Zhi Chen

Thomas M. Siebel Center, 201 North Goodwin Avenue, Urbana, IL 61801-2302
zhic4@illinois.edu ⋄ (+1) 510-345-7211 ⋄ https://zhichen98.github.io

EDUCATION

- **University of Illinois at Urbana-Champaign** *Aug.2020 - Present*
  Ph.D. in Computer Science
  Advisor: Gang Wang
- **University of California, Berkeley** *Aug.2019 - May.2020*
  M.S. in Electrical Engineering and Computer Sciences
  Advisor: Dawn Song
- **University of California, Berkeley** *Aug.2016 - May.2019*
  B.S. Honors in Electrical Engineering and Computer Sciences

RESEARCH INTERESTS

AI Security, Data-driven Security, LLM Security

PUBLICATIONS

(*The authors contribute equally to this paper (co-first authors))

- [**USENIX Security 2024**] Limin Yang*, **Zhi Chen***, Chenkai Wang, Zhenning Zhang, Sushruth Booma, Phuong Cao, Constantin Adam, Alex Withers, Zbigniew Kalbarczyk, Ravishankar K. Iyer, Gang Wang. **True Attacks, Attack Attempts, or Benign Triggers? An Empirical Measurement of Network Alerts in a Security Operations Center**. Proceedings of *The 33rd USENIX Security Symposium*, Philadelphia, PA, August 2024.
- [**IEEE SP 2023**] 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**. Proceedings of *The 44th IEEE Symposium on Security and Privacy*, San Francisco, CA, May 2023.
- [**DLSP 2023**] **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**. Proceedings of *The 6th Deep Learning Security and Privacy Workshop*, in conjunction with IEEE Symposium on Security and Privacy, San Francisco, CA, May 2023.
- [**TKDE 2023**] Jinyin Chen, Jian Zhang, **Zhi Chen**, Min Du, Qi Xuan. **Time-aware Gradient Attack on Dynamic Network Link Prediction**. Proceedings of *The IEEE Transactions on Knowledge and Data Engineering*, February 2023.
- [**TKDE 2023**] Jiajun Zhou*, **Zhi Chen***, Min Du, Lihong Chen, Shanqing Yu, Guanrong Chen, Qi Xuan. **RobustECD: Enhancement of Network Structure for Robust Community Detection**. Proceedings of *The IEEE Transactions on Knowledge and Data Engineering*, January 2023.
- [**CCS 2019**] Min Du, **Zhi Chen**, Chang Liu, Rajvardhan Oak, Dawn Song. **Lifelong Anomaly Detection Through Unlearning**. Proceedings of *The 26th ACM Conference on Computer and Communications Security*, London, UK, November 2019.

Pre-Prints

- [**CCS 2025, Under Review**] **Zhi Chen**, Pirouz Naghavi, Phuong Cao, Zhenning Zhang, Sushruth Booma, Constantin Adam, Zbigniew Kalbarczyk, Ravishankar K. Iyer, Gang Wang. **Concept Drift Forensics using a Real-world Network Dataset**.
- [**NDSS 2026, Under Review**] Han Bao, Qinying Wang, **Zhi Chen**, Qingming Li, Xuhong Zhang, Peng Gao, Zonghui Wang, Shouling Ji, Wenzhi Chen. **VMODA: An Effective Framework for Adaptive NSFW Image Moderation**.

RESEARCH EXPERIENCE

- **Backdoor Attacks against Agent Systems**, Research Assistant, UIUC *Nov.2024 - Present*

  - Propose a type of backdoor that is separately embedded in the different stage of LLM agent.
  - Design a more flexible backdoor embedding and triggering scheme.
  - Provide a stealthier method than previous backdoors to avoid being detected.

- **Investigate Adversarial Attack Angles in Learned Systems**, Research Assistant, UIUC *Jun.2024 - Present*
  - Understand tunable feature space for adversarial attackers in different learned subsystems.
  - Generate adversarial noises/workloads in congestion control systems to find useful specifications.

- **Forensics of Concept Drift in Network Traffic Datasets**, Research Assistant, UIUC *Jan.2023 - Present*
  - Built a real-world dataset from our Security Operations Center (SOC) collaborator and labeled for benign/malicious, how many days, malicious types.
  - Conducted feature engineering for continuous features (Duration, Bytes Transferred, and Time Gap) and categorical features (Protocol, Connection State, Service, History, IP Address Binning, Port Binning).
  - Analyzed drift with supervised learning models and unsupervised learning model, and obtained some interesting findings.

- **Empirical Measurement of Network Alerts in a SOC**, Research Assistant, UIUC *Jan.2023 - Jun.2024*
  - Conducted a quantitative analysis of network alert data from a real-world Security Operations Center (SOC) over a four-year period, involving 115 million alerts.
  - Correlated these network alerts with 227 confirmed successful attacks over 20 years, offering insights into the effectiveness of SOC alerts in detecting true threats.
  - Observed a significant proportion of the alerts related to "attack attempts" and "benign triggers".

- **Backdoor Attack on Malware Classifiers**, Research Assistant, UIUC *Aug.2021 - Dec.2022*
  - Proposed a new selective backdoor that only protects the author's own malware 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.

- **Concept Drift Detection and Explanation**, Research Assistant, UIUC *Jan.2021 - Feb.2023*
  - Implemented a novel detector with contrastive learning to detect concept drift in security applications.
  - Built an explanation module to offer semantically meaningful reasoning of detector's decision with new metrics.
  - Achieved 2 times faster and higher detection rate than the state-of-the-art method Transcend on Android malware and network intrusion datasets.
  - Worked well on Blue Hexagon's PE malware database and identified 161 out of 165 unseen families.

- **Lifelong Anomaly Detection Through Unlearning**, Research Assistant, UC Berkeley *Jan.2019 - Sep.2019*
  - Proposed the unlearning framework which can be applied to any deep learning-based zero-positive anomaly detection approach to turn it into a lifelong anomaly detection solution.
  - Proposed novel techniques to tackle the exploding loss and catastrophic forgetting challenges, and a theoretical framework to apply a generative model for anomaly detection.
  - Evaluated on three real anomaly detection datasets and demonstrated significant reductions in false positives and false negatives. For example, in the Hadoop File System log dataset, it achieved reductions of up to 77.3% in false positives and 76.6% in false negatives under different thresholds.

- **A Database Security Project**, Research Intern, Alibaba DAMO Academy *Dec.2018 - Jan.2019*
  - Developed a LSTM model for detection of abnormal conditions of database.
  - Assisted in parsing unstructured, free-text log entries into structured representation.

HONORS AND AWARDS

- **USENIX Security Student Grant** *Aug.2024*
- **B.S. Honors**, UC Berkeley *May.2019*
- **Dean's List**, College of Engineering, UC Berkeley *May.2018*
- **Dean's List**, College of Engineering, UC Berkeley *May.2017*
- **Finalist**, the 67th Intel International Science and Engineering Fair, Phoenix *May.2016*

TALKS

- **True Attacks, Attack Attempts, or Benign Triggers? An Empirical Measurement of Network Alerts in a Security Operations Center**, Proceedings of *The 33rd USENIX Security Symposium*, Philadelphia, PA, USA, August 2024.
- **Is It Overkill? Analyzing Feature-Space Concept Drift in Malware Detectors**, Proceedings of *The 6th Deep Learning Security and Privacy Workshop*, San Francisco, CA, USA, May 2023.
- **Feature-Space Concept Drift in Malware Detectors**, Zhejiang University Machine Learning Security Seminar, Hangzhou, China, June 2023.

TEACHING

- CS 463: Computer Security II, UIUC, Teaching Assistant *Fall 2023*
- CS 445: Computational Photography, UIUC, Teaching Assistant *Fall 2020*

PROFESSIONAL SERVICES

- **Conference Subreviewer**, ACM Conference on Computer and Communications Security (CCS) *2024*