Department of Computer Science
香港城市大學
City University of Hong Kong

COMPUTER SCIENCE COLLOQUIUM

Model Stealing Attacks and Defenses: Where are we now?

SPEAKER **Prof. N. ASOKAN**

Professor of Computer Science; David R. Cheriton Chair
University of Waterloo

DATE 19 Sep, 2024 (Thu)
TIME 10:30 AM - 11:30 AM
VENUE Y6405, 6th Floor, Yeung Kin Man Academic Building, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong

ABSTRACT

The success of deep learning in many application domains has been nothing short of dramatic. This has brought the spotlight onto security and privacy concerns with machine learning (ML). One such concern is the threat of model theft. I will discuss our work on exploring the threat of model theft, especially in the form of "model extraction attacks" -- when a model is made available to customers via an inference interface, a malicious customer can use repeated queries to this interface and use the information gained to construct a surrogate model. I will also discuss possible countermeasures, focussing on deterrence mechanisms that allow for model ownership resolution (MOR). I will touch on the issue of conflicts that arise when protection mechanisms for multiple different threats need to be applied simultaneously to a given ML model.

BIOGRAPHY

N. Asokan is a professor of computer science and a David R. Cheriton Chair at the University of Waterloo where he also serves as the executive director of the Cybersecurity and Privacy Institute. Asokan is a Fellow of the ACM, the IEEE, and the Royal Society of Canada. His research focuses on systems security. More information about his work is on his website at https://asokan.org/asokan/ or via X/Twitter @nasokan.

All are welcome!

In case of questions, please contact Prof. WANG Cong at congwang@cityu.edu.hk, or visit the CS Departmental Seminar Web at https://www.cs.cityu.edu.hk/events/cs-seminars/recent-cs-colloquiums.