Overcoming Communication Restrictions in Collectives

Kagan Tumer
Computational Sciences Division
NASA Ames Research Center
Mailstop 269-4
Moffett Field, CA 94035
E-mail: ktumer@mail.arc.nasa.gov

Adrian K. Agogino
University California, Santa Cruz
NASA Ames Research Center
Mailstop 269-4
Moffett Field, CA 94035
E-mail: adrian@email.arc.nasa.gov

Abstract—Many large distributed system are characterized by having a large number of components (e.g., agents, neurons) whose actions and interactions determine a “world utility” which rates the performance of the overall system. Such “collectives” are often subject to communication restrictions, making it difficult for components which try to optimize their own “private” utilities, to take actions that also help optimize the world utility. In this article we address that coordination problem and derive four utility functions which present different compromises between how “aligned” a component’s private utility is with the world utility and how readily that component can determine the actions that optimize its utility. The results show that the utility functions specifically derived to operate under communication restrictions outperform both traditional methods and previous collective-based methods by up to 75%.

I. INTRODUCTION

Control and coordination in a large distributed system designed that needs to achieve a collective task is a challenging area of research. Many methods exist for coordinating the actions of such a system when the components (e.g., agents, neurons) can fully communicate with one another [6], [15], [21]. In this work we focus on solution to this coordination problem based on “collectives” [17], [21]. A collective is a large distributed system of interacting agents where there is a well-defined “world utility” function rating the possible dynamic histories of the full system, and where each agent is only concerned with maximizing its own “private utility” function [21]. However, in many problems, the presence of communication restrictions significantly complicates the coordination problem[4], [8], [13]. Examples of such problems include controlling collections of rovers or constellations of satellites, and coordinating data routing across a network (because of such examples, we will refer to the components as “agents” in this article). In each of those cases, an agent may only be able to directly communicate with a small number of other agents. In addition, even if there are indirect methods for sharing information (e.g., team formation), they may be costly and a agent may be unwilling to share, if doing so would hurt its private utility (the use of teams to overcome communication restrictions in multi-agent systems is discussed in [1]). In all of these problems, the system designer faces the difficult task of providing the agents with a private utility that:

1) allows agents to work towards the common goal and not against one another, i.e., the agents’ private utility functions are aligned with the “world utility function”; and

2) does not require access to global information available through a broad communication network, i.e., agents can determine which actions are beneficial to their private utilities with the limited information at their disposal.

These issues are at odds with each other and in fact in many cases it will be impossible for the agents to achieve high values of a private utilities which is “aligned” with the world utility.1 In addition even if the world utility, computed with global information, can be broadcast to all the agents, agents may not be able to effectively use this information to select actions that will be useful to them and to the overall system. In fact many obvious methods of combining local information with the world utility can actually cause reduced performance as communication increases (Figure 1). This example shows the behavior of a system (described in detail in Section IV) where the world utility is plotted with respect to the percentage of agents with which an agent can communicate. Note that in some states of the system (e.g., low communication levels), increasing the amount of information to which agents have access has deleterious effects on the performance of the system. We will discuss the reasons for this paradox and show how some problems stemming from communication restrictions can be overcome by providing agents with carefully crafted private utility functions.

The first step in creating a distributed system that can effectively maximize world utility is to ensure that the agents work together. If the agents are not designed to work well with each other, they may not learn their task properly, may interfere with each other’s ability to contribute to the world utility, or simply perform useless repetitive work. Hand tailoring the agents’ private utility functions may offer a solution, but generally, such systems: (i) have to be laboriously modeled; (ii) provide “brittle” global performance; (iii) are not

1By “aligned” we mean that actions that improve the private utility of an agent will also improve the world utility. We will formalize this concept in Section II.
"adaptive" to changing environments; and (iv) generally do not scale well.

To sidestep these problems, yet address the design requirements listed above (i.e., utility "alignedness" and "learnability") one can use the framework of collectives. Given this framework, the crucial design problem becomes: Assuming the individual agents are able to maximize their own utility functions (e.g., through reinforcement learning or evolving neural networks), what set of private utilities for the individual agents will, when pursued by those agents, result in high world utility? The collectives framework has been successfully applied to multiple domains including packet routing over a data network, congestion games, multiple-resource job scheduling over a heterogeneous computational grid, and the coordination of multi-rovers in learning sequences of actions.

In this article, we extend the question of how to design the agents' private utilities given that centralized communication is not possible. Though this question has not been directly addressed, there is a large body of work on systems with low levels of communication. Issues such as agent communication languages and physical implementation of communication have received particular attention. At a higher level, Pynadath and Tambe have formalized many aspects of agent communication. For multi-agent Markov decision processes, Xule et al. dealt with the problem of partially hidden states of communications have received particular attention.

In this article, we show how communication restrictions in a system can be overcome by modifying the agents' utilities. Based on the work on collectives, we derive four different agent utility functions that offer different levels of alignedness and learnability for the private utility functions. Furthermore, those utilities differ in whether they allow for global broadcasts of the world utility (in some domains, though the agents will not be able to engage in realtime agent to agent communication, some global information can be broadcast at various intervals). In Section II, we summarize the theory of collectives that is needed for this article. In Section III, we describe the problem domain and derive the collective-based solution to this problem. In Section IV, we present and discuss the simulation results.

II. BACKGROUND: COLLECTIVES

In this section, we summarize the theory collectives necessary to derive the agent utility functions used in this article. Let Z be an arbitrary vector space whose elements z give the joint move of all agents in the system (i.e., z specifies the full state of the system). The world utility G(z), is a function of the full state z, and the problem we face is to find the z that maximizes G(z). In addition to G, each agent has a private utility function g_. The agents' goals are to optimize their individual private functions, even though, we, as system designers are only concerned with the value of the world utility G. We will denote the state of agent _ by z_ and the state of all other than _ by z_. In this work we take z, z_, and z_ to have the same dimensionality (e.g., for z_ all elements of z that are not dependent of _ are replaced with zeros), resulting in the notation: z = z + z_.

A. Factoredness and Learnability

For high values of G to be achieved, the private utility functions need to have two properties, which we will call factoredness and learnability. First we want the private utility functions of each agent to be aligned with respect to G, intuitively meaning that an action taken by an agent that improves its private utility also improves the world utility. Specifically, for any two states z and z' which differ only on agent _'s state, an action by agent _ that increases _ will also increase G. Formally a utility _ is factored with G when:

\[ g_\eta(z) > g_\eta(z') \iff G(z) > G(z') \]

\[ \forall z, z'. \text{ s.t. } z_ = z_ \]

In game theory language, the Nash equilibria of a factored system are local maxima of G. In addition to this desirable equilibrium behavior, factored systems also automatically provide appropriate off-equilibrium incentives to the agents (an issue rarely considered in the game theory / mechanism design literature).

Second, we want the agents' private utility functions to have high learnability, intuitively meaning that an agent's utility should be sensitive to its own actions and insensitive to actions of others. As a trivial example, any system in which all the private utility functions equal G is factored. However such
systems often suffer from low signal-to-noise, a problem that get progressively worse as the size of the system grows. This is because for large systems where $G$ sensitively depends on all components of the system, each agent may experience difficulty discerning the effects of its actions on $G$. As a consequence, each $\eta$ may have difficulty achieving high $g_\eta$. This signal-to-noise effect, called learnability is the second property that is crucial in the design of the agents’ private utility functions. Formally we can quantify the learnability of a utility $g_\eta$ by:

$$\lambda_{\eta,g_\eta}(z) \equiv \frac{\|\nabla_z g_\eta(z)\|}{\|\nabla_z g_\eta(z)\|}.$$  

So at a given state $z$, the higher the learnability, the more $g_\eta(z)$ depends on the move of agent $\eta$, i.e., the better the associated signal-to-noise ratio for $\eta$. Intuitively then, higher learnability means it is easier for $\eta$ to achieve a large values of its utility.

### B. Difference Utilities

Consider difference utilities, which are of the form:

$$DU(\eta) \equiv G(z) - G(z - z_\eta + v_\eta)$$

where $v_\eta$ is a constant vector. In the second term of $DU_\eta$ all states depending on $\eta$ are replaced by a constant, creating a virtual state. Difference utilities are factored no matter the choice of $v_\eta$ precisely because the second term does not depend on $\eta$’s state [21]. Furthermore, they usually have far better learnability than does setting $g_\eta$ to $G$ because the second term of $DU$ removes a lot of the effect of other agents (i.e., noise) from $\eta$’s utility. In this work we set $v_\eta$ to the “null” vector, (e.g., $v_\eta = \emptyset$). Note, that when $z_\eta$ is set to the null state, $DU$ is closely related to the economics technique of “endogenizing a player’s (agent’s) externalities” [12]. Indeed, $DU$ has conceptual similarities to Vickrey tolls [19], and Groves’ mechanism [10], though the Groves mechanism results in a team game.

Intuitively, one can look at $DU$ from the perspective of a human company, with $G$, the “bottom line” of the company, the agents $\eta$, the employees of that company, and the associated $g_\eta$, the employees’ performance-based compensation packages. For a “factored company”, each employee’s compensation package contains incentives designed such that the better the bottom line of the company, the greater the employee’s compensation. For example the board of a company wishing to have the private utilities of the employees be factored with $G$ may give stock options to the employees. The net effect of this action is to ensure that what is good for the employee is good for the company. In addition, if the compensation packages have “high learnability”, the employees will have a relatively easy time discerning the relationship between their behavior and their compensation. In such a case the employees will both have the incentive to help the company and be able to determine how best to do so. Note that in practice, providing stock options is generally more effective in small companies than in large ones. This makes perfect sense in terms of the formalism, since such options generally have higher learnability in small companies than they do in large companies, in which each employee has a hard time seeing how his/her moves affect the company’s stock price.

### C. Communication Restrictions

In many real world problems the computation of the difference utility requires sufficient communication among the agents to allow the agents to infer the value of the state of the entire system. In some specific domains, using difference utilities results in many elements of the system state to cancel out, allowing the agents to compute $DU$ without knowing the full state. However in general, an agent may not have sufficient communication to compute $DU$, and needs to approximate under the constraints of communication restrictions.

Mathematically we represent the communication restrictions for an agent $\eta$ as elements of the system state that are not observable. We can decompose the state $z$ into a component observable by agent $\eta$, $z^{o_\eta}$, and a component hidden from agent $\eta$, $z^{h_\eta}$ (note $z = z^{o_\eta} + z^{h_\eta}$). In this paper we will define the communication level for agent $\eta$ as:

$$B_\eta = \int_{z^{o_\eta}} dz' \int_z dz.$$  

For a problem with countable state elements, $B_\eta$ reduces to the the number of observable elements in the state divided by the total number of elements in the state. Note that $B$ is always in the range $[0,0.1]$. If the $DU$ for agent $\eta$ depends on any component of $z^{h_\eta}$ then $\eta$ cannot compute it directly. Instead we introduce different approximations to the $DU$ that vary in their balance between learnability and factoredness. In the four utilities discussed below, the first two letters of the utility represent how the two terms of the difference utility get their information. “B” stands for “broadcast” meaning that the world utility is broadcast to the system. “T” stands for “truncated” meaning that the hidden values are ignored, and “E” stands for “estimated” meaning that the hidden variable is estimated from the observed variables. Table I shows the factoredness, learnability and communication level trade-offs for $DU$ and each of the four utilities presented below (e.g., $BEU$ is fully factored, has low learnability and uses local communications as well as global broadcasts, whereas $EEU$ is partially factored, has high learnability and only uses local communications).

<table>
<thead>
<tr>
<th>Utility</th>
<th>Factoredness</th>
<th>Learnability</th>
<th>Required Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DU$</td>
<td>Full</td>
<td>High</td>
<td>Global</td>
</tr>
<tr>
<td>$BTU$</td>
<td>Full</td>
<td>Low</td>
<td>Broadcast/Local</td>
</tr>
<tr>
<td>$TTU$</td>
<td>Partial (low)</td>
<td>High</td>
<td>Local</td>
</tr>
<tr>
<td>$BEU$</td>
<td>Full</td>
<td>Low</td>
<td>Broadcast/Local</td>
</tr>
<tr>
<td>$EEU$</td>
<td>Partial (high)</td>
<td>High</td>
<td>Local</td>
</tr>
</tbody>
</table>

#### 1) Broadcast/Truncated Utility (BTU): BTU is a variant of $DU$, where the communication restrictions force agent $\eta$ to
not only set its own state, but also the states of all agents that it cannot observe to the null state:

\[
BTU_\eta(z) = G(z) - G(z - z^{h_\eta} - z_\eta) \tag{4}
\]

Note that \(BTU\), as well as \(BEU\) (discussed below), assume that the true world utility can be broadcast despite the communication restriction. In many applications, this is a reasonable assumption since the world utility can often be computed once and broadcast throughout the environment \[9\]. More complex forms of broadcasting are often used for distributed multi-agent systems \[5\], but in this paper we will assume a very simple global broadcast of a single number.

Despite creating a virtual state by setting more than \(\eta\) to the null state, \(BTU\) is still factored since it is in the form of the difference utility (e.g., the second term of Equation 4 does not depend on \(\eta\)). However, this utility generally has significantly more noise than a pure \(DU\) since the difference removes not only \(\eta\)'s contribution, but all states hidden from \(\eta\). Accordingly, in situations where a large number of agents are hidden from \(\eta\), \(BTU\) suffers from poor signal to noise problems, e.g., at the limit of agent \(\eta\) observing only its own actions, the second term becomes \(G(\vec{0})\).

2) Truncated/Truncated Utility (TTU): The second private utility is conceptually similar to \(BTU\) except that both terms are computed under the communication restrictions:

\[
TTU_\eta(z) = G(z - z^{h_\eta}) - G(z - z^{h_\eta} - z_\eta). \tag{5}
\]

Essentially, \(TTU\) is \(DU\) where \(z\) is approximated by \(z - z^{h_\eta}\). Because of this, \(TTU\) is not factored with respect to the world utility \(G(z)\). While not being factored with world utility, \(TTU\) generally has higher learnability than \(BTU\) \[20\].

Again, consider the case where a large number of agents, not interacting with \(\eta\), are hidden from \(\eta\). The contribution of those agents will not be included in either term of \(TTU\), since both terms are computed with the communication restriction. Therefore this utility will have less noise. However, if the assumption that \(G(z - z^{h_\eta})\) is close to \(G(z)\) does not hold (e.g., some hidden agents are crucial to the system's behavior) then \(TTU\) will not produce good system performance.

3) Broadcast/Estimated Utility (BEU): The third utility is similar to \(TTU\), except that instead of truncating the components of \(z^{h_\eta}\) (e.g., setting them to zero), their values are estimated given the values of \(z^{o_\eta}\):

\[
BEU_\eta(z) = G(z) - G(z^{o_\eta} + E[z^{h_\eta} | z^{o_\eta}] - z_\eta) \tag{6}
\]

where \(E[z^{h_\eta} | z^{o_\eta}]\) gives the expected hidden state given the states observable to \(\eta\). As long as this estimate is not influenced by the actions of \(\eta\) beyond \(z_\eta\), this utility is factored, since the first term of the difference equation is still \(G(\eta)\).

While both \(BTU\) and \(BEU\) are factored, \(BEU\) may have less noise, depending on how good the estimate for \(z^{h_\eta}\) is.

Again, consider a system where a large number of agents that do not interact with \(\eta\) that are hidden from \(\eta\)'s state, but that their values can be approximated from the visible components of the state. In this case the first term of \(BEU\) will contain the agents' contribution to \(G(z)\), but the second term will subtract out their inferred contribution. Even if effects of the hidden elements cannot be perfectly estimated, significant amounts of noise can be eliminated from the system. Note however that if the estimate is particularly poor, noise can also be introduced into the system.

4) Estimated/Estimated Utility (EEU): The fourth utility is similar to \(TTU\), except that instead of truncating the hidden elements, the value of \(z^{h_\eta}\) is estimated in both terms:

\[
EEU_\eta(z) = G(z^{o_\eta} + E[z^{h_\eta} | z^{o_\eta}]) - G(z^{o_\eta} + E[z^{h_\eta} | z^{o_\eta}] - z_\eta). \tag{7}
\]

Essentially, \(EEU\) is a \(DU\) where \(z\) is approximated by \(z^{0_\eta} - E[z^{h_\eta} | z^{o_\eta}]\). As was the case with \(TTU\) \(EEU\) is not factored with respect to the world utility \(G\). However, with a good estimate of \(z^{h_\eta}\), the value \(G(z^{o_\eta} - E[z^{h_\eta} | z^{o_\eta}])\) will be much closer to \(G(z)\) than \(G(z^{o_\eta})\), so this utility can be much closer to being factored with respect to \(G(z)\) than can \(TTU\).

In addition this utility retains \(TTU\)'s desirable property that both terms are using the same version of the state. Since both terms are estimating the values of \(z^{h_\eta}\) in the same way, any contribution that the non-\(\eta\) terms of \(z^{h_\eta}\) make on the first term will be subtracted out in the second term. Note that unlike with \(BEU\), even if the estimate of the hidden components is very poor, noise will not be added to the system since both terms of the utility use the same estimate. Instead, the quality of the estimate only affects how close this utility is to being factored with respect to \(G(z)\).

III. CONGESTION GAMES

Congestion games are characterized by having the world utility depend on the agents use of a particular resource (e.g., quality of an agent's action depends on the number of other agents selecting the same action) \[2, 11\]. This type of problem arises in many domains, ranging from telecommunications (e.g., response of a link depends on the number of users), transportation (e.g., value of a highway lane depends on the number of cars), power/computer grids (e.g., performance of a server depends on the number of scheduled jobs), and public good distribution (e.g., enjoyment of a park/restaurant depends on the number of people using it). In each instance of the problem, at each time step, each agent \(\eta\) has to decide whether to participate (e.g., use server, drive on a lane, attend restaurant) in the use of that resource or not. The nature of the problem produces a "congestion" (e.g., if most agents believe the resource will be under-used, they will use it and cause it to be over-used, and vice-versa).

In this work, we focus on the following instantiation of the congestion game: There are \(N\) agents, each picking one out of \(K\) actions each time step. Those actions result in a world utility, \(G\), given by:

\[
G(z) \equiv \sum_{k=1}^{K} x_k(z) e^{-z_k(z) / c_k}, \tag{8}
\]

where \(x_k(z)\) is the number of agents choosing action \(k\); \(z_\eta\) is \(\eta\)'s choice at that time step; and \(c_k\) is the optimal "capacity" of
resource $k$. At the end of the time step, the associated private utilities for each agent are communicated to that agent, and the process is repeated.

Since we wish to concentrate on the effects of the utilities rather than on the algorithms that use them, we use a very simple learning algorithm, though a number of learning methods (e.g., neural networks, Q-learning) can be used. In this simple algorithm each agent $\eta$ keeps a $K$-dimensional vector giving its estimates of the utility it would receive for choosing that action. The decisions are made using the vector, with an $\epsilon$-greedy learner with $\epsilon$ set to 0.05. All of the vectors are initially set to zero and there is a learning rate decay is 0.99.

A. Communication Restrictions

We model communication restrictions in this problem by controlling how many other agents one agent can "talk" to. Without this communication the agent cannot know what the other agents have done. We define a communication level $B$ in the range $[0.0, 1.0]$ representing the fraction of all the agents to which an agent can talk. When $B = 1.0$ an agent can talk to the all other agents, whereas when $B = 0.0$ an agent has no communication, and thus is only aware of its own action. In this problem, communication restrictions result in variations on how $x_k(z)$ is computed. For truncated versions of the DU, (BTU and TTU), $\eta$ uses $x_k(z^n)$ which provides the number of observable agents that have selected action $k$. (note since in BTU the first term is broadcast, the agent does not need to compute it). For utilities using an estimate of the state (BEU and EEU), $x_k(z^n)$ is scaled, and $\frac{1}{B} x_k(z^n)$ represents agent $\eta$'s estimate of how many agents selected action $k$. Note this is an extremely simple estimation procedure and does not take any information an agent collects to modify how it forms this estimate.

IV. EXPERIMENTAL RESULTS

We tested the performance of the four versions of the DU with varying levels of communication. The test were conducted in a congestion game with 100 agents and with $c_k = 5$ for all $k$. All of the trials were conducted for 1000 episodes, and were run 25 times.

Figure 2 shows the performance of the four utilities with different levels of communication. When the communication level is high, the utilities converge to DU. When communication is very low, the BTU and BEU have the best performance because their first term, $G$, is not affected by the communication restriction. They essentially are reduced to a team game, and give moderately good performance. Note that the performance of BTU is worse at 50% communication than at 5%. This counterintuitive result is explained by how the utility is computed in this problem. With little communication, the total number of agents that can be seen is small, and the contribution of the second term is small. With 50% communication on the other hand, the second term will be large enough to have an impact on the utility. However, because both at 5% and 50% communication levels $x_k(z^n)$ is significantly different than $x_k(z)$, neither provide a usable second term. In fact, rather than subtracting out noise, the second term adds noise.

![Figure 2](image1.png)

Fig. 2. Performance of four utility functions for a range of communication levels. For moderate communication levels EEU performs best. For very low communication BTU performs best since, it uses information from world utility.

![Figure 3](image2.png)

Fig. 3. Learning rates of four utility functions at 40% communication. EEU learns far more quickly than the others, because it provides a cleaner signal. Note that even though TTU is highly learnable, it is not close to being factored with respect to $G$, so it has a flat learning curve. Both BEU and BTU learn because they are factored, but because they have low learnability (too much noise in the signal) their learning curve is extremely slow.

For most levels of communication restriction, the EEU performs the best and performs up to 75% closer to optimal than utilities which use the same information. Recall that EEU and TTU are not factored, whereas BTU and BEU are. What helps EEU in this case is that though it is not factored, as long as the estimate for $G$ in the first term is sufficiently close to $G$, it is close to being factored. Furthermore, because both the first and second terms use the same estimate for the state, the subtraction does remove noise, as intended. The utility TTU performs the worst even though there may not be much noise in the utility. This is caused by TTU being far from factored due to the truncation of the hidden state components.

Figures 3 and 4 give a clearer view of the performances at a fixed level of communication restriction (40% and 70% respectively). EEU is clearly superior at 40% communication. It is close to being factored and because of it's high learnability it rapidly converges to a good solution. Both BTU and BEU...
provide the best solutions in helping overcome communication restrictions in a collective [1]. Future work in this area includes investigating new utility functions for the agents, dynamic team formation where agents may join and/or leave teams in an adaptive fashion, and incurring a cost for sharing information. Furthermore, we are determining the effectiveness of using the utilities as fitness evaluation functions for evolutionary computation with neural networks.

REFERENCES