University of California, Irvine
Statistics Seminar

Data-Driven Label-Poisoning Backdoor Attack

Xuan Bi
Assistant Professor
University of Minnesota

4 p.m., Thursday, April 27, 2023
6011 Donald Bren Hall

Backdoor attacks, which aim to disrupt or paralyze classifiers on specific tasks, are becoming an emerging concern in several learning scenarios, e.g., Machine Learning as a Service (MLaaS). Various backdoor attacks have been introduced in the literature, including perturbation-based methods, which modify a subset of training data; and clean-sample methods, which relabel only a proportion of training samples. Indeed, clean-sample attacks can be particularly stealthy since they never require modifying the samples at the training and test stages. However, the state-of-the-art clean-sample attack of relabeling training data based on their semantic meanings could be ineffective and inefficient in test performances due to heuristic selections of semantic patterns. In this work, we introduce a new type of clean-sample backdoor attack, named as DLP backdoor attack, allowing attackers to backdoor effectively, as measured by test performances, for an arbitrary backdoor sample size. The critical component of DLP is a data-driven backdoor scoring mechanism embedding in a multi-task formulation, which enables attackers to simultaneously perform well on the normal learning tasks and the backdoor tasks. Systematic empirical evaluations show the superior performance of the proposed DLP to state-of-the-art clean-sample attacks.

Dr. Xuan Bi is an Assistant Professor of Information and Decision Sciences at the Carlson School of Management at the University of Minnesota. His research mainly revolves around personalization, with a special interest in recommender systems and data privacy. His works have been published in leading academic journals, including *Journal of the American Statistical Association, Annals of Statistics*, and *Journal of Machine Learning Research.* Dr. Xuan Bi holds a Bachelor of Science in Mathematics from Tsinghua University, and a Ph.D. in Statistics from the University of Illinois at Urbana-Champaign. Prior to joining the University of Minnesota, Dr. Xuan Bi was a Postdoctoral Associate at Yale University.