Expertise and Trust-based formation of effective coalitions: an evaluation of the ART testbed

Sandip Sen
Mathematical & Computer Sciences Department
University of Tulsa, USA
sandip@utulsa.edu

Indranil Goswami
Mathematical & Computer Sciences Department
University of Tulsa, USA
inder@utulsa.edu

Stéphane Airiau
Mathematical & Computer Sciences Department
University of Tulsa, USA
stephane@utulsa.edu

ABSTRACT

We are interested in designing strategies by which a service provider can effectively compete to maximize its clientele when the distribution of a fixed client base over the providers is proportionate to provider performance. We assume that service providers can receive and provide assistance to other service providers to effectively serve their customers and thereby improve performance. We identify situations where effective cooperation can benefit both parties in the long run. We prescribe trust models that can be used to leverage such mutually beneficial relationships to form stable coalitions. Additionally, we derive conditions under which existing agent coalitions should welcome new agents. We study the recently proposed Agent Reputation and Trust (ART) Testbed as a model for situating our discussions. These provides a dual benefit: (i) we can discuss our strategies in more concrete terms within the specific context of the ART testbed, and (ii) we identify features of the ART testbed which are not conducive to fostering trust and suggest modifications to address those limitations.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms

Algorithms

Keywords

ART, trust, reputation, coalition formation

1. INTRODUCTION

We believe that future online environments will situate intelligent agents in dynamic, open environments posing both opportunities for collaboration with cooperative agents and dangers from malicious or exploitative agents. Mirroring interactions in human societies, such agent communities will need to utilize a variety of rich interaction models and decision mechanisms that handle a gamut of social scenarios. For example, while economic mechanisms like auctions would be appropriate for one-time interactions between self-interested parties, a negotiating agent should prefer argumentation-based negotiation when it can influence the decisions of its opponents by providing justification or explanation for its position. Agents interacting for extended periods can jointly explore alternate behaviors and learn coordinated policies, etc. We are particularly interested in studying mechanisms that can be used by agents in stable communities to leverage complementary capabilities. Such capabilities may be in the form of resources, knowledge, or expertise. If we look at corresponding equivalents of stable communities in human societies, e.g., neighbors, friend circles, members of a local, special interest club, etc. we find that the significant majority of interactions and exchanges are based on mechanisms which cannot be modeled by standard currency-based economic mechanisms. Rather, trust and reputation based exchange seems to be the dominant decision-making paradigm that sustains effective collaborative relationships in stable human groups. A typical trust-based interaction involves at least one party taking a risk, e.g., providing help, and at least one party benefiting from the interaction, e.g., receiving help. To mitigate the possible loss from individual risks, e.g., when interacting with exploitative agents, it is necessary that the longevity of the agent and the group be sufficient to recoup short term losses from long-term gains via interacting with cooperative agents in the society. For an agent community to benefit from complementary expertise, at least two conditions must be satisfied:

1. The environment presents sufficient cooperation opportunities. Such opportunities may include situations where one agent helps another one for a net utility gain (the cost of the helping agent is less than the savings of the helped agent) as well as situations that reward collaboration (where the cost of the collaborative effort of multiple agents in completing a task is less than the sum of the costs of agents working individually to complete the same task).

2. Agents possess and utilize modeling, adaptation, and reasoning methods that differentiate trustworthy collaborators from malicious or exploitative ones. Accurate identification of the cooperative nature of population member will then enable self-interested agents to nurture mutually beneficial relationships and avoid unrewarding ones.

We have been interested in studying environments and mechanisms that promote trusted interactions between self-interested agents. Most of our work in this area has assumed the presence of cooperation opportunities in the environment, and has therefore concentrated on identifying and evaluating computational mechanisms to

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.

leverage cooperation opportunities between self-interested agents [1, 4, 8, 12, 14, 13, 15, 16].

Our purpose in this paper is to analyze a representative environment to clearly delineate cooperation opportunities and predict stable coalitions that will develop in such environments. We analyze a recently proposed competition testbed, ART, which is an international collaboration with the goal of providing a common platform to evaluate, compare, and contrast alternative computational trust mechanisms [5]. We value the role of a common testbed in promoting comparative research in a field with a highly diverse set of research issues. Even though a single testbed cannot represent or evaluate all the different aspects of trust or the interaction modalities it relies on, we believe that the grounding provided by a well-specified testbed environment with clearly defined and unbiased evaluation criteria can help highlight the relative merits, costs, and scopes of different trust development and utilization schemes.

Our purpose in analyzing the ART testbed is to identify the cooperation possibilities present in the specified environment. We want to clearly identify the nature and extent of cooperation possibilities existing in the framework as well as specify what levels or frequencies of collaboration it facilitates. In the process we also hoped for recognizing features, constraints, and metrics in the ART testbed specification that can be tuned or changed to further increase cooperation possibilities. Our basic premise is that if ART is to be used as a representative testbed for evaluating alternative trust frameworks, it should be designed to not only enable and support trusted behavior but also recognize and reward mutual cooperation between individuals and groups. The ART framework is described briefly in a following section. We will study in this paper a more generic framework, of which ART is a specific instance.

The model we consider involve service providers serving a group of clients. Clients pick service providers proportionate to their past performance. Each service provider is an agent who is an expert in a specific task type in which its performance is significantly better than in other task types. An agent incurs a cost for doing a task but receives a compensatory payment from the customer who assigns it. A service provider can seek input from another service provider for an assigned task or can ask the other service provider to carry out the task on its behalf in return for a payment which is much smaller than the payment it receives from the client. An agent has only indirect knowledge of the expertise level of another agent and uses a trust metric to model this. Such models can also be based on reputation, i.e., while forming a trust estimate of agent Y, an agent X can combine its own trust value for Y with the trust values that other agents report for Y. Such reputation-based trust models can be used by an agent to decide which other agent to ask for help with a task type in which it is not an expert. The goal for an agent is to maximize its utility. Utility can be increased by garnering more customers, which can be done by improving performance. A service provider must form trusted relationships with other agents with complementary expertise to perform well over all task types. Thus, trust becomes a key ingredient in decision-making involving who to give and ask for help.

The paper has three distinct components: (i) we identify the nature and extent of cooperation possibilities in the ART framework and suggest modifications to the framework that will foster trusted cooperation, (ii) we present analytical derivation and experimental analysis of possible steady-state stable cooperative groupings under alternate client allocation schemes and ranges of system parameter values, and (iii) we recommend trust-based interaction strategies that can monitor for deviation from cooperative behavior thus suppressing exploitative tendencies in the population.

2. RELATED WORK

The Agent Reputation and Trust (ART) testbed [5] proposes a competition specification in an artwork appraisal domain, where agents evaluate paintings for clients. In order to perform better and produce accurate appraisals, an agent may gather opinions from other agents in the system. In a run, an appraiser (agent) has a set of clients, and each client will pay a fix amount f to obtain an appraisal for one painting. Each painting is classified in one of the finite set of eras (e.g. French Renaissance), and appraisers have varying levels of expertise in each era. For a given era, the expertise of an agent is modeled by an error distribution. In the ART testbed, the error between the agent's appraisal and the true value of the painting follows a normal distribution of mean zero and of standard deviation s given by

$$s = \left(s^* + \frac{\alpha}{c_g}\right) t$$

where

- s^* defines the intrinsic precision, or systematic error, of the agent for a specific era. For each era, the associated s^* is drawn from a uniform distribution, common to all agents. The smaller s^*, the more qualified in the era.
- t is the true value of the painting to be appraised. Its presence on the s enables to scale the error with the true price.
- α is a fix parameter shared by all agents. This parameter allows tuning the cost of generating an opinion with the accuracy of the opinion.
- and c_g is the cost that an agent will charge to provide an opinion. The larger c_g, the smaller the error.

An agent can ask other agents an appraisal for a painting, and also the reputation of other agents. To improve accuracy of its appraisal, an agent can seek opinions (i.e. appraisals) from other agents. For each opinion requested, it incurs a cost of c_p. Before purchasing the opinion from another agent, an agent can seek the reputation of an agent from other agents with an associated cost of c_r. The following holds true:

$$c_r \ll c_p \ll f,$$

where f is the fixed fee paid by the clients for each appraisal request. These inequalities ensure that an agent can afford to pay the opinions from other agents, and that it is affordable to ask the reputation of an agent.

The total number of clients remains fixed throughout the game. Initially, the clients are evenly distributed among appraisers. The appraisers whose final appraisals are most accurate are rewarded with a larger share of the client base in the following timestep. To compute the number of clients for the next timestep, the following quantity are computed:

the average relative error ϵ_a for an appraiser a is

$$\epsilon_a = \frac{\sum_{c \in C_a} \frac{|p_c^* - t_c|}{t_c}}{|C_a|},$$

where C_a is the set of a's clients in the current timestep, and $\frac{|p_c^* - t_c|}{t_c}$ is the error in evaluating a query from its client c (p_c^* is the appraisal of the agent for client c and t_c is the true valuation of the painting).

the relative non-error for an agent a is

$$\delta_a = 1 - \frac{\epsilon_a}{\Sigma_{b\in A}\epsilon_b}, \quad (1)$$

The client size r_a^{new} of agent a for the next timestep is given by

$$\tilde{r_a} = \left(\frac{\delta_a}{\Sigma_{b\in A}\delta_b}\right)|C|, \text{ and} \quad (2)$$

$$r_a^{new} \leftarrow q.r_a + (1-q).\tilde{r_a}. \quad (3)$$

where C is the total number of clients in the system, r_a is the current number of a's clients and q represents a *volatility factor*, which reflects the influence of current client size $|C_a|$ on the client size in the next run.

The idea of coalition formation has been investigated by the game theory community. Self interested agents form a coalition when the utility obtained in a coalition is higher than the utility obtained when they are working alone. In this paper, we focus on long term coalition as in [19], but we are only concerned with the potential increase of utility, which justifies the formation of the coalition, due to the presence in a coalition. We are not concerned with the distribution of the utility to the agents [3, 10] or the formation process [9, 17, 19].

3. TRUST ANALYSIS

We study the ART testbed and analyze the potential existence of cooperative groups or coalitions of self interested agents: by forming a coalition, agents may increase their private utility. To aid analysis, we make certain simplifications, e.g., assume extreme values of some parameters in the ART framework, while always clarifying the reason and the implications of these simplifications.

- An appraiser agent (agent from here on) is an expert in only one era (task type from here on). We consider also that s^* can take two values only: s^*_e or s^*_n. For only one task type, $s^* = s^*_e$ is a low value, and for all other task types $s^* = s^*_n$ is a higher value. By making this simplification we create the need for cooperation between appraisers as no appraiser can perform well, by itself, on multiple task types.

- There is no cost in seeking reputation information about other agents in a group, i.e., $c_r = 0$, which enables free flow of reputation information to facilitate the identification of the expertise of other assessors.

- The cost of seeking opinion from another agent, c_p, balances out when the agents in a group consult each other. This happens when there are reciprocal exchanges and allows us to ignore the opinion cost while deciding whether or not to ask for help.

- α is very close to zero, and hence the error approximately is given as s^*t. This simplification enables us to make an analytical analysis.

- A group of agents is fully cooperative: agents help when asked, and agents are not deceptive. For example, an assessor will report its honest assessment of a painting when its opinion is sought by another assessor.

- For the singleton agents, i.e., agents who do not work in a group, we assume they work on their own and do not seek opinion from other (even singleton) agents.

Thus we work with a scenario more conducive to group formation. Our next goal is to identify when agents can benefit by forming coalitions, and the performance gains compared to singleton agents.

3.1 Trust Opportunities

In order to improve its utility, an agent needs to increase its number of clients (with our simplifying assumption, there is no cost). Because the total number of client is fixed, increase a clientele requires improving its relative accuracy. Since an agent is an expert in only one of the many task types, it has to rely on other agents for task types in which it is not expert. Therefore, an agent has the incentive to build long-term, trusted relationship of mutual cooperation with agents of complementary expertise. Agents who form stable groups with agents of complementary expertise can then significantly improve performance and thereby their utility, as their client base increases proportionately. The grand coalition produces the most accurate appraisals, but the agents' utility is the same as if they all work on their own. Some beneficial coalitions lies between these two extremes.

3.1.1 Analytical Treatment

We now analyze how much an agent can benefit if another agent joins its group or coalition. We use the phrase "rewarding scheme" in our discussion to refer to the client share allocation policy for the next round used in ART. For the purpose of our discussion, performance of an agent is measured by its number of clients.

We assume that the accuracy of an appraisal is not significantly improved by consulting more than one expert in that task type[1]. This means that initially, experts of complementary expertise will try to form coalitions. Let A be the set of all agents, K be the set of task types, k be the number of task types ($k = |K|$), $e(a)$ be the expertise of agent $a \in A$, and E_G represents the set of expertise already represented in group G, i.e., $E_G = \cup_{a\in G} e(a)$. Group G prefers to welcome an agent x over another agent y, iff, $e(x) \notin E_G$ and $e(y) \in E_G$. Only when a group does not contain experts of all task types will other agents be considered for inclusion in the group (since the domain of the paintings are uniformally distributed, there is little incentive to specialize in a set of domains, and we assume that all expertise is needed).

For a painting in a given domain, each agent in G will produce an opinion: experts in the domain will perform at the expert level and non-expert will perform at non-expert. The average error of an agent who is part of group G that does not have experts for all task types ($|G| < k$) is:

$$\epsilon_{|G|} = \frac{s^*_e|G| + s^*_n(k - |G|)}{k}.$$

Recall that s^*_e is the s^* value for an expert agent, and s^*_n is the s^* value for a non-expert agent.

Let us take a concrete scenario to illustrate the performance of coalitions of experts. Let $|A| = 10$, $k = 6$, $s^*_e = 0.01$, $s^*_n = 0.9$. Then the average errors for a group of 2 agents (i.e. ϵ_2) and a group of 1 agent (non-cooperating agent, ϵ_1) are:

$$\epsilon_2 = \frac{0.01*2 + 0.90*4}{6} \Rightarrow \epsilon_2 \approx \frac{0.90*4}{6}, \; and,$$

$$\epsilon_1 = \frac{0.01*1 + 0.90*5}{6} \Rightarrow \epsilon_1 \approx \frac{0.90*5}{6}.$$

Thus the average error rate of the non-cooperating agent is 20% more than that of an agent who is part of a group of 2 experts.

[1] While this is possible in practice, the corresponding calculations are more complex and will be taken up in the future.

We now consider a situation where there is only one coalition formed by two agents, all remaining agents are working on their own. We want to find out the effect of the different error on the allocation of clients in future timesteps. To derive the performance we first calculate the relative non-errors in the ART framework from Equation 1 (in a slight notational abuse we use δ_i to refer to the relative non-error of an agent belonging to group of size i, which is somewhat different from the use of the subscript in Equation 1):

$$\delta_2 = 1 - \frac{\epsilon_2}{\sum_{i=1}^{A} \epsilon_i} = 1 - \frac{\epsilon_2}{2 * \epsilon_2 + 8 * \epsilon_8},\ and,$$

$$\delta_1 = 1 - \frac{\epsilon_1}{\sum_{i=1}^{A} \epsilon_i} = 1 - \frac{\epsilon_1}{2 * \epsilon_2 + 8 * \epsilon_8}.$$

From these values, and using Equation 2 we can see that their relative shares will be in the ratio 44:43 in the steady state (when the updates using Equation 3 have converged). Hence, we see that a 20% less error translates into only a 2.3% performance gain using the customer share allocation scheme proposed in ART. The actual situation in the ART testbed experimentation may be even more problematic as 20% error reduction can be achieved in a rather idealized scenario ignoring costs, etc. and as described in the beginning of this section.

We attribute this apparent problem to the relative non-error calculation scheme used in ART, and more specifically to the denominator in Equation 1 which depends on the size of the assessor population. In larger population, the ratio average relative error will be extremely small. More concretely, let $\epsilon_{21} = \frac{\epsilon_1}{\epsilon_2}$ be the relative error advantage of agents in a group of size 2 over singleton agents. Then the relative performances of these agents are given by the ratio

$$\delta_{21} = \epsilon_{21} \frac{|A| - 2 + \frac{1}{\epsilon_{21}}}{(|A| - 3)\epsilon_{21} + 2},$$

which is a decreasing function of $|A|$. This function is represented in Figure 1. Note that the performance advantage of coalition members rapidly vanishes in larger populations in spite of the same superiority in error rates. The maximum performance advantage of a group member over a non-group member occurs when experts of all the task types are represented in the group, and even in this case, i.e., $|G| = 5$, the group member has only a $\approx$ 1% performance advantage over non-group members when the population has 100 assessors! Thus coalition members will not perform noticeably better than individual agents, and hence there is no tangible incentive for cooperative group formation in a relatively large population of assessors in the current formulation of the ART testbed.

We now propose an alternate, client share function that rewards experts proportionately more as their error rate decreases:

$$\delta_a = \frac{\sum_{b \in A} \epsilon_b}{\epsilon_a}$$

We call this the *proportionate* customer share allocation scheme. Note that another advantage of this formulation is that δ_a can now be interpreted as the probability with which a customer agent selects a. When $\epsilon_{21} = 1.2$, for example, with this new scheme, we have $\delta_{21} = 1.25$, i.e. coalition members will receive 25% more clients in this case which provides significant incentive for coalition formation. As another example, if the only group in a population of 10 consists of four experts, i.e., $|G| = 4$, then the relative share of the experts compared to singleton agents is 1.09 using the ART share allocation scheme and approximately 2.5 using our proposed scheme.

By redefining the share allocation function, we have created a more supportive or congenial environment for cooperation. This

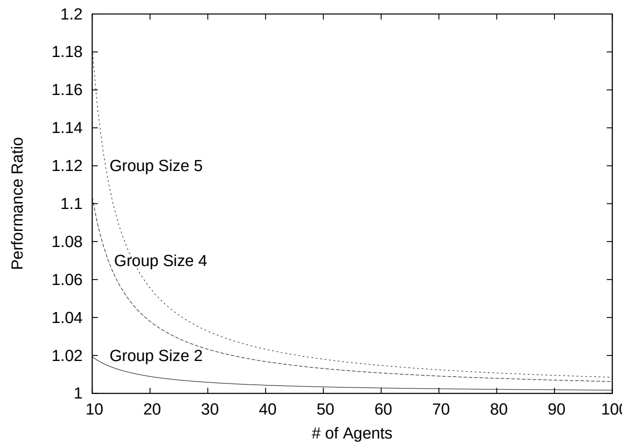

Figure 1: Ratio of customer share for each member of a group of agents to singletons for varying number of assessors and different group sizes.

suggests a small but far-reaching change in the design of the ART framework that significantly elevates the cooperation opportunities in the environment. In the following section we run some simulations of a local, simplified implementation of the ART framework[2].

3.1.2 Experimental Setup

In the following experiments, $|A| = 10$, $|C| = 10,000$, $k = 5$, $s_e^* = 0.01$, $s_n^* = 0.9$ and $q = 0.1$. We simulate group formation among cooperating agents such that agents with different types of expertise will form coalitions to improve performance. We assume only one coalition in the population (we will later relax this assumption). To compute the assignment of clients in successive iterations, we sample the error distribution ($\mathcal{N}(0, 0.01)$ when the agent is expert, $\mathcal{N}(0, 0.9)$ when it is not expert). We increase the number of cooperating agents (i.e. group size) from 2 to 10 and note the average number of clients assigned to the agents at the end of each timestep. We ran our experiment for a fixed number of timesteps, and then estimate the average client size for the cooperating and non-cooperating agents over the total number of runs.

3.1.3 Simulation Results

Figure 2 shows the result when we use ART's client share function. We observe that the number of clients obtained by a cooperating agents increases with the group size: experts for additional domains enter the group, which helps the group members to increase the accuracy of their appraisals in any domain. Non-cooperating agents, as their accuracy is fixed, loose clients in profit of the coalition of agents. Though the performance difference between group and non-group members is noticeable for larger group sizes, the difference is almost negligible for a group size of two. This may be problematic to initiate the coalition formation process. As noted above, we may not even realize this performance difference in the actual ART testbed when we incorporate the costs of cooperation, e.g., seeking reputation, opinion, etc.

We also note that the performance of the coalition stabilizes after experts of all different task types have joined the coalition, and thereafter starts falling. As only one expert is asked, new experts do not add any value to the group, hence, the number of clients remains the same, and the revenue has to be shared by more agents. The

[2]The ART testbed is not available for experimentation at the this time.

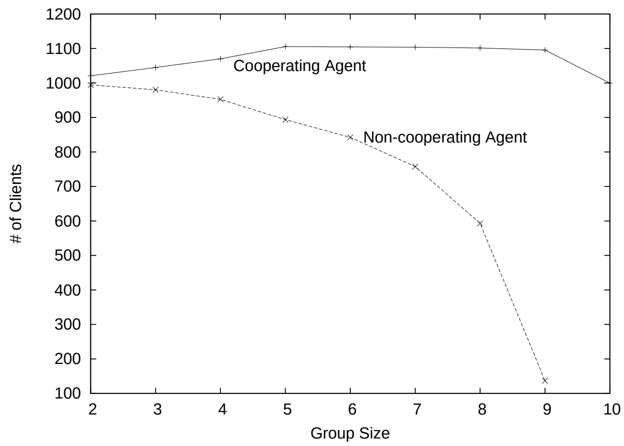

Figure 2: Group Size Vs # customers of Cooperating and Non-Cooperating Agents using the customer share scheme of the ART testbed.

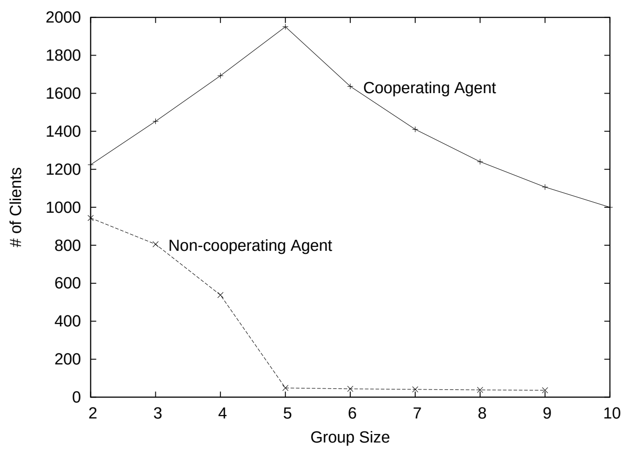

Figure 3: Group Size Vs # customers of Cooperating and Non-Cooperating Agents using the ***proportionate*** **customer share scheme.**

customer share of the non-group members decrease at a faster rate. Hence there will be an incentive to join well-established groups, i.e., to cooperate with group members to earn their trust, rather than working on one's own.

We now present the corresponding results when using our proposed *proportionate* customer share allocation function. The results (presented in Figure 3) show that the cooperating agents gain substantially over the non-cooperating ones at all group sizes. In particular, with 2 agents cooperating, the average number of clients of the cooperating agents reaches around 1200 compared to about 950 for non-cooperating agents. Unlike the ART customer share allocation, there is now a real incentive for two agents to start a coalition. When 5 agents cooperate, i.e., when the group has expertise in all the areas, the number of clients goes up to 1900. We emphasize that this substantial gain in clients (and hence revenue) is a decisive factor in promoting trusted behavior among the assessors.

Similar to the previous case though, the performance of the group degrades once the group size increases beyond k. In the next section, we analyze the question of optimal coalition sizes in such environments.

3.2 Coalition Forming Decisions

In the previous section, we discussed the variation in performance of coalition of different sizes. This leads us to the question of the optimality of a coalition size. The analytical and experimental results suggest that a group should stop accepting new agents once it contains experts for each domain of expertise. By accepting, we mean the experts in the group will trust and cooperate with this new applicant. A basic "admission decision function" or selection function used by a group G to admit agent a can be:

$$S(G,a) = \begin{cases} Yes & \text{if } \delta_{G\cup a} > \delta_G \\ No & \text{otherwise} \end{cases} \tag{4}$$

where δ_G is the union of each member's set of customer. If we assume that G is the only group in the population, a reasonable selection function, as discussed above, is the following:

$$S_1(G,a) = \begin{cases} Yes & \text{if } e(a) \notin E_G \\ No & \text{otherwise.} \end{cases}$$

The above reasoning makes the critical assumption that there would be only one group in the population of assessors. In practice, however, an assessor who is not accepted into a group is free to cooperate with any other non-group member in the population, and hence, it can start another group. A better interpretation of Equation 4 is that the group should select an individual if the marginal utility of selecting the individual is positive. In general, the marginal utility calculation should consider the average number of customers of the members of G over all possible groupings of the population with and without accepting a in G. Such a calculation is not only computationally demanding but may also not be feasible given the knowledge that members of a group possess about other groups.

We present a coarse approximation of this marginal utility calculation, which can suffice for a number of common environments including situations with a relatively small number of assessors (more likely in the ART framework). In our approximation scheme, a group G considers the situation that if rejected, a might initiate the formation of a new coalition H composed of experts of distinct expertise: G assumes that in the worse case, it will compete with another group H and $|A| - (|G| + |H|)$ singletons. Now, we can present the approximated marginal utility based decision function as

$$S_2(G,a) = \begin{cases} Yes & \text{if } \delta_{G\cup a} > \delta_{G,H} \\ No & \text{otherwise,} \end{cases} \tag{5}$$

where $\delta_{G,H}$ is the number of customers of each member in G when there is another coalition of experts of distinct types, H, in the population.

3.3 Experimental Setup

In our simulation, we create a fixed size coalition H, a coalition G of varying size, and the rest of the agents work independently. We run experiments for $|A| = 10$, $k = 5$, $|H| = 2$ and $|G| \in [5 \ldots 9]$, where G contains at most one expert in a given domain of expertise. We plot the performance of the two groups in Figures 4 and 5 for our proposed customer share function and the one used in the ART testbed respectively. For each figure, we add the performance of G when its size is $|G| \in [2 \ldots 4]$; in this range, G always accepts experts of new types, hence H will form only after $|G| \geq k$.

As we have noted before, performance of members of G peaks at $|G| = k$, i.e., when the group has all the experts. We denote this group by G^*. In Table 1, we present the expected number of clients of an agent a for its different options: joining G^*, staying on its

Action	Number of clients
Joins existing group	1630
Stays alone	50
Forms a group of 2 members	500
Forms a group of 3 members	790
Forms a group of 4 members	1129
Forms a group of 5 members	1000

Table 1: Relative comparison of the Client counts for a new member in different scenarios.

own, or forming a new coalition H. The results clearly demonstrate that a benefits the most by joining G^*.

At this point, when a new member seeks admission in the group, the group has to perform the marginal utility calculation of Equation 5. At any group size $|G| \geq |G^*|$, the group should compare $\delta_{G \cup a}$ (corresponding to the curve labeled $r_{1\text{ group}}$ in Figures 4 and 5), i.e., performance of a group of $|G|+1$ agents, with $\delta_{G,H}$, i.e., performance when another coalition of two experts compete with G (corresponds to the curves labeled $r_{2\text{ groups}}$ in Figures 4 and 5). We first explain the situation in our proposed customer sharing scheme (see Figure 4).

Starting at $|G^*| = 5$, we first observe that, if G rejects the potential new member and this one forms a new coalition with another agent, the performance of G drops to $r_{2\text{ groups}}(6)$[3]. Then, we observe that $r_{1\text{ group}}(6) > r_{2\text{ groups}}(6)$, which means that instead of letting the newcomer form a new coalition, the coalition G is better off by accepting it, even if it incurs a loss. Similar reasoning shows that a group of size 6 will also accept a new applicant. For a group of size 7, however, accepting a new member is worse than letting a new coalition of size two form. Under the assumption that the largest competitive group that can form is size 2, the marginal utility based decision in this case is to reject the applicant. Hence, under these condition and assumptions, the optimal group size in the population is 7. For a more accurate calculation, G should consider all possible competitive group sizes before denying an applicant. We surmise that under these more extensive calculations either of the following scenarios will arise: (a) the entire population will form a single group, (b) the entire population will break up into equal sized groups. In either case, all agents will receive the same share of the customers.

If we now consider the question of optimal group size for the customer share allocation scheme in the ART framework, from Figure 4 and similar reasoning as above we find that G should never reject any applicant and the optimal group size is 10! So, if the coalition can get started, without much initial incentive, it can grow to encompass the entire population.

In the above discussion we have assumed that all members are truthful and trustworthy. But an agent has the incentive to receive help but not help back. In the next section we briefly review mechanisms from our prior work which can be adapted to address the issue of trust management in the ART framework.

3.4 Whom to Trust

So far, in this paper, we have identified cooperation possibilities

[3]In these figures, the $r_{2\text{ groups}}$ curve is offset one position to allow easy marginal utility based decision making: $r_{2groups}(x) = \delta_{G,H}$ where $|G| = x-1$ and $|H| = 2$. This means the $r_{2\text{ groups}}(x)$ corresponds to the performance of a member of group G of size x-1 against a group H of size 2.

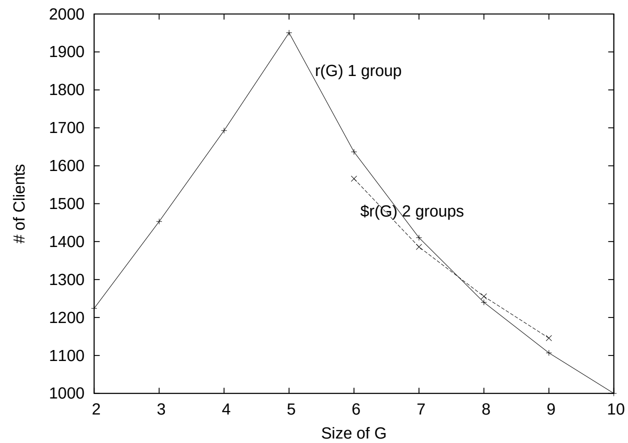

Figure 4: Performance of coalition G: when G is the only coalition (r 1 group) and when there are two coalitions (r 2 groups): G and a coalition H of size 2. (*proportionate* customer share scheme).

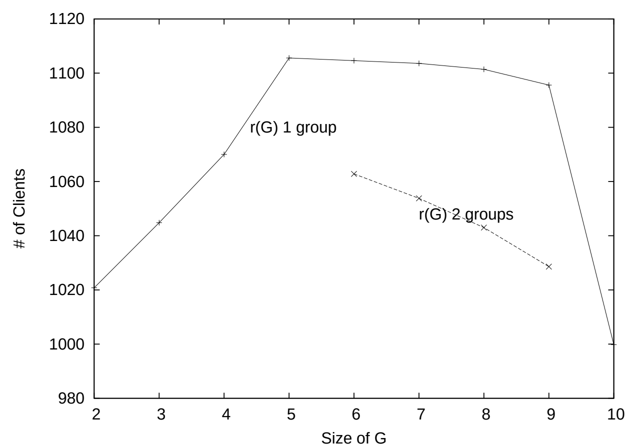

Figure 5: Performance of coalition G: when G is the only coalition (r 1 group) and when there are two coalitions (r 2 groups): G and a coalition H of size 2. (*proportionate* customer share scheme).

in the ART testbed, and identified optimal cooperative group sizes assuming that agents will cooperate in a group to maximize the average performance of the group members, i.e., group members will readily help other group members. Such a naive assumption will be invalidated assuming self-interested agents would be interested in receiving help, i.e., saving cost, but not providing help, i.e., incurring cost. This is particularly true in the absence of any omniscient arbiter to penalize deviations from cooperative behavior. So, the existence of cooperation opportunities can only be successfully utilized if agents use trust-mechanisms that identify and reward cooperative agents and shun exploitative agents. Such trust management schemes are necessary for agents to perform well in the ART testbed. We now present a set of criteria that should be satisfied by an effective trust-management scheme:

- allow agents to be inclined to help someone who has the potential to provide help,
- allow comparison between different cooperation costs,

- be able to flexibly adjust inclination to cooperate based on current work-load (e.g., more inclined to cooperate when less busy, etc.)
- be responsive to change in types of tasks and types of expert in the population.

We present an expected utility based decision mechanism, from our prior work, that can be used by the agents to decide whether or not to honor a request for help from another agent [16]. When requested for help, an agent, using this decision mechanism, estimates the utility of agreeing to the request by evaluating its chance of obtaining help from the asking agent in future. An agent, being self-interested, has the objective of earning more savings by receiving help than cost incurred by helping others in the long run. When an agent using this strategy decides whether or not to provide help, it uses a statistical summary of its past interactions with the requesting agent as a metric for evaluating its expected interaction pattern with the latter in future. Using this information, it evaluates the difference between the expected benefit and the expected cost it might incur for that agent by helping it in the future. In the following, we present the expected utility based decision mechanism that agent m uses to evaluate a help request by another agent o for helping with task type τ. The expected utility of agent m for interacting with agent o at time $\mathcal{T}$ and future time steps, $E_{\mathcal{T}}(m, o, \tau)$, is defined as:

$$E_{\mathcal{T}}(m, o, \tau) = \sum_{t=\mathcal{T}}^{\infty} \gamma^{t-\mathcal{T}} [\sum_{x \in \Upsilon} (D_m^t(x) \mathrm{Pr}_{m,o}^t(x) cost_m(x)) - \sum_{x \in \Upsilon} (D_m^t(x) \mathrm{Pr}_{o,m}^t(x) cost_m(x))] - cost_m(\tau), \quad (6)$$

where $cost_i(x)$ is the expected cost that i incurs doing a task of type x, $cost_m(\tau)$ is the cost to be incurred by agent m to help agent o in the current time instance, γ is the time discount, and Υ is the set of different task types. We assume that an agent is expert in only one of the possible task types. The evaluation of the expected utility of agent m helping agent o considers all possible interactions in future and for all task types. In equation 6, $D_m^t(x)$ is the expected future distribution of task types that agent m will receive at time instance t. We define $\mathrm{Pr}_{i,j}^t(x)$ as the probability that agent i will receive help from agent j at time step t, given it has a task of type x. While personal interaction history can aid in forming an estimate of these probabilities, the reputation of o obtained from other agents can be combined with personal experience when m has had limited interactions with o.

The term $\sum_{t=\mathcal{T}}^{\infty} \gamma^{t-\mathcal{T}} \sum_{x \in \Upsilon} D_m^t(x) \, \mathrm{Pr}_{m,o}^t(x) cost_m(x)$ represents the time discounted (with discount factor γ) expected savings of m by receiving helps from o in future. We assume that when an agent is helped by another agent, the helped agent incurs no cost for the task. Hence, when an agent m is helped with task type x, its savings is $cost_m(x)$, the cost it would have incurred to complete the same task on its own. We use an infinite time horizon and increasingly discount the impact of estimates for future interactions by the factor $\gamma^{t-\mathcal{T}}$, where $0 < \gamma < 1$, and t refers to the time period. The sum of the terms $-cost_m(\tau)$ and $\sum_{t=\mathcal{T}}^{\infty} \gamma^{t-\mathcal{T}} \sum_{x \in \Upsilon} D_m^t(x) \, \mathrm{Pr}_{o,m}^t \, cost_m(x)$ is the net expected cost that can be incurred by m for (a) helping on the current time instance and (b) incurring helping cost for o in the future. Thus, $E_{\mathcal{T}}(m, o, \tau)$ gives the net time-discounted future expected benefit that agent m has for interacting with agent o.

Our prescription is for agent m to help agent o in the current time if $E(m, o, \tau)$ is greater than the expected cost of an agent processing its next assigned task without receiving help from others, i.e., if $E_{\mathcal{T}}(m, o, \tau) + \sum_{x \in \Upsilon} D_m(x) cost_m(x) > 0$, where the summation term is the expected cost of agent m for doing a task in the next time instance. This summation term represents the initial bias or willingness for an agent m to help another agent o incurring a risk of not being reciprocated. As initial probability values are all zero, the agent will not help another without this initial bias.

In addition to the above method, a number of other approaches to trust-based reasoning that have been proposed by multiagent system researchers can also be gainfully applied to evaluate the trustworthiness of other experts in the population [2, 6, 7, 11, 18, 20]. The use of such trust schemes acts as a disincentive to deviation from cooperative behavior and can help sustain effective groups where the workload is balanced by exchanging help.

4. CONCLUSION AND FUTURE WORK

We have studied a general model where service providers, each an expert in a given task type, are processing tasks assigned by a fixed set of customers. The customers are distributed among the providers proportionate to their performance. To improve performance, service providers can seek and receive help from others to process tasks that are not in their area of expertise. As a specific example, we have considered the ART experimental testbed proposed for comparative evaluation of trust-management schemes.

We provide a detailed analysis that demonstrates the nature and frequency of cooperation opportunities in such environments. Rather surprisingly, we find that the customer share assignment scheme in the ART framework does not effectively reward high-performing providers. We suggest a modification to this scheme that will make the testbed better differentiate between effective and non-effective trust-management schemes.

We also derive a decision mechanism that allow cooperative group members to decide whether or not to include new members in the group. It is interesting to note that it might be beneficial to accept new group members that would lower group performance if not doing so can mean a worse loss from more effective competitions from other groups.

The presence of sufficient cooperation opportunities or the derivation of optimal cooperative group sizes can only be used in practice if effective trust-management schemes exist that can reward cooperative behavior and suppress exploitative ones. We identify the desirable features of such trust-management schemes, and present an adaptation of our previous work as an example scheme that possess these desirable features.

We are currently working on developing a general mechanism for deciding whether or not to accept a new member to a group that will take into consideration larger competing group sizes. While deriving the cooperation opportunities, we made simplifying assumptions that minimized the cooperation cost. We plan to develop more elaborate models that considers realistic costs and will be able to provide more accurate cooperation decisions. The ART testbed considers a fixed customer population. Hence, the performance of experts without any cooperation is the same as the performance of a group of cooperating experts. We are interested in a more open environment, where the utility of a provider is a monotonic function of its performance. We will analyze the unique trust requirements of such open environments.

5. REFERENCES

[1] A. Biswas, S. Sen, and S. Debnath. Limiting deception in groups of social agents. *Applied Artificial Intelligence, Special Issue on Deception, Fraud, and Trust in Agent Societies*, 14(8):785–797, 2000.

[2] C. Castelfranchi, R. Conte, and M. Paolucci. Normative reputation and the costs of compliance. *Journal of Artificial Societies and Social Simulation*, 1(3), 1998.

[3] V. Conitzer and T. Sandholm. Complexity of determining nonemptiness of the core. In *Proceedings of the International Joint Conference on Artificial Intell igence (IJCAI)*, 2003.

[4] P. S. Dutta and S. Sen. Forming stable partnerships. *Cognitive Science Research*, 4(3):211–221, 2003.

[5] K. K. Fullam, T. B. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter, and M. Voss. A specification of the agent reputation and trust (ART) testbed: Experimentation and competition for trust in agent societies. In *Proceedings of the Fourth Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 512–518, New York, NY, 2005. ACM Pres.

[6] A. Glass and B. Grosz. Socially conscious decision-making. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 217–224, New York, NY, 2000. ACM Press.

[7] N. Griffiths and M. Luck. Coalition formation through motivation and trust. In *Proceedings of the Second Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 17–24, New York, NY, 2003. ACM Pres.

[8] P. P. Kar, S. Sen, and P. S. Dutta. Effect of individual opinions on group interactions. *Connection Science*, 14(4):335–344, 2002.

[9] M. Klusch and A. Gerber. Issues of dynamic coalition formation among rational agents. In *Proceedings of the Second International Conference on Knowledge Systems for Coalition Operations*, pages 91–102, 2002.

[10] S. Kraus, O. Shehory, and G. Taase. Coalition formation with uncertain heterogeneous information. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 1–8. ACM Press, 2003.

[11] M. Schillo, P. Funk, and M. Rovatsos. Using trust for detecting deceiptful agents in artificial societies. *Applied Artificial Intelligence*, 14:825–848, 2000.

[12] S. Sen. Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents. In *Proceedings of the Second International Conference on Multiagent Systems*, pages 315–321, Menlo Park, CA, 1996. AAAI Press.

[13] S. Sen. Believing others: Pros and cons. *Artificial Intelligence*, 142(2):179–203, 2002.

[14] S. Sen and A. Biswas. Effects of misconception on reciprocative agents. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 430–435, New York: NY, 1998. ACM Press.

[15] S. Sen and P. S. Dutta. The evolution and stability of cooperative traits. In *Proceedings of the First Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1114–1120, New York, NY, 2002. ACM Pres.

[16] S. Sen, S. Saha, and P. S. Dutta. Helping based on future expectations. In *Proceedings of the Second Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 289–296, New York, NY, 2003. ACM Pres.

[17] L.-K. Soh, C. Tsatsoulis, and H. Sevay. A satisficing, negotiated, and learning coalition formation architecture. *Distributed Sensor Networks: A Multiagent Perspective*, pages 109–138, 2003.

[18] Y. Tan and W. Thoen. An outline of a trust model for electronic commerce. *Applied Artificial Intelligence*, 114(8):849–862, 2000.

[19] J. Vassileva, S. Breban, and M. Horsch. Agent reasoning mechanism for making long-term coalitions based on decision making and trust. *Computational Intelligence*, 18(4):583–595, 2002.

[20] B. Yu and M. P. Singh. Detecting deception in reputation management. In *Proceedings of the Second Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 73–80, New York, NY, 2003. ACM Pres.