Autonomous System Safety: Model-Checker Toolkit

Providing guarantees against risk of faults for autonomous systems

Overview

We need to build our trust for autonomous and partially-autonomous machines, with the advent of their proliferation. Unlike more standard automatic machines, autonomous systems are capable of: reasoning, independent planning, decision making and learning. Systems are becoming increasingly powerful, yet this autonomy comes with the trade-off in more bugs and many possible unpredictable outputs; hence *risks*.

Risks compound as autonomous systems must not only safely operate on their own, yet they must also co-exist with humans, as well as other autonomous machines. We thus need to be able to *guarantee* their behaviour to ensure such systems are safe. A toolkit has been developed by computer scientists at Imperial College London that tackles risk in autonomous systems through verification. Verification provides guarantees in the reliability against faults as well as against cascading of faults of autonomous machines.

Technology

Verification allows us to predict how machines will behave in all possible circumstances, irrespective of the inputs and of other interacting machines, even before the machines are deployed. The challenge arises due to the number of possible states of an *autonomous* system. Whilst a relatively simple, *automatic* system like a washing machine has 10^2 possible states, an autonomous system can have **10^{90}** possible states (by comparison, there are about 10^{80} atoms in the universe).

Model checking has been adopted by computer scientists at Imperial College to overcome the challenge of an unmanageably high number of unpredictable output states. In the past, model checking has won the coveted Turing Award, and can verify state spaces of 10^{20}. Symbolic model checking is now mature technology being used in industry for applications such as verifying computer chips, communications protocols and aircraft autopilot systems. However, it has not yet been successfully implemented for greater complexity such as smartphones or IoT systems. Difficulty compounds in messy, 'real-world' applications with an unknown number of autonomous agents, working collaboratively, and acting as independent agents. The space for verification becomes very large.

The Imperial researchers have successfully developed a software toolkit based on model checking that can be used to 'close the trustworthiness gap' with autonomous systems. The toolkit represents three methods:

1) Predicate Abstraction: Starting with complexity, automatically creating a model of the core components.

2) Parameterised Model Checking: Tackling verification of enormous swarms; it can give guarantees of millions of agents by analysing two - three free agents.

3) Epistemic Verification: Instead of verifying machine characteristics, we verify human-like attributes – e.g. beliefs, intentions, rules, regulations etc.

The use of these methods will enable the risk of autonomous systems to be better understood and verified, and as a result the dangers can be reduced. **A proof-of-concept has been deployed in autonomous submarines, and the toolkit can be customised for various applications.**

Benefits

- Help with prediction of insurance risk for autonomous systems
- Reduced risk of expensive product recalls & damage claims
- Fewer systems blocked from being put in production due to unknown output states
- Verification leading to certification of autonomous systems

Applications

- Insurance
- Certification bodies & regulators
- Agriculture
- Manufacturing
- Transport (inc. automotive, aircraft, aerospace)
- Health & Ambient Assisted Living
- Energy (inc. decommissioning)
- Autonomous space exploration

Alessandro Garcia
Technology Licensing Executive
e: alessandro.garcia@imperialinnovations.co.uk
t: +44 (0)20 3727 2059
w: www.imperialinnovations.co.uk

Technology reference number: 8322