Yuwei Han
+852 51727524 22041949r@connect.polyu.hk

EDUCATION BACKGROUND

Hong Kong Polytechnic University (PolyU) — Hong Kong
Master of Philosophy in Computer Science — 2022.09 - Present
- **Relevant Coursework:** Optimization Theory, Computational Intelligence

Jinan University (JNU)（"211 project" university of China） — Guangzhou, China
Bachelor of Engineering in Cyberspace Security — 2018.09 - 2022.07
- **Awards and Honors:** JNU Ziwen Inspirational Scholarship, Jinan University Excellent Student First-Class Scholarship, JNU 5A Excellent Student Award Program "Entrepreneurship Star" Pacemaker Award
- **Relevant Coursework:** C Program Design, Computer Composition Principle, Data Structure, Operating System Principle, Computer Network, Compilation Principle, Mathematical Analysis, Probability Theory and Mathematical Statistics, Linear Algebra, Network Security, Applied Cryptography, Machine Learning, Digital Forensics Technology.

University of California, Irvine — California, United States
Exchange Student Project — 2020.8
- **Credit Coursework:** Introduction to Machine Learning, Business Project Management-Data Analysis Branch

University of Lodz — Lodz, Poland
Exchange Student Project — 2019.7
- **Computer Science Workshop:** Studied linear algebra and used JavaScript, HTML, and CSS to build web pages

RESEARCH EXPERIENCE

Secure and Trustworthy Intelligence Laboratory led by Prof. Kai Zhou — Hong Kong
Poisoning attack to graph neural networks — 2022.9- Present
- Proposed a poisoning attack loss called Cost Aware Poisoning Attack Loss(CAPA) which is aware to adversarial perturbation/budget cost. Specifically, CAPA optimizes budget allocation by differentially weighting graph nodes based on their margins. A node with a large positive margin is theoretically more robust to graph perturbations, whereas a node with a pronounced negative margin is misclassified with a high degree of false confidence in its predicted label. This understanding directs us to prioritize nodes with smaller positive margins, given their inherent susceptibility to attacks. When integrated with various poisoning attacks, our empirical evaluations reveal that CAPA markedly bolsters the efficacy of prevailing attack methodology.(e.g. On Cora dataset, the accuracy of graph convolutional network attacked by our approach with 0.05 perturbation rate drops 6% more than Metattack).

Summer Research guided by Prof. Haohan Wang — UIUC,USA(Remote)
Defense of backdoor attack for deep learning models using data distillation — 2023.6- Present
- To design innovative data distillation algorithms aimed at mitigating backdoor threats in image data.

Guangdong Network Security and Privacy Protection Laboratory — Guangdong, China
Location privacy protection algorithms and machine learning security problems — 2019.9 – 2022.6
- Contributed to **National Key Research and Development Program** "Privacy Protection and Forensics Technology under the Internet" project to design privacy protection algorithm based on heuristic privacy metrics. Project utilized "street information" as constraint to encourage generative model to output more realistic confounding trajectory. Through obfuscation, the algorithm reconstructs the trajectory based on the original trajectory of the target object, and then generates a dummy trajectory that effectively confounds attackers. The proposed method was implemented in an Android App prototype built with Vue.js, HTML, CSS, and Spring Boot.
- Granted software copyright as first author: ***Han Yuwei,*** *Yang Haolin, Chen Jingjing: Advanced location calling service software based on dynamic spatial query, Certificate number: 8216161, registration number:2021SR1493535, October 12th, 2021.*
- Granted software copyright as the first author: ***Han Yuwei,*** *Yu Qiaobin, Liao Ziwei: Location/track privacy protection mobile application system , Certificate number: 8216143, registration number:2021SR1493517 , October 12th, 2021.*
- Researched defense methods of adversarial sample attacks against machine learning models. Published paper: *Hailiang Li, Bin Zhang, Yu Zhang, Xilin Dang,* ***Yuwei Han,*** *Linfeng Wei, Yijun Mao, Jian Weng. A defense method based on attention mechanism against traffic sign adversarial samples, Information fusion, (76),55-65,2021.* In the paper, we proposed a spatial transformation model that can extract decision-relevant parts of the image and filter out irrelevant (attacking) noise to eventually synthesize a cleaner image. The experiment results demonstrated that our method has strong defensive ability and universality in both black box and white box attacks (e.g. on MNIST dataset, our method is much better than other defense methods, reaching 98.5% accuracy of adversarial sample recognition).
- Finished the Graduate Thesis *"Adversarial Examples for Deep Learning General Object Detectors"*. In the paper, we implemented two adversarial example attack methods namely (1) APA: an untargeted attack by adding "cross- shaped" patches to the key parts of the images and (2) VA: a targeted attack method that adds invisible tiny perturbations to the images globally. The experimental results on MS COCO2017 and PASCAL VOC2012 datasets showed that both APA and VA could decrease the performance of Faster-RCNN and YOLOv5 validly.(e.g. on MS COCO2017 dataset, the mean average precision of YOLOv5 declines from 50% to 8.4% under the APA attack with 1 pixel width patches and from 50% to 0.4% under VA attack.)
- Conducted research of deep learning methods to defend APT attacks; Studied the model inversion attack based on GANs.

Cambridge AI research group led by Prof. Pietro Lio' — Cambridge, UK
Machine Learning Security Research — 2020.11 - 2021.4
- Reviewed extensive literature on convolutional neural network robustness problems and adversarial sample defense and

then implemented representative works from scratch using PyTorch.

- For final report of the research program, proposed adding LSTM module to CNN to create feature filtering mask as an "information bottleneck" that only maintains the key features of the images for classification. This method was the first one to leverage LSTM to process images and effectively reduce negative effects of noise. Our approach increases the adversarial sample classification accuracy of vanilla CNN from 0% to 97.5%.
- Detailed research report in paper: **Yuwei Han,** Qing Wang: An adversarial sample defense model based on computer attention mechanism, 2021 IEEE 3rd International Conference on Communications, Information System and Computer Engineering. (DOI:10.1109/CISCE52179.2021.9446015) (EI index).

Institute of Network Science and Cyberspace, Tsinghua University   Beijing, China
Research of automatic labeling of network traffic data   2021.3 - 2021.7

- Utilized Unity and Python to build a pre-processing and 3D visualization pipeline for network data (over 500,000 samples) that included data extraction, one hot transform, normalization and so on to observe data characteristics and distribution.
- As the first step to labeling datasets, we proposed a probabilistic model of a similarity graph defined in terms of its edge probabilities and then learned these probabilities from data as a reinforcement learning task.

Cryptography and Information Security Laboratory, Shanghai Jiaotong University   Shanghai, China
Establishment of HTTPS test platform of Juanru Li postdoctoral group   2020.7 - 2020.9

- Examined the security and compliance of educational websites by writing Python scripts that obtain website certificate dates, HSTS, md5, DNS CAA and other related information; Built a regular scanning tool to automatically filter out expired and illegal webpages.
- Developed a HTTPS test platform to display website security evaluation results.

PROJECT EXPERIENCE

Java App Development   Guangzhou, China
Health self-test app based on iris detection technology   2021.1 - 2021.4

- Developed app that detects the sclera by taking pictures through mobile phone to help users obtain more timely and accurate health management information, which satisfies their multi-level health management needs.
- Granted software copyright as third author: Tan Jingwen, Guo Peini, **Han Yuwei**, Cai Yuqing, Zeng Lingchao: *Health self-test APP V1.0 based on iris detection technology*, Certificate number: 7360226, registration number: 2021SR0637600, May 6th, 2021.

Software Development by Python   Guangzhou, China
Credit evaluation system for small and micro enterprises based on K-Means algorithm   2020.1 - 2020.4

- Leveraged unsupervised classification algorithm K-Means to train nearly 40,000 real corporate credit records; Used pyQt5 framework to create UI for the system.
- Optimized system to provide credit evaluation results based on data from small and micro enterprises within 0.1s.

Blockchain Development by Solidity   Guangzhou, China
User identity authentication system based on blockchain   2021.4-2021.6

- Utilized Solidity to realize smart contracts, symmetrically encrypt user identity information and save it on the smart contract; Signed corresponding block to facilitate verification by the verifier.
- Combined blockchain with web interface to achieve functionalities for user login and registration.
- Granted software copyright: **Han Yuwei**: Blockchain-based authentication system, Certificate number: 8197139, registration number: 2021SR1474513, October 9th, 2021.

AWARDS

- National second prize of the National University Student Information Security Competition in 2021
- First prize of Guangdong Province in the Computer Design Competition in 2021
- Second prize of JNU in the "Challenge Cup" academic technology competition in 2021
- National third prize in the China University Student Service Outsourcing Competition in 2020
- Third prize in the South China Division in the Computer Design Contest-Artificial Intelligence Challenge in 2020
- Third prize in the South China Division in the iCAN International College Student Innovation and Entrepreneurship Competition in 2020
- Third prize of Guangdong Province in the Microsoft Innovation Cup "Imagine Cup" academic technology competition in 2020

PERSONAL INTERESTS

- **Research interests:** Adversarial Machine Learning , trustworthy AI, privacy protection, federated learning
- **Hobbies:** piano, violin, swimming, volunteering