Computer Science
Seminar

Trustworthy Machine Learning: From Theory to Practice

Dr. Yuan Hong, IIT
Illinois Institute of Technology

Abstract: Machine learning has achieved many big innovations in all industries and significantly impacted our daily lives. However, machine learning can also result in severe security and privacy risks. In this talk, I will present our recent works on both theory and applications that fundamentally contribute to machine learning security and privacy. First, we designed the first differential privacy mechanism (R2DP) that universally optimizes the randomization for the maximum utility w.r.t. any utility metric. It fundamentally improves the utility of differential privacy mechanisms in all the relevant applications, such as statistical queries, classification, social network analysis, and deep learning. Second, we propose the first black-box attack framework that generates universal 3-dimensional (U3D) perturbations to subvert a wide variety of video deep neural networks (DNNs). The new attack is easy-to-launch, universal, transferable, and human-imperceptible. It can also bypass the state-of-the-art defense methods. Such novel attack motivates the video recognition systems to build and integrate more robust DNN models. ¡br¿¡br¿ Biography: Yuan Hong is an Assistant Professor of Computer Science and Cybersecurity Program Director at Illinois Institute of Technology. He received his Ph.D. degree from Rutgers University in 2014. His research interests primarily lie in the fields of security, privacy, optimization, and data science, such as differential privacy, secure multiparty computation, applied cryptography, adversarial learning, and certified robustness. He is a recipient of the NSF CAREER award, and his work has appeared in prestigious security and data science venues such as Oakland, CCS, PETS, AAMAS, CIKM, EDBT, ICDCS, TDSC, TKDE, TIFS, and TOPS. His research is supported by multiple NSF and AFOSR awards.

Friday, September 17, 2021, 1:00 pm
https://emory.zoom.us/j/98352727203

Computer Science
Emory University