Multi-Agent Systems for Deep Understanding of Cyberspace

Jeffrey M. Bradshaw and Marco Carvalho

Within the Deep Understanding of Cyberspace focus area of the Moving Target theme of the Federal Cybersecurity R&D Strategic Plan, several key challenges are outlined. At the most general level, these include the need to "understand our system state, be aware of our surroundings, know the soundness of the structures on which we rely, and know what is happening around us" (p. 9). The plan recognizes, however, that this deep understanding is useless unless we can apply it effectively in the service of making better "decisions within the tight time constraints of cyber actions" (p. 9) To achieve this objective, "we must greatly enhance the speed of our complex analytics and tighten our feedback loops" (p. 9).

In this presentation, we outline how both objectives—*deep understanding* and *resilient control* of cyberspace—can be addressed with great effectiveness through the application of multi-agent systems. By the ability of agents to function in close and continuous interaction with people in the role of teammates rather than tools [10; 16], they can address the requirement to support "human [understanding and] decision-making," while minimizing "the combination of high complexity and short processing time [that] strains human cognitive processes" (p. 9). By the ability of agents to operate independently in complex situations without constant human supervision, collaborating teams of agents can perform tasks on with a degree of scalability, flexibility, and resilience that would be difficult for other approaches to duplicate [1]. Today, to the credit of most cyber security programs, human experts are used to recognize abnormalities, identify and analyze true cyber threats, and then act to mitigate them. However, those very defenders are, by their own account, easily overwhelmed by determined attacks. Their "knowledge" and discrimination cues when implemented in a multi-agent framework can provide the speed and reliability for repetitive high-tempo operations, allowing the defenders to learn and discover the more novel information. Using a military term, the agent framework becomes a "force-multiplier" supplementing human expertise.

Multi-Agent Systems in Support of Deep Understanding of Cyberspace

Agents are typified by their active, adaptive nature. This quality is often characterized in the Artificial Intelligence literature by the word "autonomy." However, as we have argued elsewhere [4; 13], autonomy is exactly the *wrong* word for characterizing agents that are designed to assist, rather than replace, people in the kind of sensemaking tasks that contribute to deep understanding. Though continuing research to make agents more active, adaptive, and functional is essential, the point of increasing such proficiencies is not merely to make the machines more *independent* during times when unsupervised activity is desirable or necessary (i.e., *autonomy*), but also to make them more capable of sophisticated *interdependent* joint activity with people and other machines when such is required—i.e., *teamwork* [2; 11; 12; 14; 15]. The mention of *joint* activity highlights the need for autonomous systems to support not only fluid orchestration of task handoffs among different people and agents, but also combined participation on shared tasks requiring continuous and close interaction—i.e., *coactivity.*

IHMC's Luna is an agent framework designed for the demands of cyber operations [5]. Within the Sol cyber framework, Luna agents function both as interactive assistants to analysts and as continuously-running background aids to data processing and knowledge discovery. To facilitate their use as sensemaking partners to analysts, Luna agents are comprehensively governed by semantically-rich policies that are defined, analyzed, and enforced by IHMC's KAoS policy services framework [19], enabling a high level of assurance in their deployment.

We describe agent-supported sensemaking as a process of "coactive emergence" [3]. Coactive emergence is an iterative process whereby secure system configurations, effective responses to threats, and useful interpretations of data are continuously developed through the interplay of joint sensemaking and decision-making activities undertaken by analysts and software agents. The word "coactive" emphasizes the joint, simultaneous, and interdependent nature of such collaboration among analysts and agents. The coactive component of agent systems allows for a more aggregated defense. For example, today's tools allow for limit triggers, or event triggers to be flagged for cyber defenders. A coactive system of "agent observers" can be extended to correlate and synthesize events and anomalies from many diverse, and often "disconnected" trigger systems. In the ideal state, they can even learn and expand the knowledge base for the cyber defenders.

Ideally, the process of coactive emergence is synergistic, leading to progressive convergence on hypotheses relating to the current situation. Of course, competition among hypotheses is also desirable in sensemaking in order to encourage the exploration of the same space (or a wider space) from different perspectives and to avoid premature closure. Such interaction would support the "Deep Understanding" theme requirements to be able to "view the situation from alternative points of view and to get below surface indicators to determine underlying causes and conditions" (p. 9).

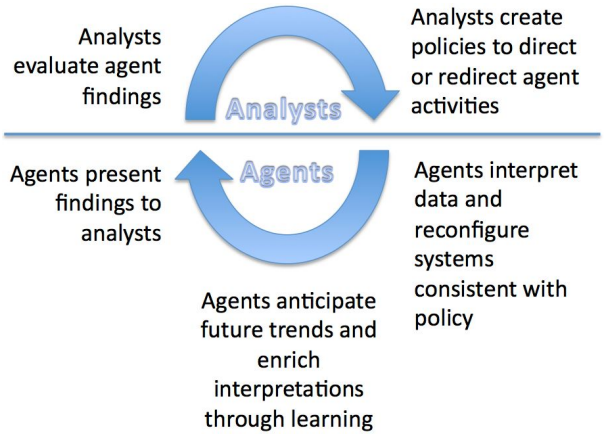

Figure 1. The Coactive Emergence Cycle

Figure 1 illustrates how a cycle of coactive emergence applies to cyber sensemaking. Note that the diagram showing a single loop is somewhat misleading, since multiple threads of agent and human activity would be operating on individual schedules rather than in lockstep as the figure implies:

1. Agents are pre-coded in Java to perform particular classes of analytic tasks. Analysts use their knowledge to encode agent behavior into high-level declarative policies that enable the agents to perform their tasks in a secure, predictable, and controllable manner.
2. Subsequently, agents interpret real-time data for presentation to analysts. Within the constraints of policy, agents may not only sense but also act—for example, manipulating system configurations to improve security in Moving Target Defense [9].
3. Agents may optionally enrich their findings with additional information gleaned through learning (e.g., hypothesized correlations between data sets of interest, anticipated future trends). Because of their built-in abilities to work together dynamically to analyze and synthesize meaningful events from the raw data, agent interpretations can be more easily made to match the kinds of abstractions found in human interpretations more closely than those that rely exclusively on low-level sensors.
4. Agents may aggregate and present their findings by visually annotating graphical displays in real-time in order to highlight and draw the attention of the analyst to anomalous or otherwise interesting elements, such as possible attacks. We call such displays *mediating*

representations — highly-communicative visual models of the situation that can be simultaneously used by mixed teams of people and software agents in order to come to a common understanding of a situation. Analysts interact with these displays in order to explore and evaluate how agent findings bear on their hypotheses.

5. As agent-derived information is presented to analysts, they may agree or disagree with agent findings, leading to further corrections and refinements of interpretations, and consideration of response options.
6. Analysts continue to direct and redirect ongoing agent activity through the construction of new agents, modification of agent policies, and extensions to lines of inquiry. As research capabilities grow, cyber analysts and systems architecture experts can allow for system "hardening" in real time by empowering agents to change any number of configurations in near real time (e.g., router ACLs, network access controls, cross validation of credentials, degrade or partially isolated suspect subnets). The opportunities and options are many, but the degree of automation and the tolerance for change can be controlled through policy.

Addressing the technology gap created by the emphasis of past research on how to make machines self-sufficient, these and similar efforts are beginning to emerge to understand and develop capabilities that would allow the participation of humans as first-class citizens in collaboration with autonomous systems. Such capabilities would enable autonomous systems not merely to do things *for* people, but also to work together *with* people and other systems—the inevitable leap-forward required in multi-agent system design and deployment.

Multi-Agent Systems in Support of Resilient Control

Within the framework of resilient systems engineering, Branlat and Woods have discussed important patterns that lead to failure in complex systems [21]. Multi-agent systems can be used to provide support for adaptive performance in the face of stressors and surprise. For example, Rieger has developed a promising multi-agent-based approach to resilient control systems [18]. Another approach is based on principles of polycentric governance [17].

A related notion of organic resilience [6] relies heavily on biologically-inspired analogues and self-organizing strategies for the management and defense of distributed complex systems. Carvalho, *et al.* have previously applied the concept for the defense of tactical communication systems [6] and mission-critical cloud applications [6]. The concept focuses on the design of emergent coordination mechanisms through local gradients and implicit signaling. Multi-layer defense frameworks following the same principles were later developed for critical infrastructure protection and distributed control systems [8]. These infrastructures included humans as an integral part of the system, working in collaboration with software agents to improve system resilience.

As with many biological systems, the goal of an approach that relies on polycentric governance is to avoid static and centralized single-point-of-failure solutions for organizing work to the greatest degree practical. Thus, although groups of agents within the system are collectively responsible for jointly executing various tasks, the specific responsibilities assigned to agents are not fully determined in advance. The goal is to allow the agents to self-organize within the constraints of their individual capabilities and current availability. As described in [6; 7], the premise of such resilience depends on understanding the advantages and disadvantages of particular techniques for self-organization for different problems within a given situation and computing environment.

The use of collective obligation policies [20] is critical for practical applications of polycentric governance. Whereas an individual obligation is a policy constraint that describes what must be done by a particular individual, collective obligations are used to explicitly represent a given agent's responsibilities within a group to which it belongs, without specifying in advance who must do what. In other words, in a collective obligation, it is the group as a whole that becomes responsible, with individual members of the group sharing the obligation at an abstract level. The self-organizing nature of the system enables the agents to revisit responsibilities and resource allocations themselves, as needed, on an ongoing basis.

The flexibility and maturity of multi-agent systems has been gained by their application to many domains. As the Cyber Security community further embraces this technology the possibilities to improve cyber defense are enormous.

References

1. Bradshaw, J.M. "An introduction to software agents." In *Software Agents*, edited by J.M. Bradshaw, 3-46. Cambridge, MA: AAAI Press/The MIT Press, 1997.
2. Bradshaw, J.M., P. Feltovich, and M. Johnson. "Human-Agent Interaction." In *Handbook of Human-Machine Interaction*, edited by G. Boy, 283-302. Ashgate, 2011.
3. Bradshaw, J.M., M. Carvalho, L. Bunch, T. Eskridge, P.J. Feltovich, C. Forsythe, R.R. Hoffman, M. Johnson, D. Kidwell, and D.D. Woods. "Coactive emergence as a sensemaking strategy for cyber operations." Pensacola, FL: IHMC Technical Report, 2012.
4. Bradshaw, J.M., V. Dignum, C. Jonker, and M. Sierhuis. "Introduction to Special Issue on Human-Agent-Robot Teamwork." *IEEE Intelligent Systems 27*, no. 2 (March-April 2012): in press.
5. Bunch, L., J.M. Bradshaw, M. Carvalho, T. Eskridge, P.J. Feltovich, J. Lott, and A. Uszok. "Human-Agent Teamwork in Cyber Operations: Supporting Co-Evolution of Tasks and Artifacts with Luna." Presented at the Tenth German Conference on Multiagent System Technologies (MATES 2012) (LNAI 7598), Trier, Germany, October 10-12, 2012, 53-67.
6. Carvalho, M., T. Lamkin, and C. Perez. "Organic resilience for tactical environments." In *Fifth International ICST Confernece on Bio-Inspired Models of Network, Information, and Computing Systems (Bionetics)*. Boston, MA, 2010.
7. Carvalho, M., D. Dasgupta, M. Grimaila, and C. Perez. "Mission resilience in cloud computing: A biologically inspired approach." In *Proceedings of the Sixth International Conference on Information Warfare and Security (2011)*, 2011.
8. Carvalho, M. and C. Perez. "An evolutionary multi-agent approach to anomaly an evolutionary multi-agent approach to anomaly detection and cyber defense." In *CSIIRW '11: Proceedings of the 7th Annual Workshop on Cyber Security and Information Intelligence Research*. New York City, NY: ACM, 2011.
9. Carvalho, M., J.M. Bradshaw, L. Bunch, T. Eskridge, P.J. Feltovich, R.R. Hoffman, and D. Kidwell. "Command and control requirements for Moving Target Defense." *IEEE Intelligent Systems 27*, no. 3 (2012): 79-85.
10. Christofferson, K. and D.D. Woods. "How to make automated systems team players." In *Advances in Human Performance and Cognitive Engineering Research, Vol. 2*, edited by E. Salas. JAI Press, Elsevier, 2002.
11. Feltovich, P., J.M. Bradshaw, R. Jeffers, and A. Uszok. "Social order and adaptability in animal, human, and agent communities." Presented at the Proceedings of the Fourth International Workshop on Engineering Societies in the Agents World, Imperial College, London, 29-31 October, 2003, 73-85.
12. Feltovich, P.J., J.M. Bradshaw, W.J. Clancey, and M. Johnson. "We regulate to coordinate: Limits to human and machine joint activity." Presented at the Proceedings of ESAW 2006, Dublin, Ireland, 6-8 September, 2006.
13. Hoffman, R.R., J.M. Bradshaw, and K.M. Ford. "Introduction." In *Collected Essays on Human-Centered Computing, 2001-2011*, edited by R.R. Hoffman, P. Hayes, K.M. Ford, and J.M. Bradshaw. New York City, NY: IEEE Press, 2012.
14. Johnson, M., J.M. Bradshaw, P.J. Feltovich, R.R. Hoffman, C. Jonker, B. van Riemsdijk, and M. Sierhuis. "Beyond cooperative robotics: The central role of interdependence in coactive design." *IEEE Intelligent Systems 26*, no. 3 (May/June 2011): 81-88.
15. Johnson, M., J.M. Bradshaw, P.J. Feltovich, C. Jonker, B. van Riemsdijk, and M. Sierhuis. "Autonomy and interdependence in human-agent-robot teams." *IEEE Intelligent Systems 27*, no. 2 (March-April 2012): 43-51.
16. Klein, G., D.D. Woods, J.M. Bradshaw, R. Hoffman, and P. Feltovich. "Ten challenges for making automation a "team player" in joint human-agent activity." *IEEE Intelligent Systems 19*, no. 6 (November-December 2004): 91-95.
17. Ostrom, E. 2008. Polycentric systems as one approach for solving collective-action problems (SSRN-id130469). In *Social Science Research Network*. http://ssrn.com/abstract=1304697. (accessed September 28, 2012).
18. Rieger, C.G. "Personal Communication." October 26, 2012
19. Uszok, A., J.M. Bradshaw, J. Lott, M. Johnson, M. Breedy, M. Vignati, K. Whittaker, K. Jakubowski, and J. Bowcock. "Toward a Flexible Ontology-Based Policy Approach for Network Operations Using the KAoS Framework." Presented at the The 2011 Military Communications Conference (MILCOM 2011) 2011, 1108-1114.
20. van Diggelen, J., J.M. Bradshaw, M. Johnson, A. Uszok, and P.J. Feltovich. "Implementing collective obligations in human-agent teams using KAoS policies." In *Proceedings of Workshop on Coordination, Organization, Institutions and Norms (COIN), IEEE/ACM Conference on Autonomous Agents and Multi-Agent Systems*. Budapest, Hungary, 2009.
21. Woods, D.D. and M. Branlat. "Basic patterns in how complex systems fail." In *Resilience Engineering in Practice*, edited by E. Hollnagel, J. Paries, D.D. Woods, and J. Wreathall, 127-143. Burlington, VT: Ashgate, 2008.