ML4Cyber: Machine Learning for Cyber Security

Instructor: Prof. Radha Poovendran

Email: rp3@uw.edu

Course Introduction

This course will study the use of machine learning for cybersecurity applications. There are many security applications which have large amount of data related to the system as well as adversarial actions. Our ability to identify the type of machine learning algorithms that are useful for specific security applications can help us to improve the defense against attacks and also anticipate the potential attack variants that may arise in the future. Even in the case when one does not know the type of attack, if the machine learning algorithms can identify any anomaly, then the next level of security checks can be performed by other means or experts.

But nothing comes in life for free and this holds for machine learning in the context of cyber too! Another point to remember is "Beware of what you add to your tool bag! You may have adversary manipulating your machine learning itself." This leads to adversarial machine learning where the machine learning could be tricked to fail the detection!

Attacks on Google Video, Toxic Comments, and Google Vision are some of the fun examples where simple modifications make the machine learning algorithms fail!

We will start with setups where the machine learning will be useful for the cybersecurity. As indicated in the title of the course, it will be hands-on course on applying machine learning for cybersecurity applications. Machine learning algorithms will be introduced as needed.

Lectures

1. Introduction to Machine Learning (ML) for Cyber Security
2. Supervised Learning Approach for Email Spam Detection (**Scheduled Release of Homework #1**)
3. Detecting Distributed Denial of Service Attack using Time series Modeling
4. Machine Learning for Solving Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) (**Scheduled Release of Homework #2**)
5. Data Dimensionality Reduction in Cyber Attack Data

6. Network Anomaly Detection Using Clustering Techniques (**Scheduled Release of Homework #3**)
7. Credit Card Fraud and Malicious Event Detection Using Decision Trees
8. Ensemble Learning for Online Ad blocking, Program Binary Analysis, and Credit Card Fraud Detection
9. Adversarial Machine Learning
10. **Student presentations of Course Projects**

Textbook

Hands-On Machine Learning for Cybersecurity by Soma Halder and Sinan Ozdemir; Packt Publisher. University bookstore has copies. UW Library also has copies. You will need to log in with your UW NetID to read for free.

References

1. *AI for CyberSecurity by Alessandro Parisi; Packt Publishers*
2. *Machine Learning for Cybersecurity by Chiheb Chebbi; Packt Publishers*
3. *Machine Learning for Penetration Testing by Emmanuel Tsukerman; Packt Publishers*
4. Adversarial Machine Learning by Vorobeychik and Kantarcioglu; Morgan and Claypool Publishers

Labs

Throughout the whole course we will have eight labs, during which you will learn and practice various machine learning algorithms and use them to solve cyber-security problems. Below is a tentative schedule for all lab sessions:

1. Python basics review and introduction of common data analysis libraries
2. Machine learning pipeline for cybersecurity problems
    - Case study: spam email detection
3. Time series analysis
    - Case study: DDoS network traffic analysis
4. A small step into deep learning and convolutional neural network (CNN)
    - Case study: breaking Captchas with neural network
5. Dimensionality reduction and data visualization
    - Case study: network anomaly detection and visualization
    - Dataset: KDD Cup 1999 dataset (We will reuse these in lab 6)
6. Autoencoder and clustering algorithm
7. Data oversampling and decision tree algorithm

    - Case study: detecting and categorizing network attacks
    - Dataset: Kaggle credit card fraud detection dataset (We will reuse these in lab 8)
8. Ensemble learning

Depending on the actual progress of lab sessions, the last two lectures may not contain lab sessions so that you can focus on the course project.

Homework

We plan to have three homework throughout the whole course. **We won't have any homework in the last two weeks so that you can focus on the project**.

Project

See course project page for more details.

Grading

- Homework: 45% (15% each)
- Project: 55%
    - Proposal: 5%
    - Checkpoint report: 10%
    - Final presentation: 15%
    - Final report: 25%

Course Announcements and Discussions

- Announcements

We will post the course announcements to both canvas and class mailing list

Discussions – Slack

    - We will use Slack channel PMP-595-W-2021 for the course discussions. Feel free use this Slack channel to post your questions on the course material, lectures, homework, projects or answer the posted questions if you can help. Instructor and TAs will also answer these questions promptly.
    - **Slack Link will be Provided**

Course Policy

- Please complete the homework by yourself and do not copy code from others or the internet without understanding what it is doing. If your homework is found identical to others or any sample code snippets on the internet, you will receive zero scores and we are mandated to report it to the College of Engineering.
- You are encouraged to discuss lab and homework content with your classmates both offline and on the discussion board. However, please limit your discussion to ideas only and do not talk about the code. Specifically, do not copy-paste any homework answers into the discussion board.
- You should submit homework and project materials online by the posted due date. Throughout the whole quarter, we will provide you with **one late day credit per homework for a total of 3 late day credits,** which you can use to submit one or more homework without incurring a penalty. If you have used all the three days, however, each additional late day will result in a 20% penalty in the corresponding homework.
- You must submit project reports and presentations strictly on time. Project does not have any credit days. **Any overdue project materials will receive a zero score** unless the instructor has allowed late submission prior to the deadline for some unavoidable cases.
- Washington state law requires that UW develop a policy for the accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The University of Washington policy, including information about how to request an accommodation, is available at Faculty Syllabus Guidelines and Resources. Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form available at https://registrar.washington.edu/students/religious-accommodations-request/ (Links to an external site.).