Good morning,
You are invited to attend our weekly ECE Graduate Seminar.

Old Dominion University
College of Engineering and Technology
Department of Electrical and Computer Engineering

All lectures to be held at 3:00pm on Fridays online at ODU DL: ECE 731 831 Grad Seminar

For more information, contact Dr. Chung Hao Chen at (757) 683-3475 or email cxchen@odu.edu.

Friday, October 15, 2021 Seminar Topic:

CLEAR: CLEAN-UP SAMPLE-TARGETED BACKDOOR IN NEURAL NETWORKS by Dr. Rui Ning, Ph.D. Research Assistant Professor in the Center for Cybersecurity Education & Research at Old Dominion University

Abstract:
The data poisoning attack has raised serious security concerns on the safety of deep neural networks, since it can lead to neural backdoor that misclassifies certain inputs crafted by an attacker. In particular, the sample-targeted backdoor attack is a new challenge. It targets at one or a few specific samples, called target samples, to misclassify them to a target class. Without a trigger planted in the backdoor model, the existing backdoor detection schemes fail to detect the sample-targeted backdoor as they depend on reverse-engineering the trigger or strong features of the trigger. In this paper, we propose a novel scheme to detect and mitigate sample-targeted backdoor attacks. We discover and demonstrate a unique property of the sample-targeted backdoor, which forces a boundary change such that small "pockets" are formed around the target sample. Based on this observation, we propose a novel defense mechanism to pinpoint a malicious pocket by "wrapping" them into a tight convex hull in the feature space. We design an effective algorithm to search for such a convex hull and remove the backdoor by fine-tuning the model using the identified malicious samples with the corrected label according to the convex hull. The experiments show that the proposed approach is highly efficient for detecting and mitigating a wide range of sample-targeted backdoor attacks.

Bio:
Rui Ning received BS in Computer Science and Engineering from Lanzhou University, China in 2011, MS in Computer Science from the University of Louisiana at Lafayette in 2016, and Ph.D in Electrical & Computer Engineering from Old Dominion University in 2020. He is a current research assistant professor in the Center for Cybersecurity Education & Research, Old Dominion University. His research interests include cybersecurity, secure AI, and machine learning-based side-channel attacks. He received the Mark Weiser Best Paper Award at the IEEE International Conference on Pervasive Computing and Communication (PerCom) 2018, IEEE INFOCOM 2019 Best In-session Presentation Award, and ECE Graduate Student Award, Ph.D. Researcher of the Year, 2019.