Carnegie Mellon University
Computer Science Department

THESIS ORAL

NATHAN FULTON

Verifiably Safe Autonomy for Cyber-Physical Systems

Friday, November 9, 2018 – 10:00 a.m. – GHC 6501

This thesis demonstrates that autonomous cyber-physical systems that use machine learning for control are amenable to formal verification.

Cyber-physical systems, such as autonomous vehicles and medical devices, are increasingly common and increasingly autonomous. Designing safe cyber-physical systems is difficult because of the interaction between the discrete dynamics of control software and the continuous dynamics of the vehicle's physical movement. Designing safe autonomous cyber-physical systems is even more difficult because of the interaction between classical controls software and machine learning components.

Formal methods capable of reasoning about these hybrid discrete-continuous dynamics can help engineers obtain strong safety guarantees about safety-critical control systems. However, existing theory and tooling does not explain how to obtain formal safety guarantees for systems that use reinforcement learning to discover efficient control policies from data. This gap in existing knowledge is important because modern approaches toward building cyber-physical systems combine machine learning with classical controls engineering to navigate in open environments.

This thesis introduces KeYmaera X, a theorem prover for hybrid systems, and uses KeYmaera X to obtain verified safety guarantees for control policies generated by both model-based and model-free reinforcement learning algorithms. These contributions enable strong safety guarantees for optimized control policies when the underlying environment matches a first-principles model and also explain how to obtain strong safety guarantees for systems even without an accurate first-principles model. These contributions provide verifiable safety guarantees for systems that are controlled by policies obtained through reinforcement learning, justifying the use of reinforcement learning in safety-critical settings.

Thesis Committee:
Andre Platzer, Chair
Jeremy Avigad
Zico Kolter
Stefan Mitsch
Goran Frehse, Université Grenoble