Limin Yang

liminy2@illinois.edu +1 (540)998-9158 https://liminyang.web.illinois.edu/ GitHub: whyisyoung

Education

University of Illinois Urbana-Champaign, Ph.D. in Computer Science, Advisor: Gang Wang	*Aug.2019 – July.2023*
Virginia Tech, Ph.D. in Computer Science, Advisor: Gang Wang	*Aug.2018 – Aug.2019*
East China Normal University, Masters Study in Computer Science	*Sep.2015 – Jun.2018*
East China Normal University, B.E. in Computer Science	*Sep.2011 – Jun.2015*

Research Interests

Machine learning security and Internet measurement.

Publications

1. **[IEEE S&P'23]** Limin Yang, Zhi Chen, Jacopo Cortellazzi, Feargus Pendlebury, Kevin Tu, Fabio Pierazzi, Lorenzo Cavallaro, Gang Wang. "Jigsaw Puzzle: Selective Backdoor Attack to Subvert Malware Classifiers". In Proceedings of *The 44th IEEE Symposium on Security and Privacy (IEEE S&P)*, San Francisco, CA, May 2023.

2. **[IEEE S&P'23]** Jaron Mink, Hadjer Benkraouda, Limin Yang, Arridhana Ciptadi, Ali Ahmadzadeh, Daniel Votipka, Gang Wang. "Everybody's Got ML, Tell Me What Else You Have: Practitioners' Perception of ML-Based Security Tools and Explanations". In Proceedings of *The 44th IEEE Symposium on Security and Privacy (IEEE S&P)*, San Francisco, CA, May 2023.

3. **[DLSP'23]** Zhi Chen, Zhenning Zhang, Zeliang Kan, Limin Yang, Jacopo Cortellazzi, Feargus Pendlebury, Fabio Pierazzi, Lorenzo Cavallaro, Gang Wang. "Is It Overkill? Analyzing Feature-Space Concept Drift in Malware Detectors." In Proceedings of *6th Deep Learning Security and Privacy Workshop*, in conjunction with IEEE Symposium on Security and Privacy (IEEE S&P), San Francisco, CA, May 2023.

4. **[USENIX Security'21]** Limin Yang, Wenbo Guo, Qingying Hao, Arridhana Ciptadi, Ali Ahmadzadeh, Xinyu Xing, Gang Wang. "CADE: Detecting and Explaining Concept Drift Samples for Security Applications". In Proceedings of *The 30th USENIX Security Symposium*, Vancouver, Canada, August 2021. Artifact Evaluated.

5. **[DLS'21]** Limin Yang, Arridhana Ciptadi, Ihar Laziuk, Ali Ahmadzadeh, Gang Wang . "BODMAS: An Open Dataset for Learning based Temporal Analysis of PE Malware", In Proceedings of *4th Deep Learning and Security Workshop*, in conjunction with IEEE Symposium on Security and Privacy (Oakland), May 2021.

6. **[USENIX Security'20]** Shuofei Zhu, Jianjun Shi, Limin Yang, Boqin Qin, Ziyi Zhang, Linhai Song, Gang Wang. "Measuring and Modeling the Label Dynamics of Online Anti-Malware Engines". In Proceedings of *The 29th USENIX Security Symposium*, Boston, MA, August 2020. Artifact Evaluated.

7. **[SafeThings'20]** Hang Hu, Limin Yang, Shihan Lin, Gang Wang. "A Case Study of the Security Vetting Process of Smart-home Assistant Applications", In Proceedings of *IEEE Workshop on the Internet of Safe Things*, in conjunction with IEEE Symposium on Security and Privacy (Oakland), San Francisco, CA, May 2020.

8. **[IMC'19]** Peng Peng, Limin Yang, Linhai Song, Gang Wang. "Opening the Blackbox of VirusTotal: Analyzing Online Phishing Scan Engines." In Proceedings of *The ACM SIGCOMM Internet Measurement Conference*, Amsterdam, Netherlands, October 2019.

9. **[USENIX Security'18]** Dongliang Mu, Alejandro Cuevas, Limin Yang, Hang Hu, Xinyu Xing, Bing Mao, Gang Wang. "Understanding the Reproducibility of Crowd-reported Security Vulnerabilities." In Proceedings of *The 27th USENIX Security Symposium*, Baltimore, MD, August 2018.

10. **[Globecom'17]** Limin Yang, Xiangxue Li, Yu Yu. "VulDigger: A Just-in-time and Cost-Aware Tool for Digging Vulnerability-Contributing Changes." In Proceedings of *IEEE Global Communications Conference (GLOBECOM)*, Singapore, December 2017.

11. **[PPNA'17]** Minhui Xue, Limin Yang, Keith W. Ross, and Haifeng Qian. "Characterizing user behaviors in location-based find-and-flirt services: Anonymity and demographics." In *Peer-to-Peer Networking and Applications (PPNA)*, 2017.

Selected Research Experience

Selective Backdoor Attack on Malware Classifiers, Research Assistant, UIUC *Feb.2021 – Jan.2022*

- Proposed a new selective backdoor that only protects a malware author's own but not any others' malware.
- Achieved high attack success rates on 10 random families against an Android malware classifier in both feature space and problem space (e.g., software code).
- Increased the stealthiness of backdoor attack and successfully evaded four defenses including one state-of-the-art detection method.

BODMAS Windows PE Malware Dataset, Research Assistant, UIUC *July.2020 – Jan.2021*

- Released a new Windows PE malware dataset (BODMAS) with a security company (Blue Hexagon).
- BODMAS contains 57,293 malware samples and 77,142 benign samples collected from 2019/08–2020/09, with timestamps and curated malware family information (581 families). Feature vectors and metadata are publicly available via `https://whyisyoung.github.io/BODMAS/`.
- BODMAS malware binaries have been requested by 108 institutions (about 150 research groups) including those from developing countries.

Concept Drift Detection and Explanation, Research Assistant, UIUC *Aug.2019 – Jun.2020*

- Implemented a novel system (CADE) with contrastive learning to detect concept drift in security applications.
- Built an explanation module to offer semantically meaningful reasoning of CADE's decision with new metrics.
- CADE is 2 times faster and achieves higher detection rate (F1 = 96%) than state-of-the-art method Transcend (F1 = 80% or lower) on Android malware and network intrusion datasets.
- CADE also worked well on Blue Hexagon's PE malware database and identified 161 out of 165 unseen families.

Reliability of VirusTotal, Research Assistant, UIUC *Sep.2019 – Nov.2019*

- Surveyed 115 papers on how researchers use VirusTotal.
- Measured the label dynamics of 14,000+ PE malware via daily snapshots over one year and analyzed the correlations and causalities between VirusTotal engines.
- Identified questionable methodologies and offered suggestions on the usge of VirusTotal.

VirusTotal Phishing URLs Scanning, Research Assistant, Virginia Tech *Jan.2019 – May 2019*

- Controlled 66 phishing websites to understand the quality and reliability of security scanners and VirusTotal.
- Submitted phishing sites to VirusTotal and 18 security scanners periodically and observe the incoming traffic.
- Provided insights on the poor detection performance of VirusTotal and scanners' own APIs and suggestions to utilize VirusTotal more properly on URL labelling.

Smart Home Assistants Cloud Spoofing, Research Assistant, Virginia Tech *Aug.2018 – May 2019*

- Understand the authentication mechanism in smart home assistant systems (Amazon Alexa and Google Home).
- Developed an Amazon Alexa skill and a Google Home action for finding authentication issues.
- Verified that replay attack and SQL injection attack are feasible with proof-of-concept experiments.

Internships

IBM Research, Visiting Scholar (Research Intern), New York, US *May.2022 – Aug.2022*

- Cleaned a noisy real-world network IDS dataset (25 million traffic/day) from National Supercomputing Center.
- Semi-automatically labeled the dataset, defined 65 features for network logs, and summarized 279 incidents.
- Built anomaly detection models (per host) with LSTM autoencoder and applied on 2 past real-world attacks.

TikTok (ByteDance), Security Engineering Intern, California, US *May.2021 – Aug.2021*

- Augmenting Lark/Feishu spam email detection with rule system and user actions.
- Added ∼25 factors and ∼20 rules to capture ∼ 50,000 spams/week with an extra gain of ∼ 5,000 spams/week.
- Added ∼300 allowlist domains with manual and partial automation, protected more than 1 million emails/week.
- Built a daily task to leverage user action (add/remove spam) to cluster and capture similar emails based on 37 hand-picked features. It helped to double the size of ground-truth pool by finding more false positives and false negatives that is missing from existing detection system.

The Pennsylvania State University, Research Intern, Pennsylvania, US *Sep.2017 – Feb.2018*

- An empirical study to unveil the reproducibility of vulnerabilities using crowdsourcing information.
- Manually reproduced 368 real-world memory corruption bugs based on 6,000+ crowd-sourced reports.
- Obtained quantitative evidence on the prevalence of missing information in vulnerability reports and low reproducibility. Validated that crowdsourcing could ease the effort of vulnerability reproduction.

Peking University, Exploit Intern, Beijing, China *Jul.2015 – Aug.2015*

- Focused on practical training like binary vulnerability discovery/exploit (Windows).
- Extracted fingerprints for industrial control systems like Siemens S7-1200 with Nmap.

Awards

- CCS Student Conference Grant *2021*
- ECNU Graduate Student Overseas Research Scholarship *2017*
- ECNU Top-notch Innovative Personnel Training Plan **(4/91)** *2013 – 2015*

Teaching

- CS-463 Computer Security II, UIUC, Teaching Assistant *Fall 2022*
- CS-4264 Principles of Computer Security, Virginia Tech, Teaching Assistant *Spring 2019*
- CS-3114 Data Structures and Algorithms, Virginia Tech, Teaching Assistant *Fall 2018*

Professional Services

- **[TDSC]** IEEE Transactions on Dependable and Secure Computing, Reviewer *2023*
- **[TSC]** ACM Transactions on Social Computing, Reviewer *2023*
- **[SecureComm]** EAI International Conf on Security and Privacy in Communication Networks, Reviewer *2023*
- **[SECURWARE]** Emerging Security Information, Systems and Technologies, Technical Program Committee *2023*
- **[ADVCOMP]** Advanced Engineering Computing and Applications in Sciences, tpc *2023*
- **[JISA]** Journal of Information Security and Applications, Reviewer *2023*
- **[Oakland]** IEEE Symposium on Security and Privacy, Student PC *2021*
- [Patterns] Patterns by Cell Press, Reviewer *2021*