

# Multi-Agent Systems in Open Environments

Myriam Abramson  
Naval Research Laboratory  
Washington, DC 20375  
abramson@ait.nrl.navy.mil

Ranjeev Mittu  
Naval Research Laboratory  
Washington, DC 20375  
mittu@ait.nrl.navy.mil

## Abstract

*This paper proposes a new approach to multi-agent systems leveraging from recent advances in networking and reinforcement learning to scale up teamwork based on joint intentions. In this approach, teamwork is subsumed by the coordination of learning agents. The intuition behind this approach is that successful coordination at the global level generates opportunities for teamwork interactions at the local level and vice versa. This unique approach scales up model-based teamwork theory with an adaptive approach to coordination.*

## Introduction

Open environments such as Peer-to-Peer (P2P) and wireless or Mobile AdHoc Networks (MANET) provide new challenges to communication-based coordination algorithms such as *joint intentions*[8] as well as the opportunity to scale-up. Our framework is based on the proxy architecture of Machinetta[9] where *proxy* agents perform the domain-independent coordination task on behalf of *real*, domain-dependent agents. This framework is extended with a coordination mechanism of individual actions based on reinforcement learning. This adaptive proxy agent architecture is illustrated in Figure 1. In this approach, local teamwork outcomes provide the feedback for learning the coordination task on a larger scale. The teamwork theory of joint intentions and its associated problems in open environments are presented first and then our tentative approach, Open-MAS, with illustration from the fire fighting example of the RoboCup Rescue competition [7].

## Joint intentions and Open Environments

Joint intentions[3, 8] form the cornerstone of teamwork theory of BDI (Belief, Desire, Intention) agents.

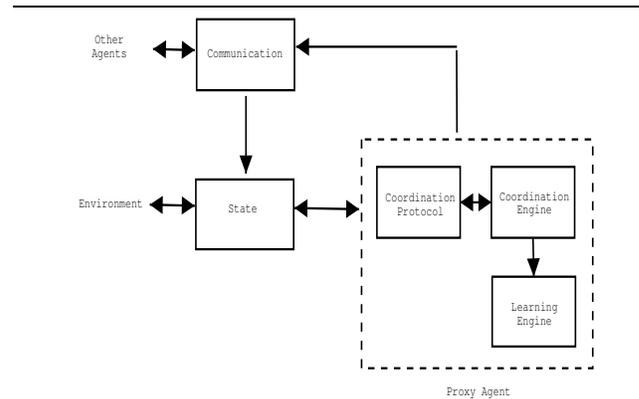


Figure 1. Adaptive proxy agent architecture

It differentiates joint actions from individual actions by the presence of a common mental state (beliefs) and the joint commitment of achieving a goal. It is based on the communication of critical information among team members. Open environments are characterized by their dynamic nature and the heterogeneity of the agents as well as asynchronous and unreliable communication on a large scale. The problems addressed can be categorized as follows: team formation, role allocation, synchronization of beliefs, and communication tradoffs.

1. **Team Formation.** An open environment gives the opportunity to find teammates appropriate for a task instead of relying on a fixed group of agents. What is the best way to find teammates? When is the best time to find teammates? In open environments, peers form “groups” by similarity of individual interests. Likewise, similarity of individual intentions is a necessary stepping stone for team formation in open environments. An intention is defined here[3] as the decision to do something in order to achieve a goal and can be construed as a partial plan.

2. **Role Allocation.** While direct point-to-point communication with any node can be expensive and uncertain, access to neighbors is readily available in open environments. P2P middleware, such as JXTA (Juxtapose)[1], provides the functionality needed to communicate reliably and cheaply with neighbors. In MANET, the possibility of disconnecting the network is another constraint in accepting a role requiring a change in location. Figure 2 describe the connection role that peers play in communication in MANET. In open environments, multiple teams are involved. How to adjust the connectivity role of the agents so that each team can accomplish its goals most effectively?
3. **Synchronization of Beliefs.** The theory of joint commitments is based on the ability to synchronize beliefs regarding “who is doing what”. Teamwork breaks down when roles do not match expected beliefs. How to adjust gracefully to delays in communication?
4. **Communication Selectivity.** The tradeoffs involve the robustness that redundancy of messages can provide in open environments versus the costs of communication to the network. When reliable communication cannot be assumed, selective communication of critical information might be detrimental to the coordination task.

Synchronization of beliefs and communication selectivity are areas that are complicated by open environments, while team formation as well as role allocation are the problems we are interested in addressing given these complicating factors.

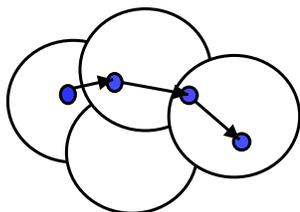


Figure 2. Multi-hop routing in a MANET

## OpenMAS Approach

Our approach consists of leveraging from the belief framework of cognitive agents at the local level but endowing the agents with the adaptative capabilities

of reinforcement learners as an additional coordination mechanism at the global level where communication is unsure and unreliable. The overarching issues addressed are (1) how to integrate general models of cooperation with reinforcement learning in distributed, open environments (2) what are good metrics for the propagation of beliefs to heterogenous agents and (3) how to integrate multiple goals.

## Methodology

Through the propagation of beliefs, the agents have some knowledge of the global situation albeit imperfect and decaying with time. This capability relaxes the invalidation of the Markov property for multi-agent reinforcement learning systems. Instead of committing to a non-local role, the agents just commit to the next individual step. This is a least-commitment approach that addresses the problems outlined above of teamwork in open environments. Local environmental beliefs on the other hand trigger a role allocation mechanism<sup>1</sup> among neighbors sharing the same beliefs. The joint actions generated have precedence over the individual actions generated by the coordination learning mechanism. Similarities between joint actions and individual actions produce the terminal rewards needed for the learning algorithm. In this approach, there is a tight integration between the local level of teamwork and the global level of coordination. The overall approach is described in Algorithm 1. Figure 3 illustrates the approach in the fire fighting example.

---

### Algorithm 1 Intention/action loop

---

INPUT: intentions  
OPENMAS-interpreter:

```

do
  <information,intention> ←receive-information()
  if similar-intentions(intention)
    accept-information()
    update-current-state()
    forget-and-predict()
    takeNextStep()
    propagate <next step,intentions> tuples
forever

```

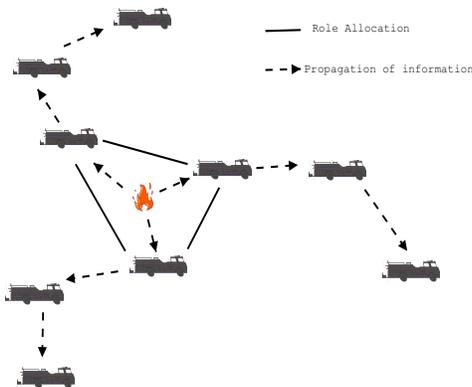
The information received includes information from peers and/or perceived information from the environment.

---

The environment of agents acting under uncertainty can be conveniently modelled as a POMDP (Partially-

---

<sup>1</sup> Role allocation of mutually exclusive tasks among agents can be modelled with a distributed resource allocation algorithm similar to the *drinking philosophers problem*[2].



The agents propagate changes of position and changes in the fires' status to their neighbors recursively according to a time-to-live (TTL) parameter. Role allocation strategies resolve local conflicts.

**Figure 3. Fire fighting example**

observable Markov Decision Process). POMDP can be reformulated as continuous-space Markov decision processes (MDPs) representing belief states[6] and solved using an approximation technique. When propagating local environmental beliefs, the redundancy of messages reinforces the current state beliefs while decaying with time. Propagated location information is updated through the same prediction mechanism used to select the next action of the agents. Forgetting and prediction are the two tools enabling the synchronization of beliefs through asynchronous and unreliable communication. The most likely state of the global situation is then modeled as an MDP and the action to take determined by a stochastic policy approximated by a policy gradient method [11].

In addition to fighting and searching for fires, the firetrucks (the agents) have the additional task of maintaining connectivity of the network. It is necessary to balance those sometimes conflicting goals. The synergy of those two goals should maintain a proper degree of dispersion among the agents. In this context, multiple MDPs model the different intentions of the agents. An MDP represents the belief map of the agents' location while another MDP represents the belief map of the location of the fires. The action to take is the best action[5] across those MDPs after a period of exploration.

## Problem Modeling

The world is modeled as the problem space:

$$W = \{S, S', A, T, R\}$$

where

- $S$  is the believed perceived local state of the world.
- $S'$  is the believed global state of the world through propagation of information.
- $A$  is the set of actions.
- $T$  is the set of transition probabilities

$$S \times A \times S \rightarrow [0, 1]$$

- $R$  is the set of roles.

and

$$S_i \times R \rightarrow A_i$$

$$S'_i \times A \rightarrow \mathbb{R}$$

where

- $A_i$  is the action determined to achieve role  $R$ .
- $A_j$  is the action determined by coordination in the believed state space  $S'$ .

A reward is obtained if  $A_i = A_j$ .

## Related Work

The dissemination of information enables agents to obtain some global, though imperfect, knowledge of the world. This capability is taken into account in scaling up teamwork approaches based on communication and our approach also takes this capability into account to enhance multi-agent learning. Our approach is different from the large-scale coordination of Machinetta proxies[10] because (1) the uncertainty due to delays in communication is taken into account and (2) individual actions lead to joint actions through online adaptation.

Our approach is also related to learning approaches of plan competencies in BDI multi-agent systems[4] where plan successes or failures trigger explanation-based learning to modify the plan. Our approach however does assume a correct and complete plan library and success of the task is dependent only on the coordination task.

## Conclusions

Open environments such as P2P and MANET forces a reexamination of teamwork in large scale systems relying more on adaptive coordination than explicit cooperation requiring synchronization points. The capability to acquire global, albeit imperfect, knowledge through the propagation of information makes it possible to use independent reinforcement learners for coordination tasks in multi-agent systems. Similarity of intentions can help relieve the burden placed on the network by selectively propagating information. A local teamwork model drives the rewards of the overall coordination task. This approach scales well to any dimensions and its precision can be modulated by the TTL parameter. This approach will be compared quantitatively with centralized and omniscient algorithms and variations in the network reliability in future work.

## References

- [1] Project jxta. <http://www.jxta.org>.
- [2] K. M. Chandy and J. Misra. The drinking philosophers problem. *ACM Trans. Programming Languages and Syst.*, 6:632–646, 1984.
- [3] Philip R. Cohen and Hector J. Levesque. Teamwork. *Nous*, 25(4):487–512, 1991.
- [4] Alejandro Guerra-Hernandez, Amal El Fallah-Seghrouchni, and Henry Soldano. Learning in bdi multi-agent systems. In *CLIMA IV Computational Logics on Multiagent Systems*, 2004.
- [5] M. Humphrys. Action selection methods using reinforcement learning. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 135–144, 1996.
- [6] L. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, (101):99–134, 1998.
- [7] H. Kitano, S. Tadokoro, I. Noda, H. Matsubara, T. Takahashi, A. Shinjoh, and S. Shimada. Robocup rescue: Search and rescue in large-scale disasters as a domain for autonomous agents research. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume VI, pages 739–743, 1999.
- [8] H. J. Levesque, P. R. Cohen, and J. Nunes. On acting together. In *Proceedings of the National Conference on Artificial Intelligence*, 1990.
- [9] P. Scerri, D. Pynadath, N. Schurr, A. Farinelli, S. Gandhe, and M. Tambe. Team oriented programming and proxy agents: The next generation. Workshop on Programming MultiAgent Systems, AAMAS 2003.
- [10] Paul Scerri, Elizabeth Liao, Justin Lai, and Katia Sycara. *Cooperative Control*, chapter Coordinating Very Large Groups of Wide Area Search Munitions. Kluwer Publishing, 2004.
- [11] R. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems*, volume 12, pages 1057–1063. MIT Press, 2000.