Limin Yang

540-998-9158 | limin.yang.cs@gmail.com | https://liminyang.web.illinois.edu | GitHub: whyisyoung | Google Scholar

Summary

PhD in Machine Learning Security with hands-on experience in building secure, scalable platforms. Led the development of threat intelligence platforms and applied deep learning techniques to real-world security challenges. Published in top-tier security venues, with research on ML backdoor attacks and concept drift detection. Skilled in applying machine learning to security applications and attacking machine learning models.

Experience

Bytedance, Senior Software Engineer, Security Operations Center Team Aug. 2023 – Current

- Led the delivery of a multi-source **Threat Intelligence Platform**, integrating 150 million indicators with a unified schema and aggregation algorithm
- **Saved over $1 million annually** by leading evaluation of external products and designing custom threat feeds
- Protected URLs accessed by **10 million customers** and delivered actionable intelligence to **10 internal teams**
- Reduced false positives by 50% from the alerts generated by the correlation of IoCs and HIDS logs (10+ TB/hour)
- Engineered a malware scanning service that improved detection rates and streamlined workflows for 3+ teams
- AI for SOC: contributed using LLM for case summary, alert verdict based on history, and URL scan service
- Skills: Golang, Python, Shell, Git, gRPC, REST API, CI/CD, K8s, Kitex, PySpark, Kafka, MongoDB, Hive, Redis

IBM Research, Visiting Scholar (Research Intern), Hybrid Cloud Team May 2022 – Aug. 2022

- Filled the gap of quantitative understanding of security incidents and associated alerts from an SOC
- Highlighted gaps between real-world data and benchmarks with LSTM autoencoder; contributed new block rules
- Skills: Python, PyTorch, Deep learning, Cloud Security, Alert analysis, Feature Engineering, Anomaly Detection

Bytedance, Security Engineering Intern, Threat XDR Team May 2021 – Aug. 2021

- Enhanced email detection systems, protected **50+ client companies** and detected 5000 additional spams weekly
- Doubled user-labeled ground-truth dataset through clustering similar emails based on user actions
- Skills: Python, Machine Learning, Rule engine, Elasticsearch

UIUC, Graduate Research Assistant Aug. 2019 – July 2023

- **Adversarial ML**: bypassed 4 recent backdoor defenses by designing a new selective backdoor attack against Android malware classifiers; published in IEEE S&P 2023 (most prestigious security venue)
- **Concept Drift Detect and Explainable AI**: leveraged contrastive learning and autoencoder to detect samples from previously unseen classes; proposed a novel distance-based explanation to explain a sample is outlier; identified 161/165 unseen families on a company's Windows PE malware database; published in USENIX Security 2021
- Skills: Python, Tensorflow, Keras, PyTorch, Machine learning security, deep learning, out-of-distribution detection

Education

University of Illinois (UIUC), Ph.D. in Computer Science, advisor: Gang Wang	Aug. 2019 – Jul. 2023
Virginia Tech, Ph.D. in Computer Science, advisor: Gang Wang	Aug. 2018 – Jul. 2019
East China Normal University, M.S. Study in Computer Science	Sep. 2015 – Jun. 2018
East China Normal University, B.Eng. in Computer Science	Sep. 2011 – Jun. 2015

Selected Publications (**Citations: 800+**, h-index: 10)

[**IEEE S&P'23**] **1st author**. "Jigsaw Puzzle: Selective Backdoor Attack to Subvert Malware Classifiers ".

[**IEEE S&P'23**] 3rd author. "Practitioners' Perception of ML-Based Security Tools and Explanations ".

[**USENIX Security'21**] **1st author**. "CADE: Detecting and Explaining Concept Drift Samples for Security Applications". **160+ citations**, **125 GitHub stars**.

[Deep Learning and Security'21] **1st author**. "BODMAS: An Open Dataset for Learning based Temporal Analysis of PE Malware". **100+ citations**. Dataset requested by **140+ institutions**.

Technical Skills

Languages:	Python, Golang, C++, C, Shell, SQL (Postgres, Hive)
ML/Deep Learning:	Keras, Tensorflow, PyTorch, Scikit-learn, LightGBM
Cloud & Infra:	Linux, AWS, Docker, CI/CD, Kubernetes
Security:	Cloud Security, Threat intelligence, Malware detection, Network IDS, Phishing
Developer Tools:	Git, Vim, tmux, gRPC, Kafka, MongoDB, Redis, Elasticsearch, Nmap, Metasploit