

## Automatic safe learning of UML Agents — Abstract —

**Johann Schumann**

RIACS/NASA Ames

email:schumann@email.arc.nasa.gov

Joint work with Jon Whittle, QSS/NASA Ames

Nowadays, agent-based technology is used in many application areas; in domains where safety and/or security issues are of great importance (e.g., aviation, traffic control, medicine), however, the design and implementation of safe agents is far from trivial. This is especially true when an agent needs to learn or adapt after its deployment.

In this talk, we will present an approach for design and adaptation of agents within the object-oriented framework UML. There, the execution of behavioral activities can modeled and implemented as UML statecharts. In our approach, an initial design can be automatically generated from a set of given scenarios in the form of sequence diagrams. All scenarios are automatically checked for consistency with a safety-policy, formulated as a set of OCL constraints. This initial design obviously obeys the safety-policy. When, after deployment of the agent, (supervised) learning has to take place, scenarios (UML sequence diagrams) describing new required behavior are sent to the agent. The agent then uses basically the same synthesis algorithm as for its initial design to extend and modify its internal statechart(s). At each step, consistency with the OCL constraints is checked, thus guaranteeing that the safety-properties are always obeyed. Because our approach is strongly tied to the OO/UML paradigm, and thus to weak agency, it is mainly intended for lower levels of agent designs, e.g., for inter-agent protocols or reactive components.

This approach has the advantage that the learning phase smoothly fits into the overall agent lifecycle: after the return of a trained agent, the information learned is directly visible and available in the refined design. This design (a highly structured, readable statechart) then can directly be used for further deployments. In this presentation, we will demonstrate our approach at design and on-the-fly refining of an inter-agent protocol, and will discuss current limitations with respect to the safety policy to ensure tractability of the synthesis algorithm as well as future enhancements.