

Interaction between Autonomous Systems and Anomaly Response Teams

Jane T. Malin

NASA Johnson Space Center
2101 NASA Parkway, ER2
Houston, TX 77058-3696
jane.t.malin@nasa.gov

Abstract

This paper discusses types of agent control and communication that are needed to integrate agents into extended operations teamwork and to provide help for teams performing anomaly response tasks. The three types of control and communication are analogous to those used for human-to-human loose coordination. First, the agents need to provide information to support team awareness of agent activities, plans, goals, assessments and status, enabled by multi-layered self-monitoring and communication. Second, agents need to provide mixed initiative interaction to enable joint problem solving and command completion, to minimize the effort needed to specify changes in agent actions and plans. Finally, support is needed to for interruption of an agent task or activity and safe resumption in a changed situation. Safety conscious agents can identify assumptions and perform system simulations to check proposed workaround configurations and operations against possible safety risks.

Supporting Anomaly Response Teams

Intelligent System Management Agents (ISMAs) need more capabilities to help teams that perform anomaly response tasks. These are tasks that cannot be explicitly coordinated beforehand. They typically extend over days or weeks. In this uncertain period of time, agents will be called on to provide pertinent knowledge about not only system status and health but also their own status, health, priorities and plans. They will also accommodate to team member status, health, priorities and plans. Agents, as team members, will be interrupted and commanded, and they in turn will need to interrupt and advise other team members. To help during anomaly response, ISMAs will need to be more controllable, predictable and communicative.

Safing, testing and workarounds may require changing the plans, procedures or expectations of the ISMA that is

managing the problem system. A well-designed intelligent agent can help, thus making the interventions more effective. Such an agent provides services in a system that not only “works for you” but also “works with you”. To provide such services, the agent will be controllable in new ways, to reconfigure its modes of operation, communication, and roles and responsibilities for system management.

The agent will also communicate and interact in new and predictable ways, to coordinate with heterogeneous specialist agents and with human team members who must occasionally intervene. While synchronization of execution is an important topic in tight coordination (Wilkins et al., 2003), that is not the issue that we are addressing. We are concerned with how autonomous agents can provide helpful support to the humans who occasionally closely supervise and direct them.

A system management agent that “works with you” is operating in a larger systems context where it depends on other systems, agents and human operators. It can interoperate and coordinate with other systems and agents. It is prepared to respond appropriately to situations where exceptions and events move the agent outside of the scope of its competence. It is flexible, modular and capable of reconfiguring to retain as much as possible of relevant functionality and safeguards in anomalous situations. It is prepared to gracefully handle interruptions and overrides. It can help the team to jointly formulate a course of action. It can evolve in safe ways.

Integrating Agents into Operations Teams

To integrate agents into operations teamwork, specific types of agent control and communication are needed. These are analogous to human-to-human loose coordination needs in operations teams.

- Information to support team awareness of agent activities, plans, goals, assessments and status, enabled by multi-layered self-monitoring and communication
- Low effort human specification of changes in agent actions and plans, where mixed initiative

interaction can enable joint problem solving and command completion

- Support for interruption of an agent task or activity and safe resumption in a changed situation, where assumption identification and system simulations can enable checking proposed workaround configurations and operations against possible safety risks

Multi-layered self-monitoring and communication for awareness of agent operations and status

A system management agent that “works with you” should communicate to support predictability and controllability. Research on asynchronous team coordination indicates that activity coordination is not sufficient for team success. This is especially true in situations where the team must solve novel problems and decide on a course of action that integrates strategies and policies from multiple disciplines (Espinosa et al., 2002). Activity synchronization by schedule or feedback is a basic means of coordination. Strategy and content coordination is more complex (Myers and Morley, 2001). To solve multi-system problems, one agent may need to know why another agent favors certain methods or activities. Is it because of recent events, capabilities or failures, intentions, guidelines or goals?

Communication strategies for loose coordination promote team awareness of important achievements and failures, significant changes, and problems and issues that may influence upcoming activities. Information is provided not only on events but also on the stance of team members: assessments, priorities and plans. Systems designed to promote awareness provide the following benefits: better management of team movement between loose and tight coordination, reduced and simplified communication, better predictive coordination of actions and more appropriate help and assistance (Gutwin and Greenburg, 1999).

A self-monitoring agent can communicate its activities at multiple levels of complexity. A model of levels of complexity of agent behavior has been developed, with levels of understanding, deciding and acting. It shows types of information processing and points of potential problems in agent activities (Malin and Johnson, 2003). This model extends and integrates concepts from the subsumption architecture of Brooks (1991) and cognitive theories of stages in automatic and conscious processing of information, intentions and actions (Norman 1981; Rasmussen 1982; Reason 1990).

What does a self-monitoring agent need to communicate about its assessment of failures of plans, actions and expectations? What types of information are needed for diagnosis of an apparent anomalous agent behavior, diagnosis of an apparent anomaly in the controlled

system, and direction of the agent to change its assessments, actions or plans? What does the supervising human need to know about the agent perception, interpretation or assessment? What about agent skills and rules? What about the models and goals that affect both analysis of situations and construction of responses? What about the agent’s assessment of what is going on now and next and in the past, and why and under what conditions? What about how this was inferred?

During anomaly response, safety issues come into focus. What should the agent know and communicate about system vulnerabilities and safeguards and about the agent’s vulnerabilities and safeguards?

Mixed-Initiative interaction for joint problem solving and command completion

Team members can shift among three types of coordination during anomaly response: assuming some or all duties of another agent (handover), participating in coordinated activity with another agent, or controlling another agent (assessing, directing, correcting). To shift between autonomy and interdependence, autonomous agents need to provide not only variable autonomy but also information to support coordination and control. Some of this information will be transmitted during mixed initiative interactions. In mixed initiative interaction, any of the players can steer communication and actions. Any can guide or be guided; any can help or be helped; any can be interrupted.

To handle anomalies, human operators will direct agent responses and reconfiguration. Mixed initiative interaction can enable loose commanding (directions that are abstract, high-level, sketched, incomplete). By opening the dialogue with a high-level command, the intervening human agent recognizes the independent knowledge of the system management agent, and depends on that agent to help develop a complete solution (Myers, 1997). By using mixed initiative communication and simulation of coordinated actions, the agent can work with the intervening human to develop a joint plan that avoids accidents. The agent can reveal and discuss goals, operational assumptions and constraints that might otherwise be missed.

Initial work on mixed initiative interaction has focused on establishing common understanding of when and under what conditions an intervention is desired and can occur. In the dialogue, the operator orders a test, and the agent asks whether to do it now or plan for it. When told to do it now, the agent confirms that it is possible to do it now, and volunteers the information about the required conditions that hold. If these required conditions change during the test, the agent can abort the test and volunteer the reason why. This dialogue orients the operator to what is necessary, what will happen and what is possible.

What types of abstract and sketched direction will be needed for anomaly response? The agent will need to guide and help the human intervener, while accommodating the request. When will the agent need requests to wait, requests for more information, or need to recommend or volunteer information about capabilities, policies, priorities, safety, and situations?

Safety Conscious Agent: Simulation for checking response plans and resuming operations

Safety issues become especially important when the managed system has changed in new and unexpected ways. Policy and rules about safety are more than just plan maintenance conditions. It is important that safety mechanisms be independent and centralized and designed to enforce safety policy (Leveson, 1995). When a human team member and an agent are working to jointly complete commands or plans for resumption from interrupted operations, safety issues need to be considered.

System accidents have frequently involved systems getting out of synchronization, when automation software resumes operation with stale or partly updated assumptions. A safety conscious agent needs to track, predict and expose assumptions and conditions for proposed operations. The agent should be able to operate safely after one or more anomaly episodes that include testing and workarounds and possible disabling of parts of agent functionality. The agent should be able to identify and communicate relevant changed assumptions and capabilities. One way this can be accomplished is by using model-based simulation to evaluate proposed operations. Simulation cases would investigate scenarios that include the hypothesized causes and impacts of the anomaly.

Acknowledgement

Thanks to Debra Schreckenghost, who designed and implemented the mixed initiative interaction case described in the paper. Thanks to Carroll Thronesbery and Kathy Johnson-Throop who have discussed these ideas and contributed to them.

References

Brooks, R.A. 1991. How to Build Complete Creatures rather than Isolated Cognitive Simulators. In K. VanLehn, ed., *Architectures for Intelligence*, 225-239, Lawrence Erlbaum Associates.

Espinosa, J. A., Carley, K. M., Kraut, R. E., Lerch, F. J., and Fussell, S. R. 2002. The Effect of Task Knowledge Similarity and Distribution on Asynchronous Team Coordination and

Performance: Empirical Evidence from Decision Teams. In *Proceedings of the Cognitive Research (CORE) Conference*, Barcelona, Spain.

Gutwin, C., and Greenberg, S. 1999. A Framework of Awareness for Small Groups in Shared-workspace Groupware, Technical Report, 99-1, Dept. of Computer Science, Univ. of Saskatchewan.

Leveson, N. 1995. *Safeware: System Safety and Computers*. Reading, Mass.: Addison-Wesley.

Malin, J. T., and Johnson, K. 2003. Information for Successful Interaction with Autonomous Systems. In *Human Interaction with Autonomous Systems in Complex Environments: Papers from the AAAI Spring Symposium*. AAAI Press, Technical Report, SS-03-04, 129-133.

Myers, K. L. 1997. Abductive Completion of Plan Sketches. [AAAI/IAAI 1997](#): 687-693.

Myers, K. L., and Morley, D. N. 2001. Human Directability of Agents. In *Proceedings of 2001 International Conference on Knowledge Capture (K-CAP '01)*. ACM, 108-115.

Norman, D. A. 1981. Categorization of Action Slips. *Psychological Review* 88(1):1-15.

Rasmussen, J. 1982. Human Errors. A Taxonomy for Describing Human Malfunction in Industrial Installations. *Journal of Occupational Accidents* 4:311-333.

Reason, J. 1990. *Human Error*. Cambridge: Cambridge University Press.

Wilkins, D., Lee, T. J., and Berry, P. 2003. Interactive Execution Monitoring of Agent Teams. *Journal of Artificial Intelligence Research*, 18:217-261.