A Study of Scalability Properties in Robotic Teams*

Avi Rosenfeld, Gal A Kaminka, Sarit Kraus

Bar Ilan University, Ramat Gan, Israel

Summary. In this chapter we describe how the productivity of homogeneous robots scales with group size. Economists found that the addition of workers into a group results in their contributing progressively less productivity; a concept called the Law of Marginal Returns. We study groups that differ in their coordination algorithms, and note that they display increasing marginal returns only until a certain group size. After this point the groups' productivity drops with the addition of robots. Interestingly, the group size where this phenomenon occurs varies between groups using differing coordination methods. We define a measure of *interference* that enables comparison, and find a high negative correlation between interference and productivity within these groups. Effective coordination algorithms maintain increasing productivity over larger groups by reducing the team's interference levels. Using this result we are able to examine the productivity of robotic groups in several simulated domains in thousands of trials. We find that in theory groups should always add productivity during size scale-up, but spatial limitations within domains cause robots to fail to achieve this ideal. We believe that coordination methods can be developed that improve a group's performance by minimizing interference. We present our findings of composite coordination methods that provide evidence of this claim.

1 Introduction

Teams of robots are likely to accomplish certain tasks more quickly and effectively than single robots [9, 12, 23]. To date, only limited work has been performed on studying how performance scales with the addition of robots to such groups. Should one expect linear, exponential, or decreasing changes in productivity as the group size grows? Previous work by Rybski et al. [23] demonstrated that groups of identical robots do at times demonstrate marginal decreasing returns. As such, their productivity curves resembled logarithmic functions; the first several robots within their group added the most productivity per robot and each additional robot added successively less. In contrast, Fontan and Matarić [26] found that robotic groups reached

* This material is based upon work supported in part by the NSF under grant #0222914 and ISF grant #1211/04 . Sarit Kraus is also affiliated with UMIACS.

a certain group size, a point they call ”critical mass”, after which the net productivity of the group dropped. Similarly, Vaughan et al. [29] wrote that the rule of ”too many cooks” applies to their groups and adding robots decreases performance after a certain group size.

Economists have studied the gains in productivity within human groups. According to their Law of Marginal Returns, if one factor of production is increased while the others remain constant, the overall returns will relatively decrease after a certain point [4]. As the size of the group becomes larger, the added productivity by each successive worker is likely to become negligible, but never negative. This classical model contains no reference to a concept similar to a ”critical mass” group size after which the added worker decreases the total productivity of the group.

Our research goal is to understand when the marginal returns predicted by the economic model would be consistently realized as work by Rybski [23] found they were, and when adding robots would decrease performance as Fontan and Vaughan [26, 29] described. Towards this goal, we first analyze several existing group coordination algorithms and empirically observe the different groups' productivity with the addition of robots. We observe that the different coordination techniques affect the productivity graphs of these groups during scale up.

To determine the cause for the differences between coordination algorithms, we define a measure of interference that facilitates comparison, and find a high negative correlation between group interference and productivity. Effective coordination algorithms maintain marginal productivity over larger groups by reducing interference levels. Using this result we are able to examine robotic group productivity in several simulated domains in thousands of trials. We find that groups in theory always produce marginally, but that competition over space causes robots to deviate from this ideal.

We believe this result can aid in studying the scalability qualities of robots. First, our interference metric is useful post-facto, for understanding the scalability qualities within robotic groups. The effectiveness of a coordination method can be judged based on its ability to minimize interference. A team's ability to scale will be hampered if interference is not kept in check. Additionally, we believe interference can be used in an online fashion to increase the group's productivity and scalability. We present preliminary results of composite coordination methods that indicate that our interference metric can be used to adapt a group's coordination activities to the needs of the domain. For future work, we plan to further study the use of this metric in improving the scalability, and performance qualities of robotic groups.

2 Related Work

The study of robotic groups is quite important for several reasons. Certain tasks require groups of robots. For example, a large hazardous item might require the combined strength of several robots to physically move it. Other tasks can be accomplished through groups of robots more quickly and robustly. Rybski et al. [23] demonstrated that groups of robots are likely to finish certain collection tasks faster than one robot. Groups of inexpensive robots are also useful in certain domains where there is a high probability damage will be incurred by any single robot. Thus, tasks such as mine clearing are well suited for groups of inexpensive robots. In this work we study the scalability qualities of these type of robotic tasks, but many of our results are likely to be useful for other categories of robotic activity as well.

We study methods for improving upon the productivity of robotic groups through improving the coordination methods in these groups. At the logical level, various formal frameworks for teamwork have been proposed such as the joint intentions theory of Cohen and Levesque [5], Grosz and Kraus' SharedPlans [11], and Joint Intentions [14] have been presented for creating a cohesive team unit. Several practical teamwork implementations have been proposed for dynamic environments based on these models. The GRATE* teamwork method [14] is based on creating Joint Recipes based on the needs of a specific domain. The STEAM [28] teamwork engine is based on creating a set of domain independent team rules. All of these frameworks revolve around having the members of the group agreeing to and maintaining a mutual beliefs among all members of the group. These beliefs are often explicitly communicated, and team members require robust sensing and communication capabilities. Finally, a behavior based approach, Alliance [20], operates through members of a robot team using impatience and acquiescence behaviors to create teamwork. This approach does not explicitly model teamwork and relies on using team behaviors within each robot to create team cohesion.

A second model of group behavior revolves around swarm group behaviors, instead of formalized teamwork. Swarm behaviors typically involve homogeneous groups of members with limited processing and operating ability. Often these models are inspired from group activity of animals [17, 21]. Such approaches are typically best suited for domains where large groups are available, the task does not require tight cooperation between group members, and robust sensing and communication abilities do not exist in group members. Dudek et al. [6] present a taxonomy of these and other possible categories.

Between these extremes lies numerous possibilities. Swarms could be created with high level reasoning and sensing abilities. These large groups could use high level team reasoning skills. For example, Scerri et al. [25] presents a scalable approach where large teams are based on dynamically evolving subteams. This work presents the challange of creating effective coordination methods that can scale. Novel coordination approaches are needed in addressing this issue.

Our research goal in this work is to understand how to increase the effectiveness of robotic groups' coordination during scale-up. Previous work by Fontan and Matarić [26] noted that proper coordination lies at the root of effective group behavior. As such, the creation of effective coordination is critical for achieving high productivity within a group. Our first step was to study how adding robots effects the groups' productivity. We wished to ascertain when adding foraging homogeneous robots hurt group performance as [26] and [29] predict they will after a certain team size, and when these robots continuously adds to the team's performance as Rybski et al. found [23].

Several coordination methods have been developed for use within the foraging domain. This domain is formally defined as locating target items from a search region *S*, and delivering them to a goal region *G* [10]. We began by studying this domain because of the wealth of existing research conducted within this environment [9, 10, 19, 23, 26, 29].

The foraging domain is characterized by a limited field of operation where spatial conflicts between group members are likely to arise. Many other robotic domains such as waste cleanup, search and rescue, planetary exploration and area coverage share this trait. In fact, this paper demonstrates that our foraging results were equally applicable within a second search domain.

We first studied the interplay between the success of group's coordination and the corresponding productivity during group scale up. Several coordination methods have been developed for use within the foraging domain. For the sake of simplifying the comparison, we initially only contrasted methods that operate on homogeneous robots, do not require prior knowledge of the domain, and do no require any communication. Arkin and Balch [1] describe a system of using repulsion schema any time a robot projects it is in danger of colliding. It additionally adds a noise element into its direction vector to prevent becoming stuck at a local minima. Vaughan et al. [29] describe an algorithm that uses *Aggression* to resolve possible collisions by pushing its teammate(s) out of the way. They posit that possible collisions can best be resolved by having the robots compete and having only one robot gain access to the resource in question. A third approach, is a dynamic *Bucket Brigade* mechanism [19]. In this method, a robot drops the item it is carrying when it detects another robot nearby. In theory, the next closest robot should retrieve the recently dropped object and carry it closer to the goal. While this last method may be effective in foraging, it is limited to certain domains. This coordination method is not appropriate for certain tasks such as searching. It also requires the robot to drop and retrieve its target without cost - an assumption that is not necessarily true in domains such as toxic cleanup.

Other foraging coordination algorithms exist that require advance knowledge of physical details of the operating domain and/or use groups of heterogeneous robots. Examples of these algorithms include the territorial allocation method developed by Fontan and Matarić [26] and the territorial arbitration scheme in Goldberg and Matarić [9]. Both methods limit each foraging robot to a specific area or zone and thus prevent collisions. Thus, these methods assume that improved performance can be achieved by specializing the robots to operate only within portions of the field. Another group of algorithms preassigns values so that certain robots inherently have a greater priority to resources than others. This group of coordination methods is similar to the Aggression method mentioned [29], but it preassigns robots to be aggressive or meek. The fixed hierarchy system within Vaughan et al. [29] and the caste arbitration algorithm in Goldberg and Matarić [9] implemented variations of this idea on foraging robots.

Other variations of these coordination methods exist within other domains. For example, Jäger and Nebel [12] presented an algorithm that can dynamically create limiting areas of operation for robots in a vacuuming domain, but require the robots to communicate locally. Within the robotic soccer domain, various groups have been created that rely on allocating each group member to a role. Communication is then needed to allocate and reallocate these roles. One example of this idea is within Marsella et. al. [18].

Because the first group of algorithms require no communication, they seem more suitable to scale to larger groups of robots. As they do not require prior knowledge of the domain, they seem better suited for working with unknown or dynamic environments. More generally, a survey work done by Kraus [16] presented various multi-agent coordination schemes and states that those requiring large overheads are typically unable to scale beyond small groups. Similarly, Jones and Mataric [15] point out that *minimal* robots, or those with low requirements for communication or sensor input from teammates are more suited to scale to large swarms of robots. Minimalistic methods have been used in collection tasks [10] and formation control [8].

To date, only limited work exists on improving robot group scalability. The work by Fontan and Matarić [26] found that groups of 3 robots performed best within their foraging domain. Adding more robots only hurt performance when using their territorial coordination method. Jäger and Nebel [13] presented a collision avoidance technique for use in trajectory planning among robot groups that requires local communication. They noted that their coordination method will not scale beyond groups of 4 robots. Rybski et al. [23] found increasing marginal productivity up to groups of 5 foraging robots, but did not study larger sizes.

Within the general agent community, Shehory et al. [27] presented a scalable algorithm for a package delivery domain suitable for groups of thousands of agents. He based his algorithm on concepts borrowed from physics. Later work by Sander et al. [24] studied how computational geometry techniques could be applied to groups in the same domain. Both found that group productivity did scale marginally with the addition of agents and that a point existed where adding agents did not significantly improve the productivity of their system. Their agents did not compete over physical space, and they never found that adding agents hurt group performance. Specific to the search domain, work by Felner et al. [7] studied the scalability qualities of their PHA* algorithm, and found that their algorithm yields marginally better results with the addition of agents. Our research goal is to understand when robotic teams would similarly scale.

The Law of Marginal Returns, also often called the Law of Diminishing Returns, is well entrenched as a central theory within economics. Most economic domains have spatial limitations and other finite production resources. These limiting factors cause the groups' performance to typically increase marginally with the addition of labor. Brue [4] demonstrated that economists from the Enlightenment Period until modern times often did not provide empirical evidence for their theories. He concluded, ”more empirical investigation is needed on whether this law is operational” within new domains, and ”conjectures by 19th century economists about input and outputs ... simply won't do!” The first goal of this paper was to provide this robust study for robotic groups.

3 Comparing Group Coordination Methods

In this section we present our initial study of scalability within groups of foraging robots. In order to minimize the factors involved in this experiment, we limited our study to groups of homogeneous robots without communication where only the coordination methods differed between groups. We were surprised to find that the coordination method strongly impacted the scalability qualities of the group. While every group demonstrated diminishing positive marginal gains up to a certain group size, the shape of this graph varied greatly between groups.

3.1 Initial Experiment Setup

We implemented a total of eight coordination methods for use on foraging robots. The Noise, Bucket Brigade and Aggression methods were based on previously published methods described in the previous section. Our implementation for the *Noise* team was included as the default team in the Teambots distribution [3]. The *Bucket Brigade* coordination behavior was initiated once a robot detected a teammate within 2 robot radii. Then, these robots would drop the target being carried, move backwards for 25 cycles, and finally revert to the random walk behavior. The *Aggression* group was based on the random function of aggressive behaviors described in Vaughan et al. [29]. For every cycle a robot found themselves within 2 robot radii of a teammate, it selected either an aggressive or timid behavior. In order to decide, we had each robot choose a random number between 1 and 100. If the random number was lower than fifty, it became timid and back away for 100 cycles. Otherwise it proceeded forward, mimicking the aggressive behavior. As all robots within two radii choose whether to continue being aggressive every cycle, one or both of the robots assumed the timid behavior before a collision occurred.

Our remaining five methods were based on variations of existing methods. Similar to the Aggression group, the *Repel Fix* group backtracked for 100 cycles but mutually repelled like the Noise group. The *Repel Rand* group moved backwards for a random interval uniform over 1 – 200 and also mutually repelled. The *Gothru* and *Stuck* groups both removed all coordination behaviors. The *Gothru* group was allowed to ignore all obstacles, and as such spent no time engaged in coordination behaviors. This ”robot” could only exist in simulation as it simply passes through obstacles such as other robots. However, this group was still not allowed to exit the boundaries of the field. We used this group to benchmark ideal performance without productivity lost because of teammates. At the other extreme, the *Stuck* group also contained no coordination behaviors but simulated a real robot. As such, this group was likely to become stuck when another robot blocked its path. Like the Stuck group, the *Timeout* group contained no repulsion vector to prevent collisions. However, these robots did add noise to the direction vector after a certain threshold had been exceeded where their position did not significantly change. The *Timeout* group moved with a random walk for 150 cycles once these robots did not significantly move for 100 cycles. If the timeout threshold was set too low, the robot may consider itself inactive while approaching a target or its home base. However, if this value was set too high, it did not successfully resolve possible collisions in a timely fashion. We experimented with various values until we found that this combination seemed to work well.

We used a well-tested robotic simulator, Teambots [3], to collect data on groups of these foraging robots. We strongly preferred using a simulator as it allowed us the ability to perform thousands of trials of various team sizes and compositions. The sheer volume of this data allowed us to make statistical conclusions that would be hard to duplicate with manually setup trials of physical robots. Using a simulator also allows us to research behaviors, such as Gothru's, that cannot exist with physical robots.

In this experiment, Teambots [3] simulated the activity of groups of Nomad N150 robots. The field measured approximately 5 by 5 meters. Our implementation of foraging followed Balch's [2] multi-foraging task in which the robots attempt to retrieve two or more types of objects. There were a total of 40 such target pucks, 20 of which where stationary within the search area, and 20 moved randomly. Each trial measured how many pucks were delivered by groups of 1 – 30 robots within 9 minutes. For statistical significance, we averaged the results of 100 trials with the robots being placed at random initial positions for each run. Thus, this experiment simulated a total of 24,000 trials of 9 minute intervals.

The simulated robots we studied were based on the same behaviors. The only software differences between the robots lay within their implementation of the previously described teamwork coordination behaviors. Each robot had three common behaviors: wander, acquire, and deliver. In the *wander* phase, the robots originated from a random initial position, and proceeded in a random walk until they detected a resource targeted for collection. This triggered the *acquire* behavior. While performing this second behavior, the robots prepared to collect the puck by slowing down and opening up their grippers to take the item. Assuming they successfully took hold of the object, the deliver behavior was triggered. At times the puck moved, or was moved by another robot, before the robot was able to take it. Once this target resource moved out of sensor range, the robot reverted once again to the wander behavior. The *deliver* behavior consisted of taking the target resource to the goal location which was in the center of the field.

3.2 Initial Results

Figure 1 graphically represents the results from this experiment. Our X-axis represents the various group sizes ranging from 1 to 30 robots. The Y-axis depicts the corresponding average number of pucks the group collected over its 100 trials.

According to the economic Law of Marginal Returns, marginal returns will be achieved when one or more items of production are held in fixed supply while the quantity of homogeneous labor increases. In this domain, the fixed number of pucks acted as this limiting factor of production. Consequently, one would expect to find production graphs consistent with marginal returns. However, only the Gothru group demonstrated this quality over the full range of group sizes. All other groups contained a critical point (CP1) where maximal productivity was reached. After the

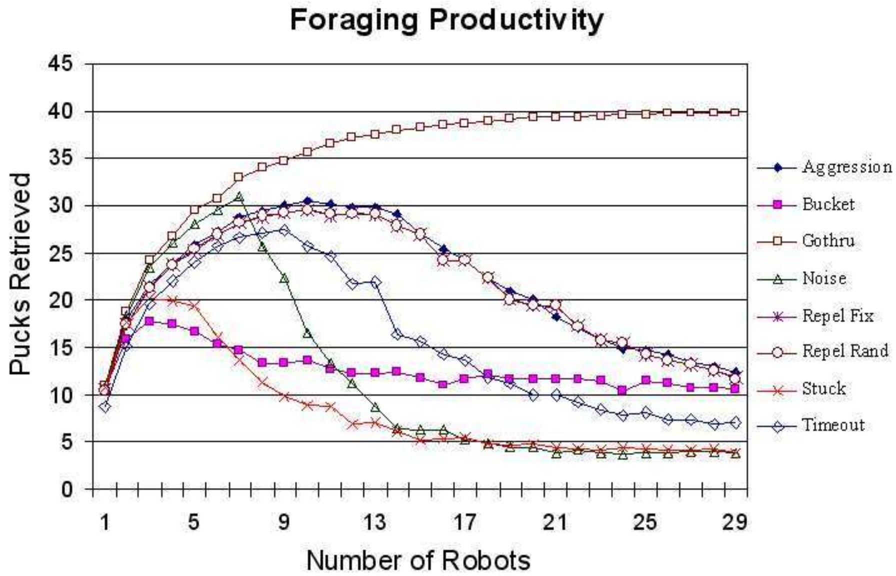

Fig. 1. Comparing Foraging Productivity Results during Group Size Scale-Up

group size exceeded this point, productivity often dropped precipitously. Eventually, the groups reached a level (CP2) where the addition of more robots ceased to significantly negatively effect the groups' performance.

With the exception of the Aggression, Repel Fix, and Repel Rand groups, all groups' productivity graphs differed significantly. For example, the Stuck group reached its CP1 point with an average of only 20.94 pucks collected with groups of 3 robots. The Aggression group reached a maximum of 30.84 pucks collected in groups of 10 robots. Even among equally sized groups, the differences were large. When comparing foraging groups of 10 robots, the Stuck group gathered only 8.58 pucks - far fewer than Gothru's 35.62 pucks, while the Aggression group collected 30.52 pucks, only 5.2 fewer than Gothru. Large differences between the level of CP2 also existed between groups. Notice how the Bucket Brigade group maintained a CP2 level near 12 pucks, while the Stuck and Noise group's CP2 level was near 4 pucks. The Bucket Brigade mechanism was more effective even in large group sizes.

Our resulting research was motivated by these results. The Gothru group was the only group capable of realizing marginal gains throughout the entire range of 30 robots. However, many groups demonstrated the positive quality of maintaining increasing productivity over a larger range of robots. For example, the Noise group only kept marginal gains until groups of seven robots, while the Aggression group kept this quality through groups of 10 robots. We also found that the positive qualities of improved performance and maintaining marginal performance over larger groups are not always synonymous. The Noise group kept positive marginal performance over a smaller range than the Aggression group, yet performed better in groups sized seven or less. A closer look at the various coordination models was needed to draw lessons about how to create groups with both properties.

4 Why does Performance Drop?

We needed a mechanism for understanding why certain coordination methods were more effective than others during size scale-up. We posited that differences among robotic groups were often sparked from clashes in spatial constraints. Specific to foraging, conflicts arose over which robot in the group had the right to go to the home base first. As the group size grew, this problem became more common. This caused the groups to deviate from the ideal marginal productivity, depicted by the Gothru group, by greater amounts. The length of time robots clashed with their teammates because of joint resources, such as the home base location, served as the basis in comparing coordination models within any domain.

Previous work by Goldberg and Matarić [9] found a connection between the level of interference a group demonstrated and its corresponding performance. They defined interference as the length of time robots collide, and we began by using this definition to equate between our coordination algorithms. This measure sufficed for some robots, such as those simulated by the Stuck group, because they did not engage in any other coordination behaviors. However, this metric of interference could not explain the differences between all groups. Many robots, such as those simulated by the Aggression group, never collided. If one takes the position that only collisions constitute interference within robotic groups, these robots do not interfere. Yet we clearly observed how the addition of robots detracted from the groups' productivity after its maximal productivity point.

In this section we present our measure of interference. We describe scale up experiments in foraging and search domains that are characterized by resources that lend themselves to group conflicts. We find that interference and productivity are strongly negatively correlated in such domains, and use this metric to explain differences in productivity between all teams. We posit that in the absence of spatial conflicts, all teams should consistently demonstrate marginal gains during scale up. We confirm this idea by easing the ”space crunch” in our domains and notice how all groups consistently demonstrate marginal returns. We conclude that any domain with group spatial conflicts will suffer from deviations in marginal performance once the causes of interference cannot be resolved.

4.1 Interference: Measure of Coordination

We define interference as the length of time an agent is involved with, either physically or computationally, projected collisions, real or imaginary, from other robots and obstacles. This period of involvement often extends well beyond the actual collision between two robots. Any time spent before a supposed collision in replanning and avoidance activities must also be recorded. Similarly, all post-collision resolution activity must be included as well. Thus, according to our definition, the Gothru group has zero interference because it never engages in any interference resolution behaviors and represents idealized group performance. The Aggression group engages in interference resolution behaviors before a collision ever happens. Its various timid and aggressive behaviors to avoid collisions all constitute interference by our definition. The Bucket Brigade group demonstrates that interference can exist after a collision is prevented. For this group, one needs to measure the productivity lost by handing off the resource from one robot to the next. Many times this group lost productivity during this process because the second robot never properly took the dropped target. Only this measure takes into the account the total interference resolution process.

According to our hypothesis, we expected to see a negative correlation between levels of interference and productivity in three respects. We reasoned that the degree to which a group deviates from the idealized marginal gains is proportional to the amount of average interference within the group. This can impact where the group hits maximal performance. Those groups which reached CP1 with a small number of robots spiked high levels of interference much faster than those where this point was delayed. Second, even before groups hit their maximum productivity point, we hypothesized that the more productive groups have lower levels of interference than their peers. Finally, we expected that differences in where the productivity of the groups eventually plateau can be attributed to the group's saturation level of interference. Those robots that more effectively deal with interference even in large groups will have CP2 values at higher levels.

In order to confirm this hypothesis, we reran our eight foraging groups and logged their interference levels according to our definition. The Gothru group never registered any interference. For all remaining groups, we used the simulator to measure the number of cycles the robots in the groups collided. For all groups other than the Stuck and Gothru groups, we additionally measured the number of cycles the robots triggered interference resolution behaviors when they were not colliding. In the Noise and repulsion groups, this represented the number of cycles spent in repelling activities. In the Aggression group, it was the number of cycles spent in timid and aggressive behaviors. In the Timeout group, this was the cycles spent trying to resolve a collision once the robot timed out. In the Bucket Brigade group, internal behaviors alone did not suffice to measure interference by our definition. We only recorded cycles spent when the robots came close to another and consequently dropped the resource they were carrying. However, we could not measure the time lost when the second robot did not effectively take that resource as we did not have omnipotent knowledge of such events. As a result, our measurement for interference for this group did not necessarily represent an exact measurement, but an underestimate.

Figure 2 represents the result from this trial. The X-axis once again represents the group size, and the Y-axis represents the average number of interference cycles that each robot within the group registered over the 100 trials.

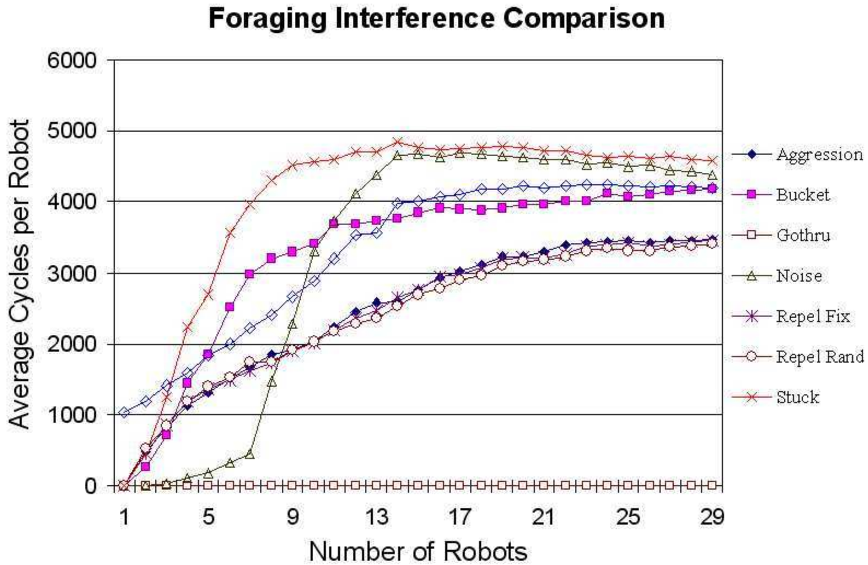

Fig. 2. Interference Levels in Foraging Domain

We found that CP1 typically occurred for all groups when the average interference level within each robot of the group reached a level between 1500 and 2500 cycles. The longer the group was able to maintain classically diminishing returns, the more cycles of interference were needed to cause the critical point. This is because CP1 will only be reached once the productivity lost due to interference is larger than the total marginal productivity of the group. Before this point, the total production of the group increases, albeit marginally. For example, the Stuck group, which reached its critical point with only four robots, needed closer to only 1500 cycles to cause this critical point. The Aggression group hit CP1 with 10 robots, and consequently needed approximately 2200 cycles to counter the productivity of more robots.

Even when viewing the differences between productivity among equally sized groups, interference differences were significant. We found a very strong average negative correlation of -0.94 between the groups' performance and their interference level over the entire range of 1 to 30 robots. For example, the Noise group most closely followed the idealized Gothru productivity graph for groups up until 7 robots, and registered significantly less interference than the other groups. This interference resolution mechanism had little overhead, and needed fewer cycles to resolve a possible collision. However, this method didn't scale well beyond this point. When the group size became larger than seven, its interference levels grew exponentially and the group's performance quickly decayed. In contrast, the Aggression and other repelling groups had significant levels of interference from the onset, but interference levels only grew linearly with respect to the group size. As a result, this group proved more effective with larger group sizes.

We also found that the eventual performance plateau (CP2) was strongly correlated with interference. Some groups leveled off at significantly smaller interference levels than other groups. For example, even in group sizes above 20 robots, the Bucket Brigade group registered an average interference level of 400 fewer cycles less than the Stuck group. Consequently, it collected on average over 5 pucks more than this group at this level.

As one would expect, most groups performed equally well with one robot, as coordination behaviors should only be triggered in groups of two robots or more. The one exception was the Timeout group which collected on average 8.7 pucks with one robot, or about 2 pucks fewer than the other groups. As we defined interference as the time spend on resolving collisions, or even perceived collisions, such a result is quite plausible. At times these robots timed out while slowing down to pick up a puck or avoid an obstacle even by themselves. As we defined such internal reasoning as interference, these robots interfered with themselves in the amount of about 1000 average cycles per trial.

Two of our groups have slight underestimates for interference; however, this did not change our overall results. As previously mentioned, the Bucket Brigade group interfered if a second robot did not successfully receive the resource handed off to it. We found that this did occur at times when there were relatively small groups of these robots. Thus, the correlation between their productivity and that of other groups' among groups of 2–6 robots dropped to -0.80. By discounting this range, the average overall correlation reached almost -0.97. However, after 6 robots we found that there were enough robots in the area to ensure a second robot would quickly take the resource, and the amount of this underestimate was less significant. The Noise group also registered an underestimate for interference. These robots actually used two repulsion fields for collision resolution. They triggered a strong repulsion field when they sensed another robot or obstacle 0.1 meters away. We only measured the number of times this repulsion field was triggered. However, a second, much weaker repulsion field was triggered from 1.5 meters away. In this instance, our underestimate did not seem to significantly statistically detract from our results. With or without the data from this group, the average correlation between groups was -0.94.

4.2 Competing over Spatial Resources

We proceeded to study if our results were limited to foraging or were a general phenomenon seen when robotic groups are faced with restriction production resources. We created a new spatially limited search domain where the task goal was to find the exit out of the room as quickly as possible. We placed groups of robots within a room of 1.5 by 1.5 meters with one exit 0.6 meters wide. We reasoned a critical productivity point would once again form in this domain. Too few robots would result in a long search time until the exit was found. However, too many robots would cause interference as the exit was only physically wide enough for one robot.

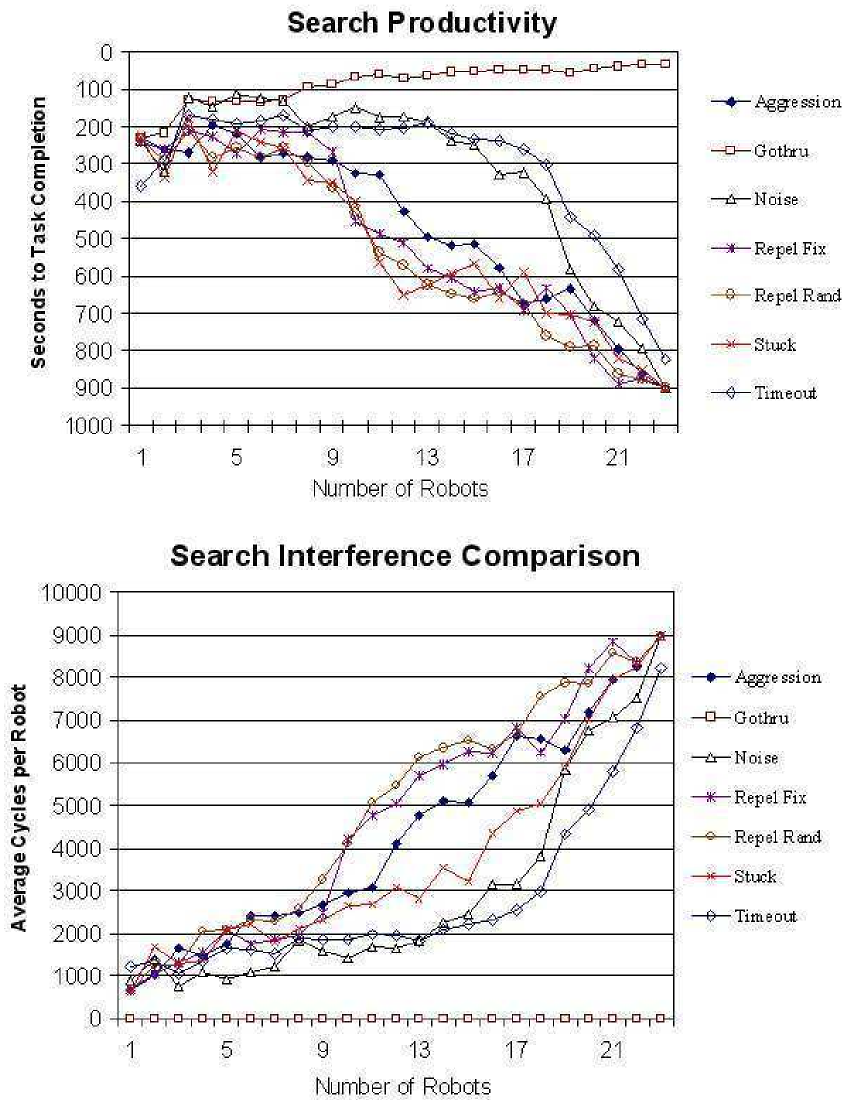

Fig. 3. Search Time and Interference Measurements during Group Size Scale-Up

We ran simulated trials of seven of our eight foraging groups ranging in sizes from 1 - 23 robots (the room holds 23 robots) and averaged the results from 100 trials for statistical significance. We omitted the Bucket Brigade group as this coordination method was not relevant to this domain. We then measured the length of time it took the first robot from each group to completely exit the room. We ended the trial at that point and recorded the time elapsed. Thus, this experiment constitutes over 16,000 trials of variable length.

Figure 3 presents our productivity graphs and corresponding interference levels from this experiment. The X-axis in both graphs depict the size of our groups. In the upper section, we flipped the Y-axis to represent the search t ime of zero as the highest point. As in our foraging graphs, we represent better performance as higher values in this graph. In the lower graph the Y-axis represents our average measurement of interference per robot in the group.

We found that the time to complete the search task was strongly negatively correlated in our new domain as well. We observed that with the exception of the Gothru group, all groups ceased to demonstrate marginal returns at some point. In the Repel Fix group this point occurred with only 5 robots, while the Noise group reached this point with 10. The Noise group had the lowest level of interference through groups of 13 robots, and was able to most closely approximate Gothru's performance until this group size. After this point the Timeout group fared the best. We found that certain interference resolution mechanisms work best in specific domains. While the repulsion methods were quite effective in foraging, the interference levels in these groups grew exponentially in this domain. Overall, the average statistical correlation for groups of 1-23 robots between the time elapsed to exit the room and their corresponding interference level was -0.94.

4.3 Easing Spatial Restrictions

According to our hypothesis, deviations of productivity in robot groups are strongly correlated with interference. Once our foraging and search groups ceased to effectively resolve interference they reached their critical group sizes. Adding more robots only hurt the groups' performance. We posit that the physical space limitations existent within many robotic groups often cause this interference. The one home base area within the foraging domain and the one exit within the search domain create a competition over space between robots that cannot always be properly resolved.

We were able to confirm that our robotic groups always demonstrated marginal returns once restrictions over physical space were eased. We changed the foraging group requirement of returning the pucks to one centralized home base location. Instead, they were allowed to consider the pucks to be in the home base immediately. With the exception of the Bucket Brigade group, we reused all 8 previously studied foraging groups. Once again, we omitted this method because it was not applicable to our new domain. We left all other environmental factors such as the number of trials, the size and shape of the field and the targets to be delivered identical. Thus, Teambots [3] simulated 21,000 trials of 9 minute intervals in this experiment.

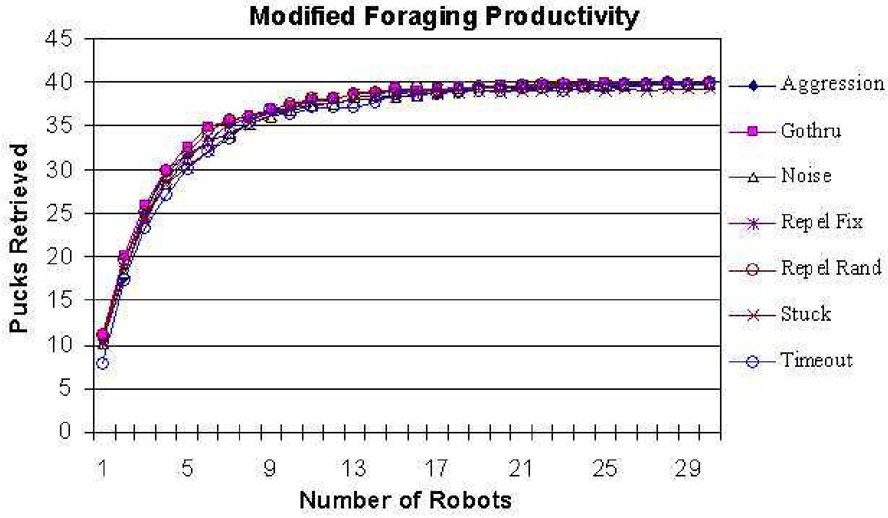

Fig. 4. Productivity of Groups in Modified Foraging Domain during Size Scale-Up

As figure 4 shows, all groups did indeed always achieve marginal returns in the modified foraging domain. While Gothru still performed the best, the differences between it and other groups' coordination methods were not as pronounced. The level of interference all groups demonstrated was also minimal, and thus not displayed. We concluded that not every foraging domain needed to have a critical point for productivity where marginal gains during scale up ceased.

Within the search domain, we hypothesized that limitations in the room size and width of the exits created the large amounts of interference during scale up. In order to ease this restriction, we doubled the size of the room to become approximately 3 by 3 meters, and widened the exit to allow free passage out of the room by more than one robot. Once again, we measured the time elapsed (in seconds) until the first robot left the room and averaged 100 trials for each point. This experiment also constituted over 16,000 trials of varying lengths. Figure 5 graphically shows that our modified domain consistently realized marginal increases in faster search times with respect to group size. Once again, interference levels were also negligible in our new domain. Thus, we concluded that achieving marginal productivity gains was always possible once competition over spatial resources was removed.

5 Improved Scalability through Coordination Combination

Our next step was to apply lessons based on our understanding of the coordination methods we studied towards creating methods with improved productivity and scalability properties. In this section we present our *Composite Coordination Methods*. We found that it was possible to combine methods with different scalability prop-

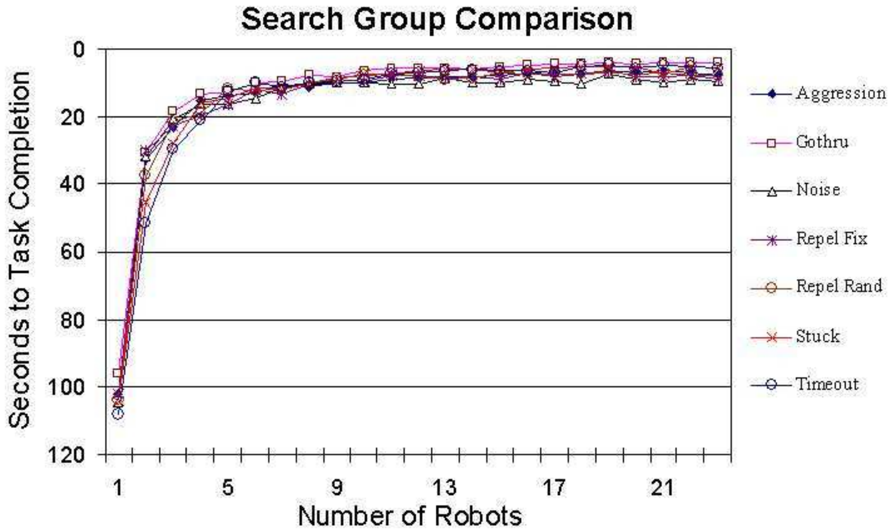

Fig. 5. Productivity of Groups in Modified Search Domain during Size Scale-Up

erties to create a new composite method. This method achieved higher productivity levels in the foraging and search domains we studied. Surprisingly, we found that our new composite method at times far exceeded the productivity levels of even the that highest levels of productivity from the groups they were based on. We believe that using multiple methods in tandem allowed robots to more effectively deal with the spatial limitations that characterized their operating domain. This allowed for the gains we found in these groups' scalability properties.

5.1 Composite Coordination Methods

Our composite coordination methods combined the two best coordination methods for any given domain. Our previous study demonstrated that it possible to order coordination methods based on groups sizes where they are most effective. In the foraging domain, the Noise group had the highest productivity in small groups, while the Aggression group had higher productivity in larger groups. In the search domain, the Noise group again had the highest productivity in the small groups with the Timeout group faring better in larger group sizes. In both domains, our implementation for the composite method was based on allowing these two simpler methods to be triggered under different domain conditions.

Our implementation of the composite method in the foraging domain revolved around using two different methods to attempt to prevent collisions. Robots first used the Noise method, but if this method proved insufficient opted for the more robust Aggression method. Once a robot detected that another teammate came within two robot radii away, it attempted to resolve a possible collision by inserting a slight repulsion and noise element into its trajectory. In cases when the probability of collisions was low, as was the case in small group sizes, this behavior alone sufficed. However, at times the spatial conflicts in the domain could no t be resolved through this simple coordination behavior. For example, in large group sizes, the probability that two or more robots mutually blocked became substantial. In these cases, the robots continued to move closer despite the use of this method. Once the robots came within a second, closer threshold, which we set to one robot radii, the second, more robust Aggression method was triggered. The timid and aggressive behaviors in this method were more successful in resolving spatial conflicts t han the simpler behaviors in the Noise method. However, the interference overhead in the Aggression behavior was higher, and not justified in situations where the simpler behavior sufficed. Thus, by two different thresholds we attempted to match the correct collision prevention behavior to the domain conditions.

We found this approach to be very effective within our foraging domain. Figure 6 displays the productivity of the composite foraging group, Noise + Aggression, compared to the two methods it is based on. In the top portion of the graph we display the average number of pucks retrieved (Y-axis) over different group sizes (X-axis). The bottom graph displays the varying interference levels (Y-axis) as a function of the group size. Notice how the composite group significantly outperformed the two groups it was based on. We performed the two-tailed t-test between our composite group and the two static ones it was based on. Both p-scores were well below 0.05 needed to establish the statistical significance, with the higher score of 0.003 found between the Aggression group and the composite one. We also found that the relationship between interference and productivity applies to this new group with a strong negative correlation of -0.92 between all three group's productivity and the corresponding interference level averaged over the interval of 1 – 30 robots.

Our motivation in the search domain was similar, but our composite coordination method was implemented slightly differently. In this domain we also created our composite method between two methods – Noise and Timeout. These two methods resolve collisions with different mechanisms. The Noise method attempts to prevent collisions before they occur through repulsion. In contrast, the Timeout behavior was purely reactive in nature and its behavior only was triggered after collisions already occurred. Thus, a composite coordination method between these two methods was able to created without two different distance thresholds. The Noise method behavior was fully implemented to attempt to prevent collisions. The Timeout behavior was also fully implemented. In cases when the Noise behavior did not prevent a collision, this second behavior was effective in then resolving the conflict.

We also found that this approach yielded marked improvement in performance and scalability properties for our search domain. Figure 7 displays the productivity of the composite foraging group, Noise + Timeout, compared to the two methods it is based on. In the top portion of the graph we display the average time to complete the search task (Y-axis) over the different group sizes (X-axis). The bottom graph displays the varying interference levels (Y-axis) as a function of the group size. Notice how the composite group again significantly outperformed the two groups it was based on, especially in larger group sizes. We performed the two-tailed t-test between

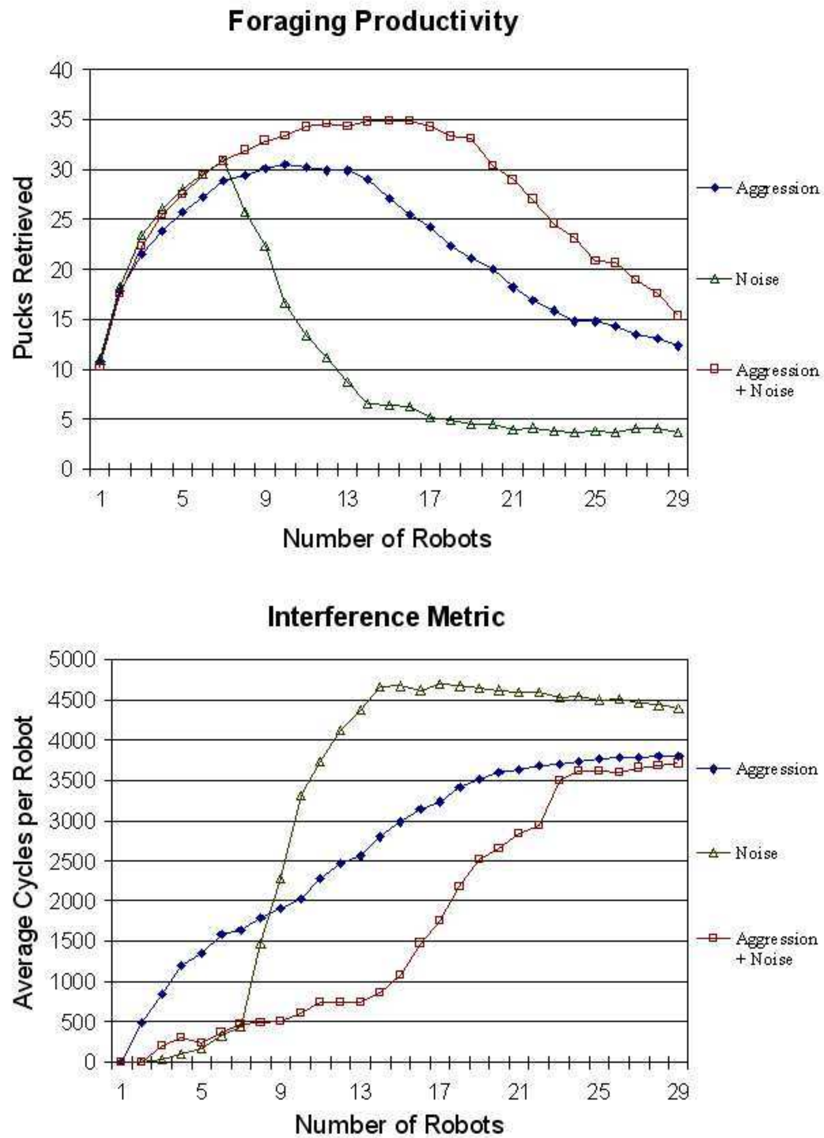

Fig. 6. Comparing a Composite Foraging Method to its Two Base Methods

our composite group and the two static ones it was based on. Both p-scores were well below 0.05 needed to establish the statistical significance, with the higher score of 0.004 found between the Noise group and the composite one. We also confirmed that the relationship between interference and productivity applies to this new group with a strong negative correlation of -0.98 between the three groups' productivity levels and their corresponding interference levels over the interval of 1 – 23 robots. It is important to note that the composite method in the search domain was able to eliminate the critical group size that existed in every group we studied except for the theoretical Gothru group. As such, this group demonstrated the best scalability quality from all methods we studied – the group's average productivity never significantly dropped with the addition of robots. Further study was needed to understand why these composite groups had significantly better productivity and scalability qualities than the methods they were based on.

5.2 Studying How to Improve Scalability

Our interference metric was useful for understanding why the composite methods we created were able to significantly outperform the simpler methods they were based on. These composite methods had significantly lower levels of interference, allowing marginal gains and larger productivity over larger groups. However, we believe that coordination methods can be developed to improve the scalability capabilities of robots. It is possible that our interference metric is not only useful post-facto, but can facilitate online adaptation to improve performance even in dynamic and changing environments. We have begun to study how to create adaptive methods based on interference and have presented our initial results in [22].

We believe coordination methods that respond to the triggers of interference can minimize the time spent resolving those instances. Throughout the course of one trial, many spatial conflicts are likely to occur. The speed w ith which the robots resolve these conflicts will determine the success of the rob ots to achieve higher productivity and scalability properties. As such, we posit that a causal relationship exists between a robot's interference level and the corresponding productivity that robot is able to contribute to its group. The more time spent on resolving coordination conflicts, the less time will be left to perform the desired ac tion. Thus, if robots could reduce their interference levels, they will consequently be able to achieve higher productivity.

Our working hypothesis is that groups that effectively deal with interference episodes are going to improve their productivity levels. While coordination behaviors themselves constitute interference, at times they are needed for achieving cohesive group behavior. Effective behaviors cannot realistically eliminate interference. Optimal coordination methods behaviors can only minimize interference levels given domain conditions. For example, in the foraging domain we studied, the Noise method's simpler coordination method contained little overhead. However, as collisions within the domain became frequent, this method did not suffice, and robots were not capable of successfully resolving space conflicts a nd thus loss productivity. The Aggression method had an overhead that made it more effective in larger group

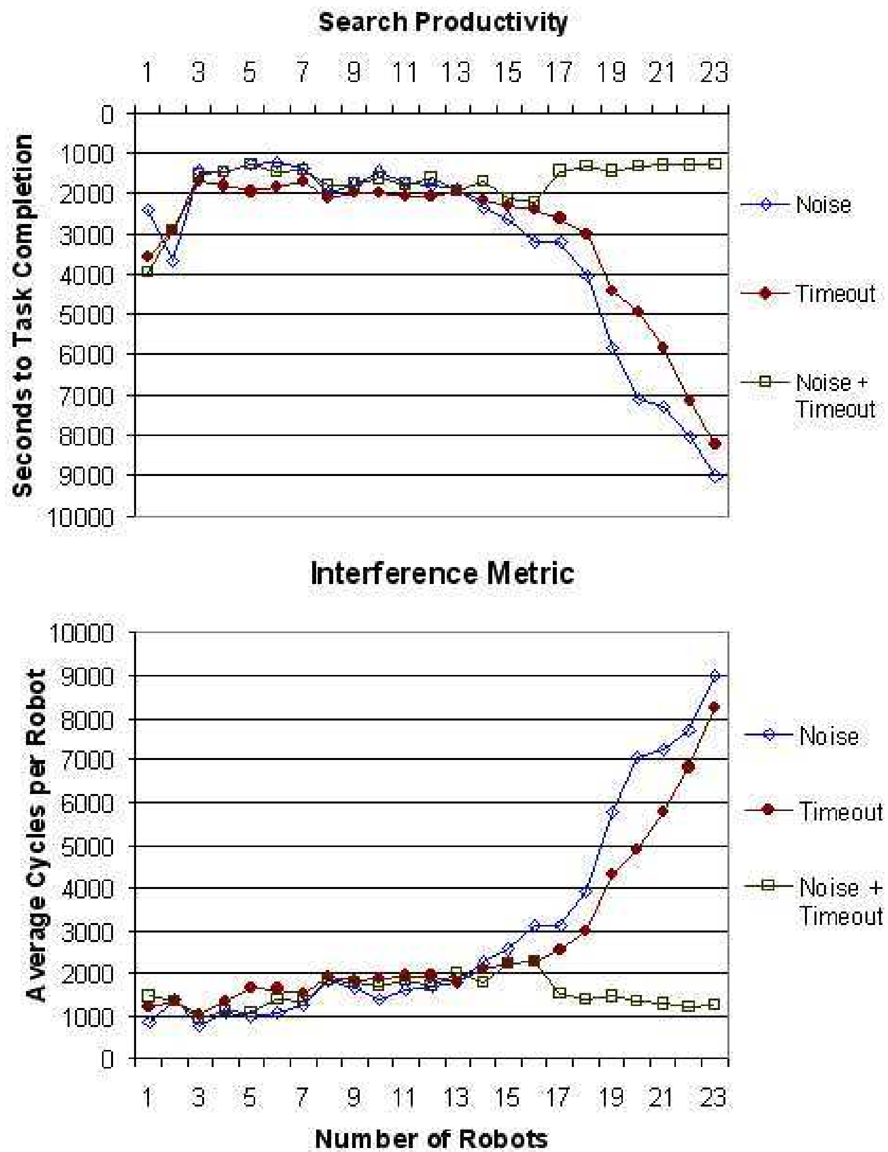

Fig. 7. Comparing a Composite Search Method to its Two Base Methods

sizes, but the larger interference overhead in this method made it less effective in smaller group sizes.

We believe that our composite methods outperformed the static method because of their improved ability to effectively match their coordination efforts to the needs of their domain. This allowed these robots to change the time spent on resolving coordination conficts based on the needs of the domain. Figure 8 dem onstrates the ability of our composite method to resolve conficts in a more timely f ashion. The graph represents the percentage of foraging robots that on average collided throughout the course of three trials (540 simulated seconds) in groups of 20 robots. The X-axis in this graph represents the number of seconds that elapsed in the trial (measured in ten second intervals), while the Y-axis measures the percentage of robots colliding at that time in the Noise, Aggression, and Noise + Aggression methods. Notice that the Noise group was ineffective in resolving collision instances in this group size and thus throughout the trial nearly all robots were exclusively engaged in collision resolution behaviors. As a result, this group had the highest interference levels and the poorest productivity. The Aggression group was able to more effectively deal with collisions, but on average consistently spent more than half of their time resolving spatial conficts. In contrast, robots in the composite grou p were able, on average, to resolve conficts and thus reduce their interference levels . This resulted in the significantly higher productivity levels in this group over the two static ones it was based upon.

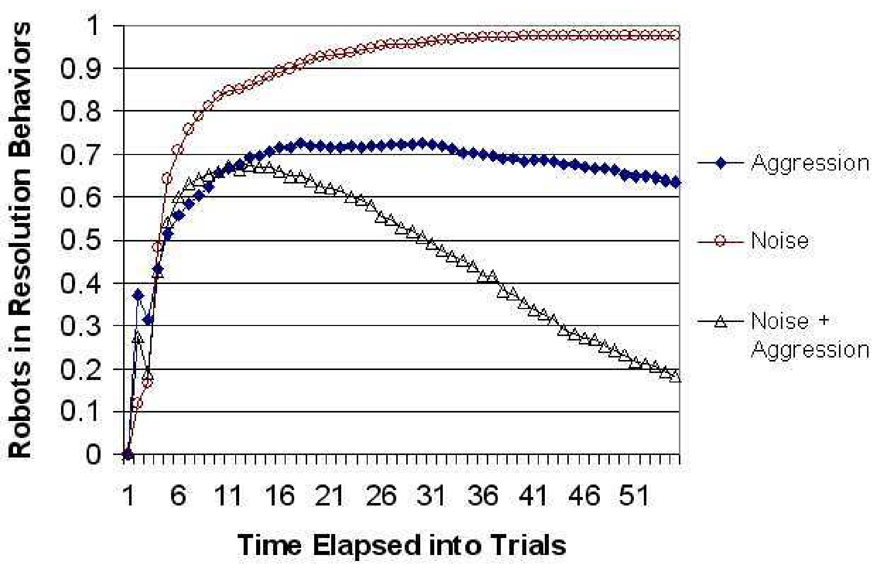

Fig. 8. Average Percentage of Robots Colliding as a Function of Time

When viewing spatial conflicts on a per trial basis, the fluctu ations in the instances of interference and the robot's ability to react to those fluctuations are even more pronounced. We posit that the composite method used the Aggression method in reaction to collisions becoming more frequent within the domain. To support this claim we viewed the internal state of these robots over the course of our trials. Figure 9 displays three individual foraging trials of the composite group, again in groups of 20 robots. In the upper graph we mapped the percentage of robots that were engaged in resolution behaviors (Y-axis) over the course of the trials (the Y-axis). The bottom graph represents the internal coordination state of these robots as a number between 1 and 2. A value of 1 represents all robots being engaged in the Noise behavior, and a value of 2 corresponds to all robots in the Aggression behavior. Groups on average typically have a value between these extremes with robots autonomously choosing different states based on how close its closest teammate is at that moment. Notice the relationship between these two graphs with the composite robots using the Aggression behavior (an average state closer to 2) when collisions are more frequent. On average over the entire time period, we found a strong negative correlation of -0.90 between these two graphs. This supports our claim that changes in interference can be sensed autonomously by robots. We believe this allowed the composite groups to achieve such a strong improvement in the productivity and scalability qualities of these teams.

6 Conclusion and Future Work

In this paper we presented a comprehensive study on the productivity of robotic groups during scale-up. As the size of robotic groups increased, effective coordination methods were critical towards achieving effective team productivity. The limited space inherent in many environments, such as the foraging and search domains we studied, makes this task difficult. Using our novel, non-domain specific definition of interference, we were able to equate between the effectiveness of various existing coordination algorithms. Our interference metric measured the total time these robots dealt with resolving team conflicts and found a strong negati ve correlation between this metric and the corresponding productivity of that group. Groups demonstrated marginal gains only when their interference level was low. If the new robot added too much interference into the system, it detracted from the group's productivity and marginal productivity gains would cease. Gains during scale-up would always be achieved if interference was not present. We present our composite coordination methods as an example of how to achieve improved scalability through minimizing interference.

Many robotic domains also contain the limited space and production resources that our foraging and search domains exemplify. We predict robotic groups involved with planetary exploration, waste cleanup, area coverage in vacuuming, and planning collision-free trajectories in restricted spaces will all benefit from use of our interference metric. We plan to implement teams of real robots in these and other domains in the future.

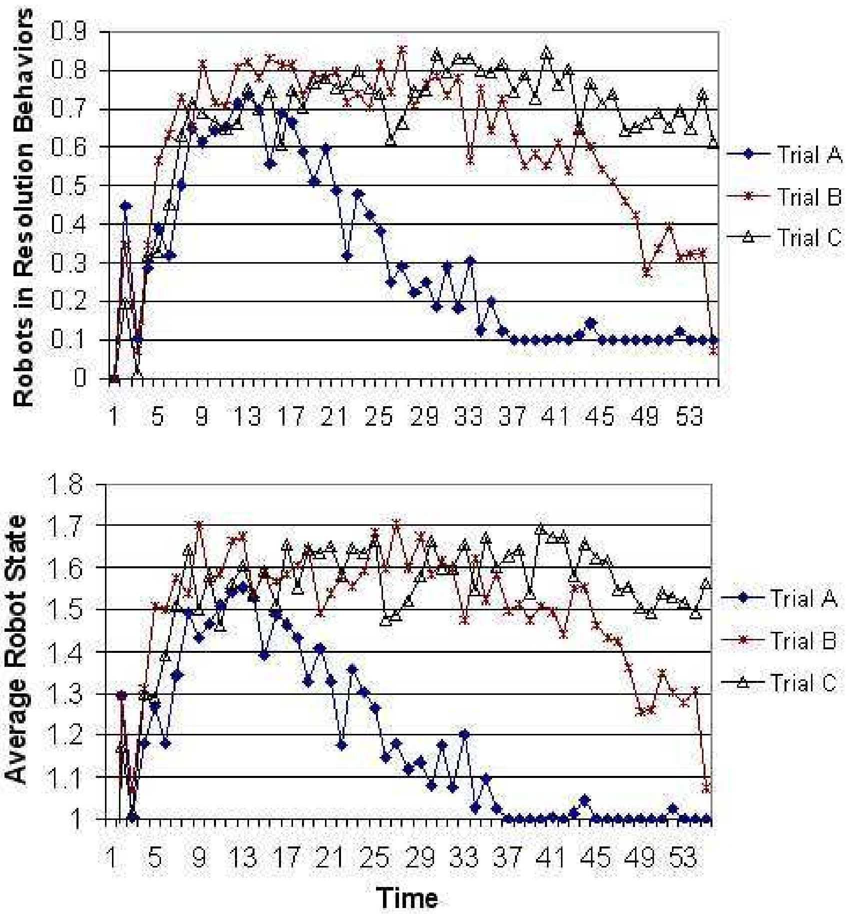

Fig. 9. Fluctuations in Collisions over Time and the Corresponding Foraging Method

We demonstrated in our paper that the spatial restrictions within robotic domains often prevented marginal gains from being realized as group sizes grew. The corollary of this hypothesis is that marginal returns will always be achieved in domains that do not clash over resources. It is not surprising that groups of agents should therefore always realize marginal returns during scale up once group interference issues have be resolved or are not applicable.

Many applications and extensions to our interference metric are possible. For future work, we hope to address several directions for possibly extending our metric. This paper limited its study to homogeneous robots without communication. Additionally, we did not study coordination methods which require pre-knowledge of their domain or algorithms that use other forms of preprocessing. We leave the study of how to widen our metric to allow contrasting robots with differing capabilities such as communication, foreknowledge of domains, and preprocessing requirements for future work. We are hopeful that our interference metric will be useful for a range of applications.

References

1. R.C. Arkin and T. Balch. Cooperative multiagent robotic systems. In *Artificial Intelligence and Mobile Robots*. MIT Press, 1998.
2. T. Balch. Reward and diversity in multirobot foraging, In IJCAI-99 Workshop on Agents Learning About, From and With other Agents, 1999.
3. Tucker Balch. www.teambots.org.
4. Stanley L. Brue. Retrospectives: The law of diminishing returns. *The Journal of Economic Perspectives*, 7(3):185–192, 1993.
5. Phil R. Cohen and Hector J. Levesque. Teamwork. *Nous*, 25(4):487–512, 1991.
6. G. Dudek, M. Jenkin, and E. Milios. A taxonomy for multi-agent robotics. *Robot Teams: From Diversity to Polymorphism, Balch, T. and Parker, L.E., eds., Natick, MA: A K Peters*, 3:3–22, 2002.
7. Ariel Felner, Roni Stern, and Sarit Kraus. PHA*: Performing A* in unknown physical environments. In *AAMAS 2002*, pages 240–247.
8. J. Fredslund and M. Matarić. A general, local algorithm for robot formations. In *IEEE Transactions on Robotics and Automation, special issue on Multi Robot Systems*, pages 837–846, 2002.
9. D. Goldberg and M. Matarić. Interference as a tool for designing and evaluating multi-robot controllers. In *AAAI/IAAI*, pages 637–642, 1997.
10. D. Goldberg and M. Matarić. Design and evaluation of robust behavior-based controllers for distributed multi-robot collection tasks. In *Robot Teams: From Diversity to Polymorphism*, pages 315–344, 2001.
11. Barbara J. Grosz and Sarit Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.
12. M. Jager and B. Nebel. Dynamic decentralized area partitioning for cooperating cleaning robots. In *ICRA 2002*, pages 3577–3582.
13. Markus Jager and Bernhard Nebel. Decentralized collision avoidance, deadlock detection, and deadlock resolution for multiple mobile robots. In *IROS*, 2001.
14. Nicholas R. Jennings. Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. *Artificial Intelligence*, 75(2):195–240, 1995.
15. C. Jones and M. Matarić. Adaptive division of labor in large-scale minimalist multi-robot systems. In *Proceedings of the IEEE/RSJ International Conference on Robotics and Intelligent Systems (IROS)*, pages 1969–1974, 2003.
16. Sarit Kraus. Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, 94(1-2):79–97, 1997.
17. C. Ronald Kube and Hong Zhang. Collective robotic intelligence. pages 460–468, 1992.
18. Stacy Marsella, Jafar Adibi, Yaser Al-Onaizan, Gal A. Kaminka, Ion Muslea, and Milind Tambe. On being a teammate: experiences acquired in the design of RoboCop teams. In Oren Etzioni, Jörg P. Müller, and Jeffrey M. Bradshaw, editors, *Proceedings of the Third International Conference on Autonomous Agents (Agents'99)*, pages 221–227, Seattle, WA, USA, 1999. ACM Press.

19. E. Ostergaard, G. Sukhatme, and M. Matari. Emergent bucket brigading - a simple mechanism for improving performance in multi-robot constrainedspace foraging tasks. In *Proceedings of the 5th International Conference on Autonomous Agents*, 2001.
20. L. Parker. Alliance: An architecture for fault-tolerant multi-robot cooperation. *IEEE Transactions on Robotics and Automation*, 14(2):220–240, 1998.
21. Craig W. Reynolds. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4):25–34, 1987.
22. A. Rosenfeld, G. Kaminka, and S. Kraus. Adaptive robot coordination using interference metrics. In *The Sixteenth European Conference on Artificial Intelligence*, Valencia, Spain, August 2004.
23. P. Rybski, A. Larson, M. Lindahl, and M. Gini. Performance evaluation of multiple robots in a search and retrieval task, In Workshop on Artificial Intelligence and Manufacturing, 1998.
24. Pedro Sander, Denis Peleshcuk, and Barbara Grosz. A scalable, distributed algorithm for efficient task allocation. In *AAMAS 2002*, pages 1191–1198.
25. Paul Scerri, Yang Xu, Elizabeth Liao, Justin Lai, and Katia Sycara. Scaling teamwork to very large teams. In *AAMAS '04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 888–895. IEEE Computer Society, 2004.
26. M. Schneider-Fontan and M. Mataric. A study of territoriality: The role of critical mass in adaptive task division. 1996.
27. O. Shehory, S. Kraus, and O. Yadgar. Emergent cooperative goal-satisfaction in large scale automated-agent systems. *Artificial Intelligence journal*, 110(1):1–55, 1999.
28. M. Tambe. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7:83–124, 1997.
29. R.T. Vaughan, K. Støy, G.S. Sukhatme, and M.J. Matarić. Go ahead, make my day: robot conflict resolution by aggressive competition. In *Proceedings of the 6th int. conf. on the Simulation of Adaptive Behavior, 2000*.