Adversarial Machine Learning Attacks and Countermeasures

Summary

The ML models induced for these critical tasks can be bypassed or manipulated by the adversarial ML attacks. The attackers may create adversarial drifts, The defenders should detect these drifts in addition to the usual drifts occurring in the data distributions. This Ph.D. position requires the development of adversarial attacks and countermeasures for the cyber security solutions (e.g., malware detection, security monitoring) that are based on machine learning models. These countermeasures would include the detection of the adversarial drifts and inducing models that have the resistance to such attacks. It is possible that research may cover other adversarial attacks.

Research field:	Information and communication technology
Supervisors:	Prof. Dr. Hayretdin Bahsi Dr. Risto Vaarandi
Availability:	This position is available.
Offered by:	School of Information Technologies Department of Software Science
Application deadline:	Applications are accepted between June 01, 2022 00:00 and June 30, 2022 23:59 (Europe/Zurich)

Description

Main Motivation and Research Problem

It has been demonstrated in a huge body of literature that machine learning (ML) can be a promising solution in solving cyber security problems [1] such as detection of network attacks, identifying malware, or security review of the software codes. However, the ML models induced for these critical tasks can be bypassed or manipulated by the adversarial ML attacks [2]. Malicious actors can design and launch sophisticated attack campaigns that intelligently merge conventional cyber attacks with adversarial ML attacks to compromise high-stake applications. It is important to defend the ML models, which are supposed to provide security, against these types of attacks.

Although adversarial ML has been widely researched in some application domains (e.g. computer vision), still, there is a significant research gap in the area of cyber security. One of the interesting problems is to develop attacks and countermeasures for the concept drift solutions that handle the evolving nature of the cyber threat landscape. It is essential for ML models to adapt to the changes in the cyber attacks to have an operational system with continuous high performance, thus, they achieve this aim by detecting the drifts in data distribution and retraining the system. On the other side, the attackers may create adversarial drifts in such systems to manipulate the security function [3]. The defenders should also detect these adversarial drifts in addition to the usual concept drifts occurring in the data distributions.

This Ph.D. position requires the development of adversarial attacks and countermeasures for the cyber security solutions (e.g., malware detection, security monitoring) that are based on machine learning models. These countermeasures would include the detection of the adversarial drifts and inducing models that have the resistance to such attacks. It is possible that research may cover other types of adversarial attacks for the problem domain. The university has a real-life organizational security monitoring system which could support the research.

This Ph.D. position is for 4-years.

Support for Teaching and Supervising Activities

We assess that the Cyber Security Master Program requires a machine learning course that is adapted to the specific needs of the program. This course should give the main theoretical background information with practical cyber security implementations. Students who take this course would be able to cooperate with the data scientists in their professional life, explain the needs of the problem domain and understand the possible contributions and limitations of

machine learning methods. Additionally, we plan to create an advanced research course that addresses the students (from Cyber Security Program, Informatics, and other programs as well) who would like to conduct research in line with machine learning and its application to cyber security. The Ph.D. candidate will contribute to the development and running of these courses in addition to supervising MSc and BSc level theses.

Candidate's Background and Knowledge

The candidate should have a solid background regarding machine learning and familiarity with cyber security concepts.

References

1. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surv. tutorials*, vol. 18, no. 2, pp. 1153–1176, 2015.

1. G. I. Webb, R. Hyde, H. Cao, H. L. Nguyen, and F. Petitjean, "Characterizing concept drift," *Data Min. Knowl. Discov.*, vol. 30, no. 4, pp. 964–994, 2016.
2. T. S. Sethi and M. Kantardzic, "Handling adversarial concept drift in streaming data," *Expert Syst. Appl.*, vol. 97, pp. 18–40, 2018.

To get more information or to apply online, visit https://taltech.glowbase.com/positions/519 or scan the the code on the left with your smartphone.