ISTeC
The Information Science and Technology Center

COLORADO STATE UNIVERSITY

Distinguished Lectures

Fall 2024

Dr. Corina Pasareanu

Principal Scientist, CMU CyLab, and
Technical Professional Leader - Data Science
NASA Ames/KBR

Compositional Verification and Run-time Monitoring for Learning-Enabled Autonomous Systems

Monday, Oct. 7, 2024
Reception with Refreshments: 10:30 a.m.
Lecture: 11:00 a.m. - 12:00 noon
Lory Student Center Rm. 386

Attacks and Defenses for Large Language Models on Coding Tasks

Tuesday, Oct. 8, 2024
Lecture: 10:00-10:55+ a.m.
CSB 130

Sponsored by
Colorado State University's Information Science and Technology Center (ISTeC)

In conjunction with the Department of Computer Science and Department of Electrical and Computer Engineering Seminar Series

Abstracts

Compositional Verification and Run-time Monitoring for Learning-Enabled Autonomous Systems

Providing safety guarantees for autonomous systems is difficult as these systems operate in complex environments that require the use of learning-enabled components, such as deep neural networks (DNNs) for visual perception. DNNs are hard to analyze due to their size, lack of formal specifications, and sensitivity to small changes in the environment. We present compositional technigiques for the formal verification of safety properties of such autonomous systems. The main idea is to abstract the hard-to-analyze components of the autonomous system, such as DNN-based perception and environmental dynamics, with either probabilistic or worst-case abstractions. This makes the system amenable to formal analysis using off-the-shelf model checking tools, enabling the derivation of specifications for the behavior of the abstracted components such that system safety is guaranteed. We also discuss how the derived specifications can be used as run-time monitors deployed on the DNN outputs. We illustrate these ideas in a case study from the autonomous airplane domain.

Attacks and Defenses for Large Language Models on Coding Tasks

Modern large language models (LLMs), such as ChatGPT, have demonstrated impressive capabilities for coding tasks, including writing and reasoning about code. They improve upon previous neural network models of code, such as code2seq or seq2seq, that already demonstrated competitive results when performing tasks such as code summarization and identifying code vulnerabilities. However, these previous code models were shown vulnerable to adversarial examples, i.e., small syntactic perturbations designed to "fool" the models. In this talk we discuss the transferability of adversarial examples, generated through white-box attacks on smaller code models, to LLMs. Further, we propose novel cost-effective techniques to defend LLMs against such adversaries via prompting, without incurring the cost of retraining. Our experiments show the effectiveness of the attacks and the proposed defenses on popular LLMs.

Speaker Biography

Corina Pasareanu is an ACM Fellow and an IEEE ASE Fellow, working at NASA Ames. She is affiliated with KBR and Carnegie Mellon University's CyLab. Her research interests include model checking, symbolic execution, compositional verification, probabilistic software analysis, autonomy, and security. She is the recipient of several awards, including ETAPS Test of Time Award (2021), ASE Most Influential Paper Award (2018), ESEC / FSE Test of Time Award (2018), ISSTA Retrospective Impact Paper Award (2018), ACM Impact Paper Award (2010), and ICSE 2010 Most Influential Paper Award (2010). She has been serving as Program / General Chair for several conferences including: ICSE 2025, SEFM 2021, FM 2021, ICST 2020, ISSTA 2020, ESEC/FSE 2018, CAV 2015, ISSTA 2014, ASE 2011, and NFM 2009. She is on the steering committees for the ICSE, TACAS and ISSTA conferences. She is currently an associate editor for IEEE TSE and for STTT, Springer Nature.

ISTeC (Information Science and Technology Center) is a university-wide organization for promoting, facilitating, and enhancing CSU's research, education, and outreach activities pertaining to the design and innovative application of computer, communication, and information systems. For more information please see ISTeC.ColoState.edu.

To arrange a meeting with the speaker, please contact Ravi Mangal <Ravi.Mangal@colostate.edu>.