CS4274 Secure Computing Capstone

Fall 2021

1 Course description

This course will focus on emerging threats impacting machine learning (ML) systems powering large online services today. Students will focus on: (1) Designing and building tools to defend against attacks that aim to evade state-of-the-art ML security defenses in the computer vision, NLP, and networking domains. Attackers can craft adversarial inputs targeting such defenses by carefully perturbing input samples (with or without knowledge of the defense system), and/or by tampering with the ML system (poisoning attacks). You will focus on building robust defenses against such attacks. (2) Designing and building tools to defend against *deepfakes* or ML-generated synthetic content. We heavily rely on ML systems that continually scan large online platforms to detect and filter unwanted content (text, images, videos) used for phishing attacks or to spread disinformation, and other malicious online activities. Today, such online filters are under threat because advances in deep learning have enabled automatic, controlled generation of convincing synthetic content or deepfakes that can be misused for online abuse. You will develop tools to robustly detect such deepfake content.

At the end of the class, you will have experience reasoning about security of ML systems, building defense tools effective against an adaptive adversary, and also learn about real-world applications of deep learning to improve security.

2 Reference materials

Most reading material will be drawn from research papers published at venues such as IEEE S&P, USENIX Security, CCS, NDSS, IMC, SIGCOMM, NSDI, CoNeXT, WWW, WSDM, and KDD.

3 Prerequisites

Prerequisites for the course include courses on information systems, data mining, machine learning, and computer security. Students are expected to have a basic understanding of data mining and machine learning algorithms, computer networks and security. Knowledge of a scripting language such as Python is essential. Knowledge of deep learning frameworks such as PyTorch or Tensorflow is required, and students are expected to learn how to use such tools, if not already familiar.

4 Grading

Grading will be based on a term-long project, which will include the following components:

- Oral presentations
- Project reports
- Teamwork