14th ICCRTS: C2 and Agility

Decentralized Command and Control: Self-Organization in a Simple Model for Emergency Response

Collective Endeavors

Michael I. Bell

Alidade Incorporated

31 Bridge Street
Newport, Rhode Island 02840

(301) 871-6247

mike.bell@alidade.net

Decentralized Command and Control: Self-Organization in a Simple Model for Emergency Response

Abstract

Recent discussions of command and control, particularly in connection with interagency or combined (coalition) operations, have emphasized the concept of "unity of effort" at the expense of "unity of command." This recognizes the reality that in many situations unity of command may not be possible, and it assumes that cooperation and coordination can be achieved through common intent and purpose. Implicit in this approach is the assumption that, given shared situational awareness and common goals, forces can "self-organize" to produce behavior that is similar to (or at least as effective as) the result of unified command.

The present work attempts to test this assumption using a very simple computer simulation to represent an interagency task force providing rapid response in an emergency situation where no centralized command exists. In the model, a small number of identical responders are provided with complete information on the status and location of a large number of simultaneous, geographically dispersed incidents requiring attention. The responders operate under a common set of behavior rules, and several types of self-organized (emergent) behavior are observed. It is found that subtle changes in the behavior rules can transform beneficial collective behaviors into ones that are counter-productive.

Keywords: self-organization, decentralization, incident response

Introduction

Military organizations have traditionally demanded unity of command, meaning "that all forces operate under a single commander with the requisite authority to direct all forces employed in pursuit of a common purpose" (Department of Defense 2008). This goal is rarely if ever attained today in operations that require cooperation or support to or from organizations belonging to other government agencies, the private sector, or foreign nations. Recognition of this fact has led to increased emphasis on unity of effort, a closely related but significantly different concept defined as "coordination and cooperation toward common objectives, even if the participants are not necessarily part of the same command or organization" (Department of Defense 2008). Achieving unity of effort is in many ways the more difficult task. Unity of command can be regarded (at the risk of over-simplification) as an organizational issue, to be resolved by establishing suitable structural arrangements. On the other hand, unity of effort is inherently a dynamic property of an activity, subject to change over time and difficult to quantify. Creating unity of effort in the absence of unity of command is and will remain a crucial part of the conduct of interagency and multinational operations. The present study attempts to explore some of the issues surrounding this effort by examining a highly simplified model of a single task associated with emergency or disaster response.

Operational Background

Modern military organizations must engage in many activities that have been described as "complex endeavors," in the sense that the participants may differ widely in organizational type, culture, and objectives and may interact with each other and the operating environment in complex ways (Alberts and Hayes 2007). Examples range from domestic interagency operations (disaster relief or terrorist incident response) that are relatively focused in time and space to multinational civil-military campaigns of much greater extent and duration. Command and control (C2) arrangements for such endeavors, regardless of scope, must deal with the fact that the participants invariably come equipped with their own information sources, chains of command, and operating procedures.

Doctrinal Background

The military doctrine cited above (Department of Defense 2008) recognizes that "during multinational operations and interagency coordination, unity of command may not be possible, but the requirement for unity of effort becomes paramount." The civil agencies participating in the National Incident Management System (NIMS) take a different approach, in which unity of command is not sought, but rather when appropriate a "unified command" may be used:

> In incidents involving multiple jurisdictions, a single jurisdiction with multiagency involvement, or multiple jurisdictions with multiagency involvement, unified command allows agencies with different legal, geographic, and functional authorities and responsibilities to work together effectively *without affecting individual agency authority, responsibility, or accountability* [emphasis added]. (Department of Homeland Security 2008)

The focus on jurisdictional and agency authority is a natural consequence (and legal requirement) of our Federal system. The NIMS and the National Response Framework (formerly National Response Plan), which it supports, "are designed to ensure that local jurisdictions retain command, control, and authority over response activities for their jurisdictional areas." (Department of Homeland Security 2008). Nevertheless, this approach can be expected to make unity of effort more rather than less difficult to attain. The NIMS structure provides substantial guidance with respect to organizational structure and command relationships but cannot be expected to mandate processes or procedures for developing unity of effort. Joint military doctrine is not much better in this regard, although there is a clear recognition of the dynamic aspects of the problem. It "emphasizes unified action — the synchronization, coordination, and/or integration of the activities of governmental and nongovernmental entities with military operations — to achieve unity of effort." (Department of Defense 2008). This formulation leaves us with a "chicken and egg" problem, however, since "synchronization, coordination, and/or integration" can reasonably be regarded as effects of unity of effort rather than causes. Fortunately, several recent developments in the theory of military operations and the study of complex systems offer suggestions as to how these behaviors might be achieved.

A Network Centric Approach

The concept of Network Centric Warfare (NCW) represents a shift from traditional attrition-based warfare to an approach that emphasizes speed of command and self-synchronization (Cebrowski and Garstka 1998). In developing a theoretical basis for NCW, Alberts (2002) described NCW as resting on four tenets:

1. A robustly networked force improves information sharing.
2. Information sharing and collaboration enhance the quality of information and shared situational awareness.
3. *Shared situational awareness enables self-synchronization* [emphasis added].
4. These, in turn, dramatically increase mission effectiveness.

These provide a causal, or at least logical, link between improved information sharing (through networking) and increased effectiveness. The concept of self-synchronization plays a central role in this linkage. Although the term refers explicitly to the organization of behavior with respect to time, the argument applies to the spatial organization of behavior as well, and the more general term self-organization will be used here.

Self-organization is a central concept in the theory of complex adaptive systems, which are dynamic collections of many interacting adaptive agents (Gell-Mann 1995). Such systems can spontaneously exhibit order or behavior that is not easily predicted by examination of the properties of individual agents or their interactions. For the present work, the exact mechanism producing this emergent order or behavior is less important than the fact that it arises internally and is not imposed from outside the system or by design (Bar-Yam 1997). In the context of military operations, self-organization (or self-synchronization) represents an attractive approach to decentralized C2. In complex endeavors, it could provide a means to achieve unity of effort without unity of command.

The Model

To gain some insight into self-organization and its potential application to emergency response, we have developed a highly simplified model that permits us to explore some of the implications of the decentralized C2 and restricted information flow that often characterize interagency or multinational operations.

The model is based on the following scenario. An event has caused a number of simultaneous incidents, randomly distributed over a geographical area. An example might be storm-related electrical power outages. A force of first-responders is available, initially distributed randomly across the operating area. In the example, these might be utility service trucks, perhaps belonging to different companies or jurisdictions. The goal is to provide a response to each of the incidents in the shortest possible time. We impose the constraint that no central planning or C2 is available. In the interest of simplicity, we assume that all responders have the same capabilities and are provided the same information and that all incidents are identical. We further assume that the time required to service an incident is negligible compared to the transit time between incidents. These simplifying assumptions could be replaced by more realistic ones, but the additional complexity would not add significantly to the insight obtained from the model.

A Mathematical Digression

Before describing the computer simulation of this model and its results, it is worth noting that in the case of a single responder the present problem reduces to the well-known traveling salesman problem (TSP). Given a set of cities and the cost of travel between each pair of cities, solving the TSP requires finding the cheapest way to visit each city exactly once. Usually, the cost of travel between two cities is taken to be the same in either direction and may be interpreted as the distance between them. The salesman may be required to return to his starting point without increasing the difficulty of the problem. All known solutions of the TSP require computational effort (time) that grows at least exponentially with the size of the problem (the number of cities, n). The "brute force" approach, examining every possible tour of the cities, takes effort that increases proportional to $(n - 1)!/2$. On the other hand, if the solution is provided, the difficulty of verifying that it is correct grows only as a polynomial in the problem size. In the terminology of computational complexity theory, such problems are nondeterministic polynomial (NP) time problems. If it can be shown that an efficient (polynomial time) solution to a given NP problem leads to an efficient solution of all NP problems, the given problem is termed NP-complete (NPC). The TSP is known to be NPC.

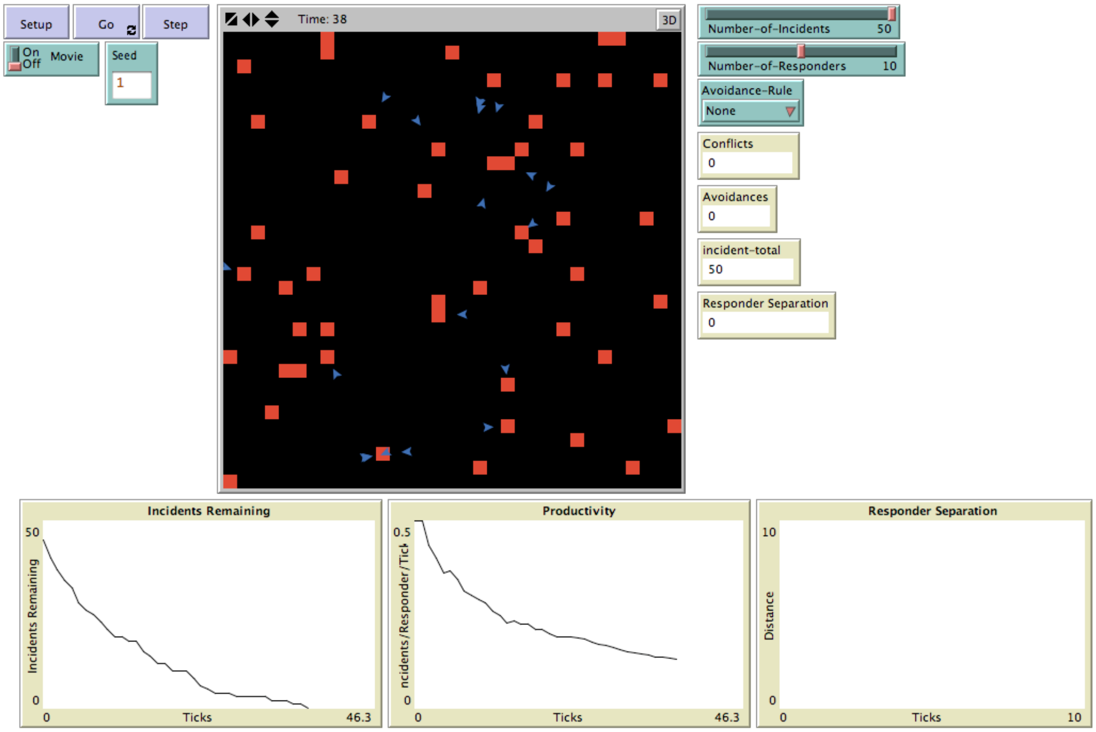
Other classic problems are still more closely related to the present one. These are multiple traveling salesman problems, the most widely studied of which is the group of variants known collectively as the vehicle scheduling problem (VSP). The VSP requires determination of the best (lowest cost) route for multiple vehicles (with differing capabilities) making deliveries to multiple customers (with differing requirements). Typically, the vehicles all begin and end their routes at a common location. The VSP and other multiple traveling salesman problems are known to be NP hard, since every multiple-salesman problem has a single salesman problem as a special case. On the other hand, these problems have not been shown to be NPC, which would require that they could be solved efficiently given an efficient solution to the TSP.

Agent-Based Simulation

We have used the NetLogo agent-based modeling environment (Wilensky 1999) to construct a numerical simulation of the model described above. The operating area contains 1089 potential incident sites, arranged on a 33 x 33 square grid. The baseline simulation starts with 50 incidents placed at random on this grid and ten responders, also distributed at random. The number of incidents and responders can be varied in order to explore the behavior of the model. Each responder is able to move a distance equal to the grid spacing (in any direction) at each time step.

Reflecting the absence of central C2, each responder selects the incident to which it will respond based only on its own information and rule set. For the present work, we have chosen to give each responder a complete and current list of the locations of all the pending incidents but no additional information regarding the activities of the other responders. This can be seen as a crude representation of a situation in which an incident is able to report or broadcast its needs, but no communication is possible between the responders. Given the difficulty of the scheduling problem even when complete information is available, we have programmed the responders to use a "greedy" algorithm, simply moving as quickly as possible toward the nearest incident. In the case of the TSP, the greedy algorithm performs well on the average, but there are pathological cases in which it does very badly, sometimes even finding the longest possible route. Less is known about its performance in multiple traveling salesman problems, but it is reasonable to expect similar behavior.

Figure 1 is a view of the user interface of the NetLogo simulation at the end of a typical model run, providing an indication of the input variables that can be adjusted by the user and the outputs than can be tracked. (Some of these inputs have not yet been mentioned and will be described below.) A history of all the internal variables can be maintained, if necessary. The animation provided by the simulation has more than entertainment value; as we will see, it can permit the user to detect patterns of self-organized behavior that are not readily visible in other outputs. NetLogo also provides a means of automating multiple model runs ("BehaviorSpace"), allowing the user to program many replications of a given run, vary the input parameters, and log the results. The animation is usually disabled in the interest of execution speed during these production runs. Unless otherwise noted, all the results presented represent the average of 1000 replications of the model for a given set of parameters and random initial configurations of the incidents and responders.



Seed 1

Time: 38

Number-of-Incidents 50

Number-of-Responders 10

Avoidance-Rule None

Conflicts 0

Avoidances 0

incident-total 50

Responder Separation 0

Figure 1. NetLogo simulation user interface

Simulation Results

Since the network centric hypothesis would imply a high degree of effectiveness for a force self-organized around a well-defined mission and shared situational awareness, the central question to be addressed is how well the performance of multiple responders compares to that of a single responder, given the information constraints and decision rules we have imposed.

Baseline

For a single responder, the mean time to respond to all 50 incidents is 210.3 time steps. For the baseline case of ten responders, the completion time is 43.4 steps. This performance improvement by a factor of only 4.8 for a ten-fold increase in available effort is not encouraging. Roughly 57% of the effort of the additional responders appears to be wasted, and much of our analysis has been devoted to determining the cause of this behavior and seeking remedies.

Measures of Effectiveness

To learn what limits the effectiveness of additional responders, we calculated the completion time as the number of responders is increased from one to ten. The results are shown in Figure 2.

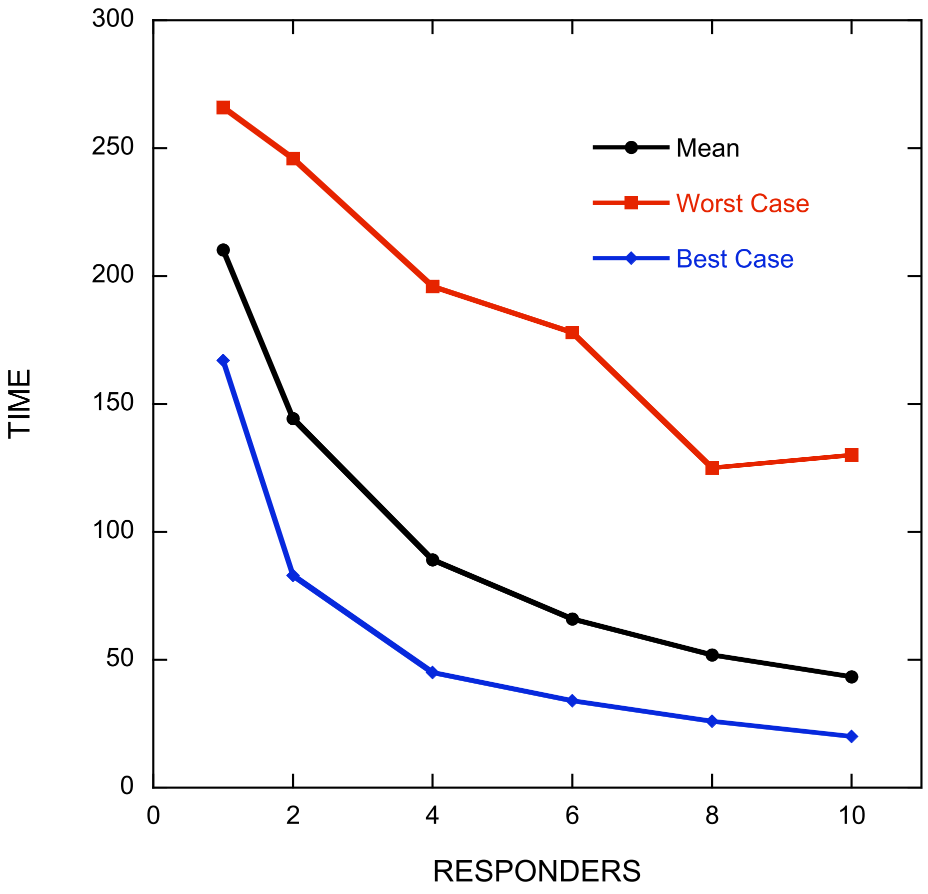


Figure 2. Completion time as a function of the number of responders

Unfortunately, this representation of the results is not very informative. It is difficult to separate any effect of diminishing returns from the increase in performance due to the added responders, especially since the completion time is expected to be inversely proportional to the number of responders. A more suitable measure of effectiveness (MOE) is the number of incidents serviced per unit time. This MOE is shown in Figure 3, where the unit of time has been defined to be the completion time for one responder.

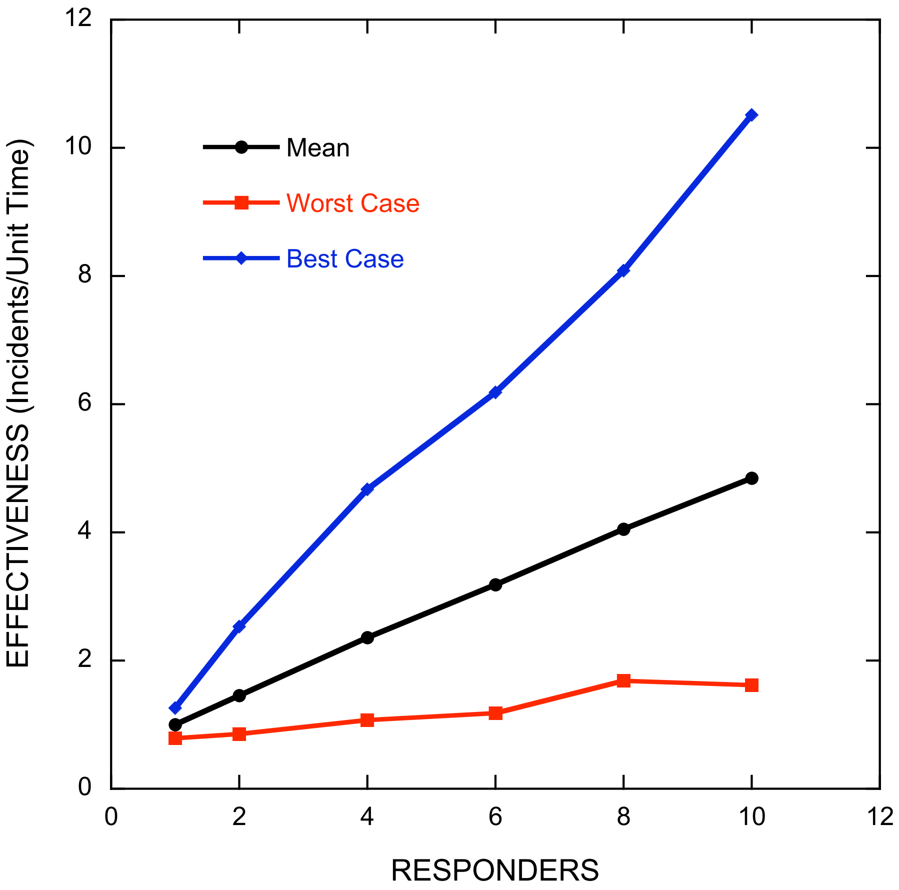


Figure 3. Measure of effectiveness as a function of number of responders

Two important facts are revealed by Figure 3. First, the performance increase produced by each additional responder is essentially constant in this range. The mean value of the MOE is almost perfectly linear. There is no bottleneck or knee in the curve of the kind usually associated with a law of diminishing returns. Second, the best-case performance reflects an essentially 100% effectiveness of the additional responders, in contrast to the constant 47% average marginal effectiveness. The gap between the best case and the worst or average case increases with the number of responders, reflecting the fact that it is worst (or bad) case performance that is lowering the mean. The distributions of completion times for various numbers of responders provide further confirmation of this conclusion. Figure 4 shows this distribution for one responder and for ten responders. For one responder, the distribution is symmetric, with a nearly Gaussian distribution about the mean. For ten responders, it is clearly skewed toward longer times, with a significant tail of very long completion times.

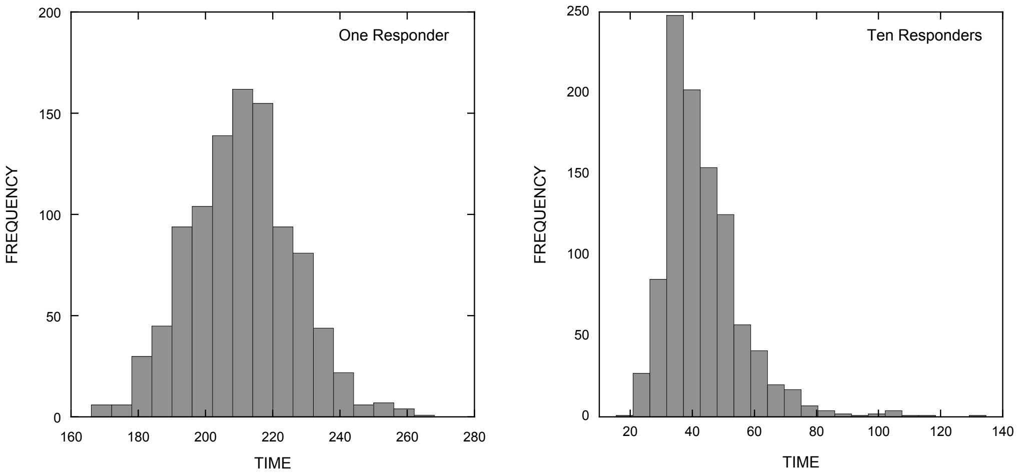


Figure 4. Completion times for one and ten responders

Dysfunctional Self-Organization

The source of this increasingly poor worst-case performance (relative to the best or average case) is readily found by examining the animations produced by the simulation for these cases. Since the animations cannot be reproduced here, a brief description will have to suffice. When cases with very long completion times are examined carefully, we typically find that at some point many (or even all) the responders have formed a tight cluster that travels together with its members competing for the same nearest incident. The origin of this behavior is easily seen:

> Initially, responders that happen to choose the same incident approach one another.
>
> The first one to reach the common goal deals with it, but the two responders are now closer than when they started.
>
> It is now more likely that the responders will again choose the same goal.
>
> Eventually, groups of responders travel together, seeking the same goals and reducing effectiveness.

For the case of two responders, it is easy to track the distance between the two during the course of a response scenario and to compare cases of long and short completion times. Although not as dramatic as an animation, a plot of this separation (Figure 5) clearly demonstrates the effect we have described.

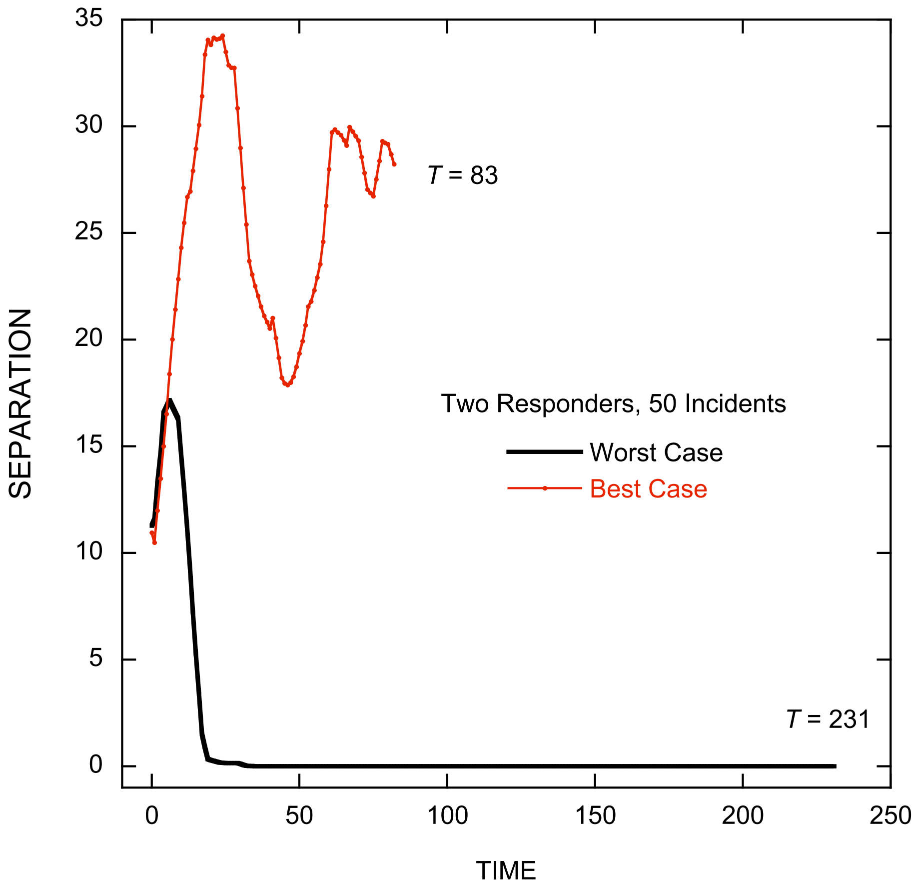


Figure 5. Separation of two responders in units of grid intervals as a function of time

In the worst case, the two responders operate independently for the first 15 time steps, but at $T = 16$ they select identical goals and begin to approach one another. By $T = 18$ they are separated by less than one grid interval, and from $T = 35$ to the end of the simulation their separation is negligible. For most of the simulation, they behave as if there were only one responder present, and the completion time (231 steps) is somewhat worse than the average for one responder (210.3 steps) and not much better than the worst-case time for one responder (266 steps). In contrast, the best-case performance (83 steps) is substantially less than half the average time for one responder. This result is obtained when the responders maintain a significant separation, in the neighborhood of 25 grid intervals, which is comparable to one-half of the diagonal of the operating area (23 intervals).

Most cases are intermediate between these extremes, but it is important to note that once a pair of responders have approached each other so closely that they select the same incident as the next goal, it is very unlikely that they will separate again during the rest of the simulation. The result is the bias toward long completion times illustrated in the distribution for ten responders shown in Figure 4. We have termed this behavior dysfunctional self-organization, since it is clearly an emergent result of the behavior rules and information flow, but it works against the design objective of the system.

Improved Behavior Rules

Since it is the greedy behavior of the responders that produces dysfunctional self-organization, it is reasonable to ask whether the rules can be modified to prevent it. Without violating the requirement that responders not communicate with one another and without providing any information not already available, it is possible to devise mechanisms for the responders to avoid one another.

"Simple" Avoidance

In the baseline model, a responder will chose a new goal if the incident toward which it is heading is reached first by another responder and so disappears from the list of pending incidents. If both responders then select the nearest incident as their next goal, the self-organization mechanism described above is triggered. On the other hand, a responder that fails to reach the goal first will know that another is in the vicinity because its goal has been serviced. (The first on the scene has no such information, however.) It is reasonable for a responder that is not first on the scene to assume that the nearest remaining incident to its current position is also the one nearest to the location of its previous goal. It can (usually) avoid making the same choice of goal as the other responder by choosing the *second* nearest incident to its present location. As we shall see, this logic is not correct in every case, but a revised behavior rule based on it avoids cluster formation and improves performance. (We call this rule "simple," with the implication that more elaborate rules will follow.)

Figure 6 shows the distribution of completion times for ten responders, with and without the simple avoidance rule. The left half of the figure is identical to the right half of Figure 4, except that the vertical scale has been changed to permit direct comparison between the two sides of Figure 6. The result of the avoidance rule is striking. The average completion time has improved by 15% (from 43.4 to 37 steps). More important, the shape of the distribution has changed dramatically, becoming narrower and more symmetric. With this narrowing, the worst-case time has improved by 37% (from 130 to 82 steps). By greatly reducing the likelihood of extremely long completion times and improving the predictability of the process, a rule of this type can make the practical use of self-organization significantly more attractive.

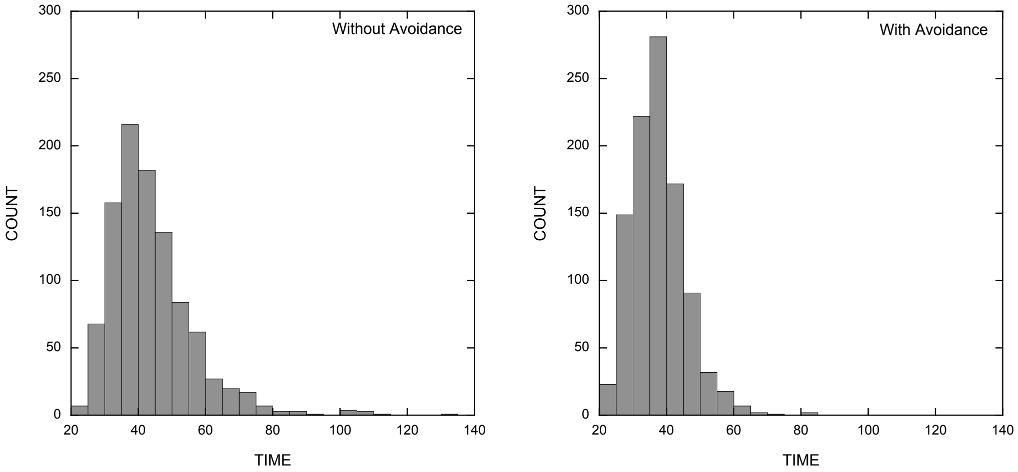


Figure 6. Distribution of completion times for ten responders, with and without application of the simple avoidance rule

It is also instructive to examine the effects of the avoidance rule on the step-by-step evolution of a simulated response. The worst case for two responders, shown in Figure 5, was run again, using the simple avoidance rule, and the results are shown in Figure 7. Once again, the improvement is substantial, with the completion time improving by 40%. The improvement mechanism is quite visible; there is a rapid increase in distance between the two responders on five occasions when they approach nearer than ten grid intervals. Inspection of the simulation log reveals that the avoidance rule was invoked at each of these times.

Conflict Avoidance

When two responders approach the same incident, one of them will necessarily arrive first and service it. As noted above, it is probable, but not certain, that both responders will then select the same incident as their new goal. The simple avoidance rule is designed to prevent this. There will be instances, however, when the rule is invoked unnecessarily, because the two responders would have chosen different goals in any event. This can lead to reduced effectiveness, because one of the responders is now diverted from its nearest possible goal to the second nearest. It is not difficult to modify the rule to correct this, since the required information is already available to the responder that is not first on the scene. This responder knows the location of the incident just serviced by the first responder to arrive, as well as the locations of all other pending incidents. It can therefore determine which goal will be selected by the other responder and can invoke the avoidance rule (chose the second nearest incident) only when choosing the nearest would create a conflict, that is, both responders choosing the same goal.

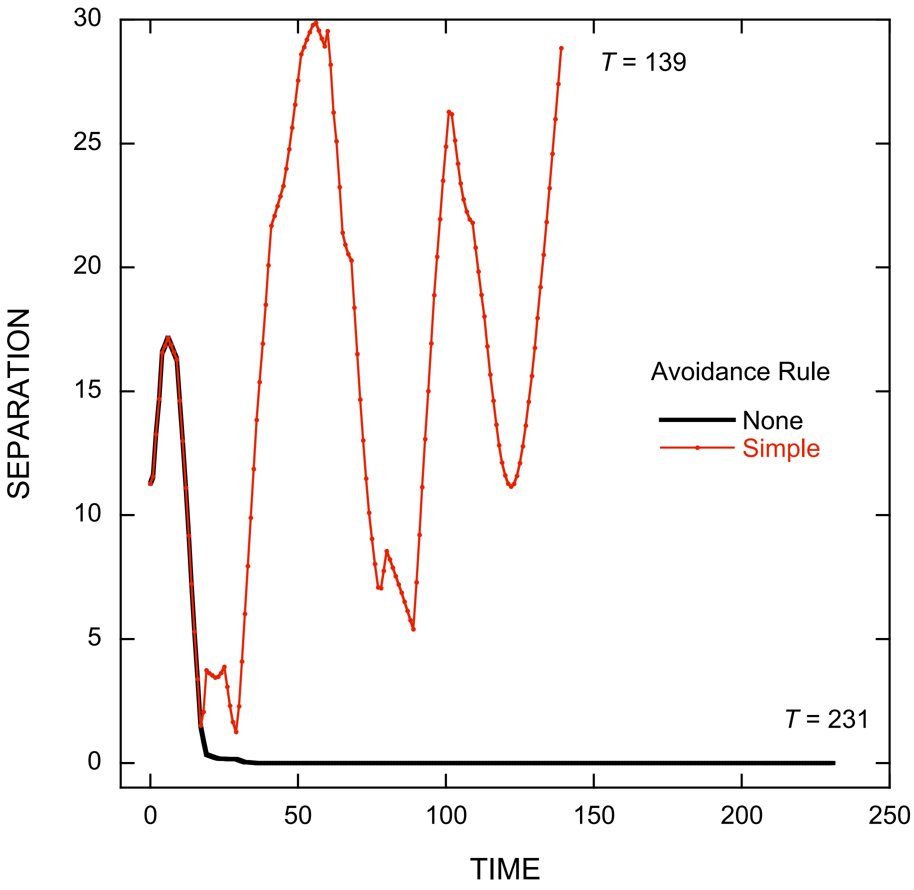


Figure 7. Separation of two responders, with and without application of the simple avoidance rule, for the worst-case scenario without avoidance

The conflict avoidance rule produces a small but significant increase in performance when compared to the simple avoidance rule. The two avoidance rules are compared in Figure 8, where the effectiveness is shown as a function of the number of responders for the average and worst-case simulations. For ten responders, the conflict avoidance rule provides improvement in effectiveness of 18% over the case with no avoidance. The ten responders now perform on average nearly six times better than one.

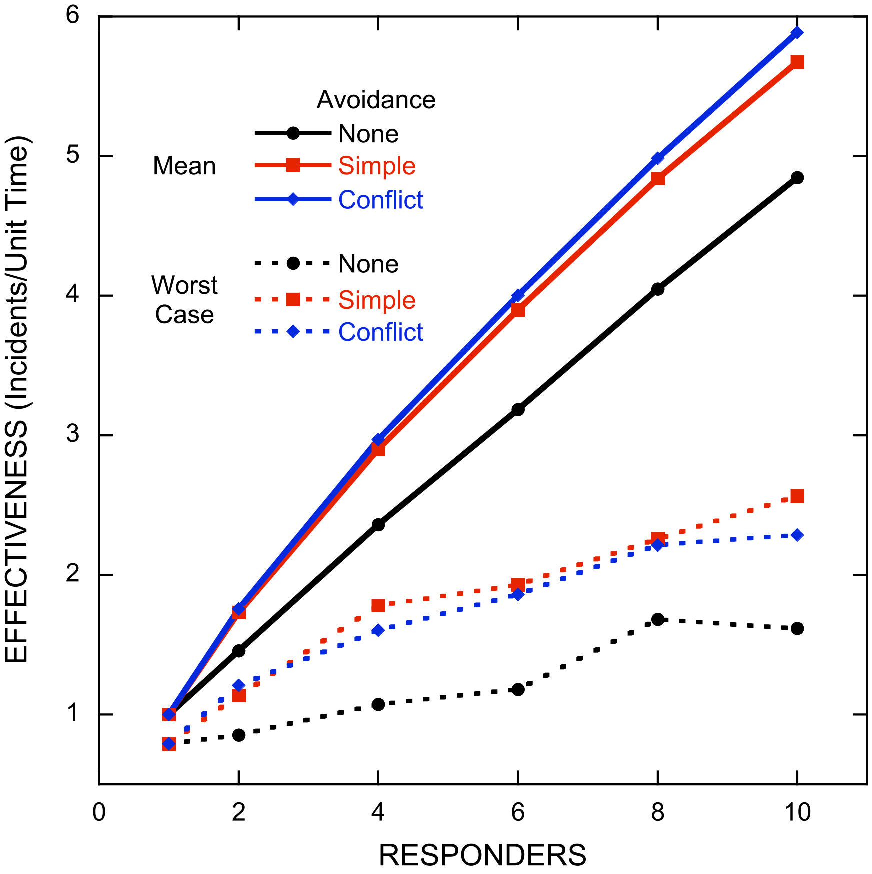


Figure 8. Effectiveness as function of the number of responders (average and worst case) without avoidance and with the simple and conflict avoidance rules

Observing Self-Organization

As noted above, the clustering of responders that contributes to their reduced effectiveness was discovered by observing an animation of the simulated behavior. This self-organization, while dysfunctional, is easily detected. Ironically, when an avoidance rule is imposed, the behavior appears much more random in the animation, but strictly speaking it is equally self-organized. Instead of a small set of cases (defined by long completion times) that exhibit a pathological behavior, we now have a much larger set (essentially all simulations) from which that behavior has been excluded. This is a recurring issue in the theory of complex systems, where the degree of order (or organization) is defined formally in terms of the number of states that are accessible or inaccessible to the system but may not be readily visible in the appearance or behavior of the system (Bar-Yam 1997).

Discussion

The reader will have noticed that in the model we have analyzed responders are not allowed to communicate directly with one another or through a central command, but they do in fact communicate indirectly by observing the effects of one another's activities. Using the list of pending incidents, a responder is able to detect the presence of other responders. The avoidance rules we have discussed here and others we have explored make more or less extensive use of this information. This kind of implicit communication through modification of the environment is known as stigmergy, and it is fairly common in biological systems, the most familiar example being the pheromone trails left by ants and other social insects. Self-organization or self-synchronization through observation of others is common in human group activities, including team sports, workplace collaboration, and warfare. Yet, despite intense interest in the analysis of team formation and performance, this aspect of group communication has attracted little study. Agent-based models such as the one presented here provide an opportunity to make explicit the information flows that occur during such activities and to examine them quantitatively.

Finally, it is worth noting that a previous discussion of centralization and decentralization in C2 used the TSP as an example (Dekker 2006). That study concluded that there is no difference between centralized and decentralized planning of a route for the traveling salesman. This should not be surprising, since in this case the central planner and the salesman have the same information available.

Questions for Further Study

An important issue that could not be addressed in this study is the origin of the remaining inefficiency in the self-organized incident response. Even with the conflict avoidance rule, ten responders typically waste 40% of their effort. On the other hand, their best-case performance reaches or exceeds ten times the average for a single responder. Two approaches might be tried. One is to seek further refinements of the greedy algorithm, using supplementary rules similar to the avoidance rules studied here. The other would examine alternatives to the greedy algorithm itself, while maintaining decentralized command and employing only information available to the responders.

It would also be useful to explore the effect of additional constraints on the responders. These might include: limits on the total distance traveled by any one responder or the number of incidents it can service, delays in updates of the list of pending incidents, errors in the incident list, or variations (random or systematic) in the speed with which the responders can travel. Algorithms that are robust to some or all of these perturbations could be very valuable.

Conclusion

We have presented and analyzed a simple model of emergency response involving multiple incidents and multiple responders, where no centralized C2 is available. The goal was not to demonstrate any advantage associated with decentralized control. In fact, the observed scaling of performance with the number of responders was not altogether encouraging. On the other hand, a very simple algorithm using only local information and decision making was shown to produce results that could very well be satisfactory under emergency conditions or when a more centralized command structure would not be acceptable.

References

Alberts, David S. 1996. *Information Age Transformation*. Washington: Command and Control Research Program.

Alberts, David S. and Richard E. Hayes. 2007. *Planning: Complex Endeavors.* Washington: Command and Control Research Program.

Bar-Yam, Yaneer. 1997. *Dynamics of Complex Systems.* Reading: Perseus.

Cebrowski, A. K. and J. J. Garstka (1998). Network-Centric Warfare: Its Origin and Future. United States Naval Institute Proceedings. 124: 28.

Dekker, Anthony H. 2006. Centralisation vs. Self-Synchronisation: An Agent-Based Investigation. 11th International Command and Control Research and Technology Symposium. http://www.dodccrp.org/events/11th_ICCRTS/html/papers/030.pdf

Department of Homeland Security (US). 2008. National Incident Management System. http://www.fema.gov/pdf/emergency/nims/NIMS_core.pdf

Department of Defense (US). 2008. *Joint Operations.* Joint Publication 3-0.

Gell-Mann, Murray. 1995. What is Complexity? Complexity 1:1.

Wilensky, Uri. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.