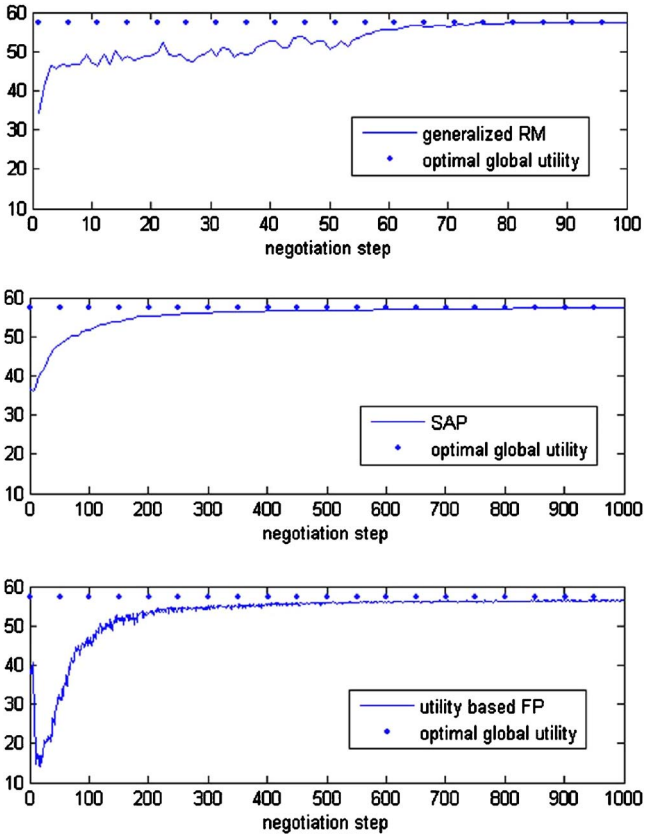

Fig. 4 Evolution of global utility during typical runs of negotiations

CPU time. In an actual implementation, individual vehicles will call their utility function evaluators in parallel. The "parallel" CPU time in Table 1 is the overall CPU time divided by the number of vehicles. It is a rough reflection of what would be the actual implementation time in a parallel implementation. We see that generalized RM is actually *faster* than SAP. The parallel time in SAP is the same as the sequential CPU time because only one vehicle updates its strategy per iteration.

In the case of utility-based FP, all vehicles call their utility function evaluators at each negotiation step but only once for each vehicle. This can be contrasted with generalized RM, which requires a utility function evaluation for every possible target. Utility-based FP took 1000 negotiation steps to approach the optimal global utility, but using 67 s CPU time, on average (or 0.33 s in parallel), which is also faster than the average CPU time used by RM, despite utility-based FP requiring more iterations.

For this scenario, action-based FP would impose enormous computational burden on each vehicle since a vehicle using action FP would have to keep track of the empirical frequencies of the choices of 199 other vehicles and compute its expected utility over a decision space of dimension 200^{200} at every negotiation step. However, the numerical results presented above verify that autonomous vehicles can quickly negotiate and agree on an assignment profile that yields near optimal global utility when vehicle utilities and negotiation mechanisms are chosen properly.

Scenario 2. In this scenario, we consider a more general instance of the weapon target assignment problem, where we have virtually no way of computing the optimal global utility. The setup in this scenario is similar to the one in Scenario 1, except that the vehicles are not identical and are also range restricted. More specifically, each vehicle still has zero value, but the probability p_{ij} that target $\mathcal{T}_j$ gets eliminated when only vehicle $\mathcal{V}_i$ engages target $\mathcal{T}_j$ differs from vehicle to vehicle. Each of the elimination probabilities, p_{ij}, $0 \le i,j \le 200$, is once independently chosen according to uniform probability distribution on $[0,1]$ and thereafter kept constant throughout the simulations. Each vehicle $\mathcal{V}_i$ has 20 targets in its range $\mathcal{A}_i$ and the targets in $\mathcal{A}_i$ are chosen from the set of all targets with equal probability and independently of the other vehicles. Therefore, a pair of two vehicles may have some common as well as distinct targets in their ranges. As in Scenario 1, the target values $V_1, \ldots, V_{200}$ are chosen independently and according to uniform probability distribution on $[0,1]$. Therefore, as a function of the assignment profile a, the utility generated by the engagement with target $\mathcal{T}_j$ is given by

$$U_{\mathcal{T}_j}(a) = V_j\left[1 - \prod_{i:\mathcal{T}_j \in \mathcal{A}_i} (1 - p_{ij})\right]$$

which leads to the following global utility function

$$U_g(a) = \sum_{j=1}^{n_t} V_j\left[1 - \prod_{i:\mathcal{T}_j \in \mathcal{A}_i} (1 - p_{ij})\right]$$

Using the same computational resources and the same setup as in Scenario 1, we simulated the vehicle negotiations on the above model. Evolution of global utility during typical runs of generalized RM, SAP, and utility-based FP negotiations is shown in Fig. 5. The global utility corresponding to the assignment profile at the end of each run of negotiations and the CPU time required for each run were recorded. A summary of these numerical results is provided in Table 2.

All negotiations eventually settled at some assignment profiles, leading to comparable global utility as shown in Fig. 5 and Table 2. The convergence in this scenario was slower for all negotiation mechanisms. The reason for this is that the vehicles in this scenario are not identical and range restricted, and as a result, computing each vehicle's utility is computationally more demanding. The relative timings in both CPU time and convergence rates are similar to those in Scenario 1.

Action-based FP was computationally infeasible for this scenario as well for the same reasons stated earlier, i.e., its enormous computational burden on each vehicle.

The numerical results presented above show that autonomous vehicles can quickly negotiate and agree on an (possibly near-optimal) assignment profile when vehicle utilities and negotiation mechanisms are chosen properly. In all cases, vehicles only communicate with their "neighbors," i.e., those vehicles that share a common target. The difference between algorithms is in the number of vehicles that communicate per iteration. In SAP, only the vehicle revising its assignment must communicate with its neighbors. In generalized RM and utility-based FP, all vehicles must communicate with their neighbors in every iteration. In Scenario 1, all vehicles share the same targets and thus, all vehicles are

Table 1 Summary of simulation runs

	Generalized RM	SAP	Utility-based FP
Average global utility/Optimal global utility	0.99	0.99	0.98
Minimum global utility/Optimal global utility	0.99	0.99	0.96
Average CPU time (s)	593 (≈3.0 parallel)	49	67 (≈0.33 parallel)

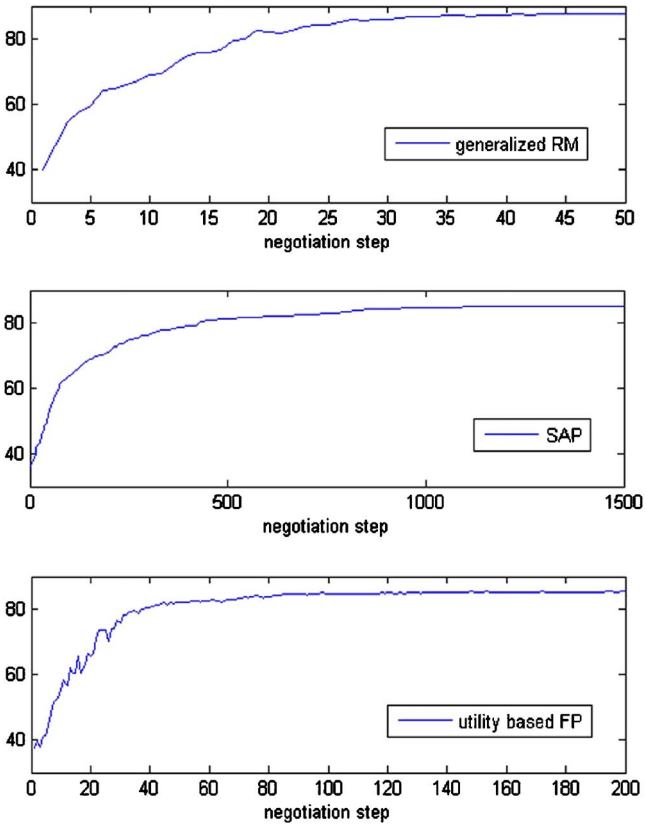

Fig. 5 Evolution of global utility during typical runs of negotiations

neighbors. In Scenario 2, the communication pattern is much more sparse because of the limited vehicle ranges and distribution of targets. The most communications savings per iteration is for SAP. However, SAP showed more iterations required for convergence.

6 Conclusions

We introduced an autonomous vehicle-target assignment problem as a multiplayer game where the vehicles are self-interested players with their own individual utility functions. We emphasized rational decision making on the part of the vehicles to develop autonomous operation capability in uncertain and adversarial environments. To achieve optimality with respect to a global utility function, we discussed various aspects of the design of the vehicle utilities, in particular, alignment with a global utility function and localization. We reviewed selected multiplayer learning algorithms available in the literature. We introduced two new algorithms that address the informational and computation requirement of existing algorithms, namely, generalized RM with fading memory and inertia and selective spatial adaptive play, and provided accompanying convergence proofs. Finally, we discussed these learning algorithms in terms of convergence, equilibrium selection, and computational efficiency, and illustrated the achievement of a global utility in a near-optimal fashion through autonomous vehicle negotiations.

We end by pointing to a significant extension of this work, the case where the vehicle-target assignments need to be made sequentially over a time horizon [2]. In this case, the assignment decisions made by the vehicles at a given time step (probabilistically) determines the future games to be played by the vehicles. Therefore, the vehicles need to take the future utilities into account in their negotiations. A natural framework to study such problems of sequential decision making in a competitive multiplayer setting is the framework of Markov games [39,40]. Extending the approach taken in this paper to a Markov game setup requires significant future work.

Table 2 Summary of simulation runs

	Generalized RM	SAP	Utility-based FP
Global utility	87.62	85.24	85.49
Average CPU time (s)	2707	382	529
	(≈13.5 parallel)		(≈2.64 parallel)

Acknowledgment

Research supported by NSF Grant No. ECS-0501394, AFOSR/MURI Grant No. F49620-01-1-0361, and ARO Grant No. W911NF–04–1–0316.

Nomenclature

$|A|$ = number of elements in A, for a finite set A

$I\{\cdot\}$ = indicator function

$\mathbb{R}^n$ = n dimensional Euclidian space, for a positive integer n

$\mathbf{1}$ = vector $\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} \in \mathbb{R}^n$

$(\cdot)^{\mathrm{T}}$ = transpose operation

$\Delta(n)$ = simplex in $\mathbb{R}^n$, i.e., $\{s \in \mathbb{R}^n \,|\, s \geq 0$ componentwise, and $\mathbf{1}^{\mathrm{T}} s = 1\}$

$\mathrm{Int}(\Delta(n))$ = set of interior points of a simplex, i.e., $s > 0$ componentwise

$\mathcal{H}: \mathrm{Int}(\Delta(n)) \to \mathbb{R}$ = entropy function $\mathcal{H}(x) = -x^{\mathrm{T}} \log(x)$

$\sigma: \mathbb{R}^n \to \Delta(n)$ = "logit" or "soft-max" function $(\sigma(x))_i = e^{x_i}/(e^{x_1} + \cdots + e^{x_n})$

$[x]^+ \in \mathbb{R}^n$ = vector whose ith entry equals $\max(x_i, 0)$, for $x \in \mathbb{R}^n$

References

[1] Olfati-Saber, R., 2006, "Flocking for Multi-Agent Dynamic Systems: Algorithms and Theory," IEEE Trans. Autom. Control, **51**, pp. 401–420.

[2] Murphey, R. A., 1999, "Target-Based Weapon Target Assignment Problems," *Nonlinear Assignment Problems: Algorithms and Applications*, Pardalos, P. M., and Pitsoulis, L. S., ed., pp. 39–53, Kluwer, Dordrecht.

[3] Ahuja, R. K., Kumar, A., Jha, K., and Orlin, J. B., 2003, "Exact and Heuristic Methods for the Weapon-Target Assignment Problem," http://ssrn.com/abstract=489802

[4] Fudenberg, D., and Tirole, J., 1991, *Game Theory*, MIT Press, Cambridge, MA.

[5] Basar, T., and Olsder, G. J., 1999, *Dynamic Noncooperative Game Theory*, SIAM, Philadelphia.

[6] Wolpert, D. H., and Tumor, K., 2001, "Optimal Payoff Functions for Members of Collectives," Adv. Complex Syst., **4**(2&3), pp. 265–279.

[7] Monderer, D., and Shapley, L. S., 1996, "Potential Games," Games Econ. Behav., **14**, pp. 124–143.

[8] Fudenberg, D., and Levine, D. K., 1998, *The Theory of Learning in Games*, MIT Press, Cambridge, MA.

[9] Young, H. P., 1998, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton University Press, Princeton, NJ.

[10] Wolpert, D., and Tumor, K., 2004, "A Survey of Collectives," in *Collectives and the Design of Complex Systems*, K. Tumer and D. Wolpert, eds., Springer-Verlag, New York, NY, p. 142.

[11] Miettinen, K. M., 1998, *Nonlinear Multiobjective Optimization*, Kluwer, Dordrecht.

[12] Rosenthal, R. W., 1973, "A Class of Games Possessing Pure-Strategy Nash Equilibria," Int. J. Game Theory, **2**, pp. 65–67.

[13] Mas-Colell, A., Whinston, M. D., and Green, J. R., 1995, *Microeconomic Theory*, Oxford University Press, London.

[14] Benaim, M., and Hirsch, M. W., 1999, "Mixed Equilibria and Dynamical Systems Arising From Fictitious Play in Perturbed Games," Games Econ. Be-

hav., **29**, pp. 36–72.
[15] Brown, G. W., 1951, "Iterative Solutions of Games by Fictitious Play," *Activity Analysis of Production and Allocation*, Koopmans, T. C., ed., Wiley, New York, pp. 374–376.
[16] Monderer, D., and Shapley, L. S., 1996, "Fictitious Play Property for Games With Identical Interests," J. Econ. Theory, **68**, pp. 258–265.
[17] Curtis, J. W., and Murphey, R., 2003, "Simultaneous Area Search and Task Assignment for a Team of Cooperative Agents," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, August, Austin, Texas, AIAA, pp. 2003–5584.
[18] Hofbauer, J., 1995, "Stability for the Best Response Dynamics," University of Vienna, Vienna, Austria, http://homepage.univie.ac.at/josef.hofbauer/br.ps
[19] Krishna, V., and Sjöström, T., 1998, "On the Convergence of Fictitious Play," Math. Op. Res., **23**, pp. 479–511.
[20] Hofbauer, J., and Sandholm, B., 2002, "On the Global Convergence of Stochastic Fictitious Play," Econometrica, **70**, pp. 2265–2294.
[21] Lambert, T. J., III, Epelman, M. A., and Smith, R. L., 2005, "A Fictitious Play Approach to Large-Scale Optimization," Oper. Res., **53**(3), pp. 477–489.
[22] Marden, J. R., Arslan, G., and Shamma, J. S., 2005, "Joint Strategy Fictitious Play with Inertia for Potential Games," *Proc. of 44th IEEE Conference on Decision and Control*, Dec., pp. 6692–6697.
[23] Fudenberg, D., and Levine, D., 1998, "Learning in Games," *European Economic Review*, 42, pp. 631–639.
[24] Fudenberg, D., and Levine, D. K., 1995, "Consistency and Cautious Fictitious Play," J. Econ. Dyn. Control, **19**, pp. 1065–1089.
[25] Sutton, R. S., and Barto, A. G., 1998, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA.
[26] Bertsekas, D. P., and Tsitsiklis, J. N., 1996, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA.
[27] Leslie, D., and Collins, E., 2003, "Convergent Multiple-Ttimescales Reinforcement Learning Algorithms in Normal form Games," Ann. Appl. Probab., **13**, pp. 1231–1251.
[28] Leslie, D., and Collins, E., 2005, "Individual *Q*-Learning in Normal Form Games," SIAM J. Control Optim., **44**(2). pp. 495–514.
[29] Leslie, D. S., and Collins, E. J., 2006, "Generalised Weakened Fictitious Play," *Games and Economic Behavior*, Vol. 56, issue 2, pages 285–298.
[30] Hart, S., and Mas-Colell, A., 2000, "A Simple Adaptive Procedure Leading to Correlated Equilibrium," Econometrica, **68**(5), pp. 1127–1150.
[31] Hart, S., and Mas-Colell, A., 2001, "A General Class of Adaptative Strategies," J. Econ. Theory, **98**, pp. 26–54.
[32] Hart, S., and Mas-Colell, A., 2003, "Regret Based Continuous-Time Dynamics," Games Econ. Behav., **45**, pp. 375–394.
[33] Marden, J. R., Arslan, G., and Shamma, J. S., 2007, "Regret Based Dynamics: Convergence in Weakly Acyclic Games," *Proc. of 6th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, ACM Press, New York, NY, pp. 194–201.
[34] Bertsekas, D., and Gallager, R., 1992, *Data Networks*, 2nd ed., Prentice-Hall, Englewood Cliffs., NJ.
[35] Hofbauer, J., and Hopkins, E., 2005, "Learning in Perturbed Asymmetric Games," *Games and Economic Behavior*, Vol. 52, pp. 133–152.
[36] Wolpert, D. H., 2004, "Information Theory—The Bridge Connecting Bounded Rational Game Theory and Statistical Physics," http://arxiv.org/PS-cache/cond-mat/pdf/0402/0402508.pdf
[37] Aarts, E., and Korst, J., 1989, *Simulated Annealing and Boltzman Machines*, Wiley, New York.
[38] van Laarhoven, P. J. M., and Aarts, E. H. L., 1987, *Simulated Annealing: Theory and Applications*, Reidel, Dordrecht.
[39] Raghavan, T. E. S., and Fillar, J. A., 1991, "Algorithms for Stochastic Games—A Survey," Methods Models Op. Res., **35**, pp. 437–472.
[40] Vrieze, O. J., and Tijs, S. H., 1980, "Fictitious Play Applied to Sequence of Games and Discounted Stochastic Games," Int. J. Game Theory, **11**, pp. 71–85.