UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc.CS Seminar

Wednesday May 12, 2021 at 1pm on Zoom

Detecting Adversarial Attacks on Image-based and Graph-based Malware Classifiers

Bardia Esmaeili

Advisor: Dr. Ali Dehghantanha
Co-Advisor: Dr. Hadis Karimipour [Engineering]

ABSTRACT:

Machine and deep learning techniques have exponentially boosted state-of-the-art performance in many areas including computer vision, natural language processing, speech recognition, healthcare, and cybersecurity. Although deep models are the leading approaches in various fields, they are prone to a class of vulnerabilities, known as adversarial attacks. In this work, we propose an adversarial detection framework for the malware classification task.

While adversarial detection in malware classification has been explored before, it has not been studied against graph-based and image-based malware classifiers. Hence, we seek to propose a detection mechanism against attacks on such classifiers as our main contributions. Each branch of the framework contains a detection model that filters out adversarial samples and feeds the remaining samples to the corresponding graph-based or image-based classifier. We further provide a novel malware graph dataset as a secondary contribution, which has also been used for the development of our graph-based branch.