Electrical & Computer Engineering

S E M I N A R

Louisiana State University

Securing Computer Systems Using AI Methods and for AI Applications

Mulong Luo

The University of Texas at Austin

Abstract—Securing modern computer systems against an ever-evolving threat landscape is a significant challenge that requires innovative approaches. Recent developments in artificial intelligence (AI), such as large language models (LLMs) and reinforcement learning (RL), have achieved unprecedented success in everyday applications. However, AI serves as a double-edged sword for computer systems security. On one hand, the superhuman capabilities of AI enable the exploration and detection of vulnerabilities without the need for human experts. On the other hand, specialized systems required to implement new AI applications introduce novel security vulnerabilities.

In this talk, I will first present my work on applying AI methods to system security. Specifically, I leverage reinforcement learning to explore microarchitecture attacks in modern processors. Additionally, I will discuss the use of multi-agent reinforcement learning to improve the accuracy of detectors against adaptive attackers. Next, I will highlight my research on the security of AI systems, focusing on retrieval-augmented generation (RAG)-based LLMs and autonomous vehicles. For RAG-based LLMs, my ConfusedPilot work demonstrates how an attacker can compromise confidentiality and integrity guarantees by sharing a maliciously crafted document. For autonomous vehicles, I reveal a software-based cache side-channel attack capable of leaking the physical location of a vehicle without detection. Finally, I will outline future directions for building secure systems using AI methods and ensuring the security of AI systems.

When: Thursday, **13 February 2025,** 10:00 - 11:00
Where: Room 3316E Patrick F. Taylor Hall
Info: `https://www.lsu.edu/eng/ece/seminar`
Food: ***Coffee and doughnuts (or the nutritional equivalent) will be served.***