Safety of Agents (SagA)

Supervisors: **Martin Steffen** and **Olaf Owe** and **Christian Johansen**
Collaborators: **Chalmers (SE)**, **Liverpool (UK)**, **Tellu AS**

1 Main objective and summary of the project

THE THEME of this project are the future information agents, also known as ubiquitous or pervasive agents. Ubiquitous information agents have computation abilities and the power to act on their physical environments. Examples include complex systems such as robot companions, self-driving cars (e.g., Stanford or Google), simpler agents as in smart elderly homes and AAL, mobile travel devices, smart refrigerators, or electronic transactions agents. Ubiquitous agents operate in the living environments of people, interacting with (helping) people in their daily life. In consequence, the safety of the decisions and actions of such information agents is of paramount importance.

THE OBJECTIVE is to automatically control ubiquitous agents for achieving a desired goal under the restrictions of safety contracts. This would imply a paradigm shift, from the current rule-based systems to goal-based systems.

Example: consider a system that shall contain patients in a certain area, unless they are in view of a nurse.

- A rule-based system would require the human expert to describe through rules how to restrict the movements of the patients and nurses in order to achieve this goal.
- A goal-based system would simply infer/synthesise suitable rules by itself from the goals set by the user.

Safety contracts

will be used to describe restrictions on the allowed behaviours of the agents. Safety contracts are essential for making sure that the agents do not harm the humans or damage their operating environment. The behaviour of agents is supposed to always conform with their contracts.

Ideally, the project leads to a user friendly way of defining the goals and a quality measure, thus allowing non-experts to use the system. For defining the safety contracts, we target semi-experts like system administrators or decision makers. To incorporate legacy systems, we will allow rules to be described as particular types of goals.

THE MINIMUM VIABLE OUTCOMES in terms of research and innovation are:

- **a formalism** for expressing the goals of ubiquitous agents and the safety contracts;
- **algorithms** for verifying the consistency of goals and inferring a rule system to control the agents; and
- **a prototype** that implements such algorithms, integrated in the SmartTracker rule-based framework of the SME partner Tellu.

2 Project background and scientific basis

Ubiquitous information agents have started to pervade our daily lives, albeit in simplistic forms s.a., travel guides or entertainment systems. But more complex agents are envisaged to penetrate our society, facilitated also by the large funds allocated to programs like Ambient Assisted Living (AAL) or smart cities. This is well motivated by the promise of these systems to reduce the economic impact that the increasingly ageing European society is announced to provoke. Complex agents like self-driving cars or robot companions for elderly or dementia have extended computing and decision capabilities, as well as the power to act on their surrounding environment, with the purpose to help humans in their daily chores. The technological advances in this direction are impressive, but the slow adoption of such systems is partly due to the lack of trust in the safety of these systems, which is understandable when thinking of how many 'bugs' common software has.

3 Research questions and scientific challenges

THE CHALLENGES taken up in this project are meant to advance the current technology towards the goal of producing controllers for ubiquitous agents to be safe w.r.t. predefined safety guidelines. At the same time, we want to elevate the burden of designing controllers at a low level, e.g., by using rule-based systems, to a more high-level method. This high level method will focus on the goal (rather than how to achieve it), thus removing a main source of errors.

Our endeavours can be seen as trying to answer the following question:

How can we develop safer and more complex information agents that are trustworthy and widely accepted?

The research and innovation efforts in this project will focus on achieving the following objectives:

Objective 1. develop languages for defining safety contracts and goals for information agents, expressive enough for the practical needs identified together with our partner organisations Tellu and ITU Copenhagen.

Objective 2. develop algorithms to automatically obtain controllers from high-level descriptions.

Objective 3. test the effectiveness of the algorithms through prototyping, and their compatibility with existing systems through integration with the SmartTracker rule-based framework of our SME partner Tellu. Test the user-friendliness of the high-level specification languages in real applications.

4 Scientific methods to be employed

Many expert systems (i.e., systems taking more complex decisions) and controllers for autonomous agents or robots (i.e., making plans of actions) are based on rule-engines and do rule-based reasoning. Therefore, rule-based systems make our starting point, and the choice of our SME partner reflects this, as their main product SmartTracker has a rule-engine at its core. But a major drawback, which Tellu and their clients confirmed, is the high degree of expertise such rule-based systems require of their users. Clients find it difficult to express correctly what their system should do, using rules. This leads to errors in designs which diminish the safety their AAL systems should have.

We replace rule-based systems by goal-based systems and hybrids between both. The difference between them is:

- **A rule-based system** describes the restrictions and actions to be taken in order to achieve the goal.
- A goal-based system simply states the goal and infers suitable rules by itself.

OUR APPROACH can be summarised as defining goal-based systems as a more usable alternative to rule-based systems. Our goal-based system has specification languages for describing goals and safety contracts, which are fed into a synthesis algorithm to produce rule-based descriptions of controllers that respect the safety contracts and direct the agent to meet the defined goal. To make sure that our goal-based system is compatible with existing legacy products, we will integrate the developed algorithms with the SmartTracker tool suite of Tellu. Thus,

- **A language for describing goals** needs to be developed, which is high-level and can be used by people with limited engineering expertise. The knowledge of the host researchers in this area will be heavily exploited.
- **Safety contracts will explicitly define safety requirements**, which will no longer be hard-coded in rule-based implementations. To achieve this, we will devise models and logics, starting from the Fellow's work.
- **Expressiveness and usability** are the two key aspects we have in mind while developing this formalism.
- **A synthesis algorithm** will take input using the above languages for describing goals and safety contracts, with the purpose to automatically generate optimal or near optimal descriptions of behaviours for the ubiquitous agents. The field of synthesis is generally concerned with generating controllers in the form of classic finite state machines. In this project, we will investigate how to generate rule-based descriptions of controlle from goals. The expertise and prior work of the supervisor is the starting point for this part of the project.

ADVANTAGES for switching to a goal-based system are multiple, e.g.:

- **Detecting inconsistency in the goals**, which at an early stage may result in re-thinking the goals, or in planning the resources differently. The change in the paradigm, however, is a general one: instead of checking whether or not a rule system complies with the goals, we can check if these goals are satisfiable.
- **It allows for optimisation**, i.e., besides defining goals, we can also define a measure for the quality of a solution. It is then natural to seek a rule system that offers an optimal control for an agent.

5 Project timeline

Progress through milestones. **M1** (month 1): **Candidate fully established**; **M2** (m.6): **Training activities** under way / **Prototyping started** / **Safety contracts**; **M3** (m.12): **Algorithms for synthesis** / **Tool integration** / Follow-up drafting; **M4** (m.18): **Languages for Specifications**; **M5** (m.24): Prototype finalizing. **M6** (m.48): PhD completed.

Secondments and collaboration plans:

S1 (m2+m11; at **Tellu**) Identifying case-studies in eHealth and current best-practices; Learning the Smart-Tracker rule-based engine and applications;
S2 (m6-m8; at **Chalmers**) Working on Contracts and goals
S3 (m16-m19at **Liverpool**) Working on synthesis of safety controllers.

Tellu AS develops software solutions meant to provide simple forms of safety. Their solutions have been deployed in, e.g., security companies, high-risk sports competitions, emergency hospital wards, or for people with disabilities. The purpose of these secondments is to provide a development medium where the prototyping would take place, as well as offer hands-on experience with existing technology deployed in existing applications of ubiquitous agents.

Liverpool University, with Dr. Sven Schewe, placed at Centre for Autonomous Systems Technology and the Logic and Computation Group. The group is internationally renowned for its work on BDE systems, automata and game theory, modal and temporal logics and their application in AI and computer science. Sven Schewe is an expert in the analysis and automatic construction of reactive computer systems and protocols. Sven's research has focused on the construction of optimal control strategies. His PhD thesis on the automated construction of distributed controllers has won the prestigious GI dissertation award and the interdisciplinary Dr. Eduard Martin prize.

Chalmers University, with Prof. Gerardo Schneider and the group on Formal Methods, for collaboration and training on tools and techniques for the analysis and implementation of electronic contracts. Gerardo has a broad range of expertise, including formal specification and analysis of contracts, formalization of privacy policies, model checking, verification of embedded systems, semantics, logics for computer science.