TEXAS A&M UNIVERSITY
Department of Electrical & Computer Engineering

TRANSFORMING ENGINEERING EDUCATION

ENERGY & POWER GROUP SEMINAR

From Pixels to Policies: Securing Multi-Agent Systems Against Adversarial Attacks

Abstract

As multi-agent systems increasingly rely on machine learning (ML) models for perception and control, the vulnerability of ML models to adversarial attacks becomes a critical threat to system safety. The design space for ensuring resilience in such systems involves not only achieving robustness to adversarial perturbations, but also mechanisms to detect and adapt to attacks across multiple levels of the system. At the level of agents, ML models must detect adversarial attacks in real-time and at the same time they should perform well despite adversarial attacks. At the system level, agents should be able to detect and identify anomalous agents in a decentralized manner to achieve situational awareness. In this talk, we explore recent advances in making key computer vision tasks, object detection and classification, robust to physically realizable adversarial attacks through ensemble saliency analysis and ensemble guided reconstruction. Focusing on the system level, we then examine how agents can identify adversarial behavior in their peers through decentralized anomaly detection in cooperative multi-agent reinforcement learning, in the framework of sequential hypothesis testing. We present methods that enable agents to detect and respond to attacks in complex discrete as well as continuous-action environments. Together, through these lines of work we aim to highlight a broader principle: in safety-critical settings it is not sufficient for learning systems to be robust to adversarial activities, they must also detect and respond to adversarial input, ensuring true resilience.

György Dán
Professor, Teletraffic Systems
KTH Royal Institute of Technology

Friday, November 21
11:30 am
241 ZACH

Biography

György Dán is professor of teletraffic systems at KTH Royal Institute of Technology, Stockholm, Sweden. He received the M.Sc. in computer engineering from the Budapest University of Technology and Economics, Hungary in 1999, the M.Sc. in business administration from the Corvinus University of Budapest, Hungary in 2003, and the PhD in Telecommunications from KTH in 2006. He worked as a consultant in the field of access networks, streaming media and videoconferencing 1999-2001. He was a visiting researcher at the Swedish Institute of Computer Science in 2008, a Fulbright research scholar at University of Illinois at Urbana-Champaign in 2012-2013, and an invited professor at EPFL in 2014-2015. He served as area editor of Computer Communications 2014-2021, as editor of IEEE Transactions on Mobile Computing 2019-2023, serves on the TPC of conferences like IEEE Infocom and ACM e-Energy, and is vice-chair of the steering committee of IEEE SmartGridComm. He has received several best paper awards from IFIP and IEEE in recent years. His research interests include the design and analysis of mobile computing systems, game theoretical models of networked systems, and cyber-physical system security and resilience.

FACULTY CONTACT:
Xin Chen
xin_chen@tamu.edu