

## Wenjie Qu

Email: wen\_jie\_qu@outlook.com

Website: <https://quwenjie.github.io/>

## EDUCATION

**Huazhong University of Science and Technology**, Wuhan, China

2019.9-2023.6(Expected)

B.E. in Automation, Honor Class

GPA: 3.88/4.0

## PUBLICATIONS

- [1] **EncoderMI: Membership Inference against Contrastive Learning**  
Hongbin Liu\*, Jinyuan Jia\*, **Wenjie Qu**, Neil Gong  
*ACM Conference on Computer and Communications Security (CCS)* 2021
- [2] **jTrans: Jump-Aware Transformer for Binary Code Similarity Detection**  
Hao Wang\*, **Wenjie Qu\***, Gilad Katz, Wenyu Zhu, Zeyu Gao, Han Qiu, Jianwei Zhuge, Chao Zhang  
*International Symposium on Software Testing and Analysis (ISSTA)* 2022
- [3] **MultiGuard: Provably Robust Multi-label Classification against Adversarial Examples**  
Jinyuan Jia\*, **Wenjie Qu\***, Neil Gong  
Submitted to NeurIPS 2022
- [4] **MPass: Bypassing Learning-based Static Malware Detectors**  
Jialai Wang, **Wenjie Qu**, Yi Rong, Chao Zhang, Han Qiu, Qi Li, Zongpeng Li  
Submitted to AAAI 2023
- [5] **A Certified Radius-Guided Attack Framework to Image Segmentation Models**  
**Wenjie Qu\***, Youqi Li\*, Binghui Wang  
Submitted to NDSS 2023
- [6] **REaaS: Enabling Adversarially Robust Downstream Classifiers via Robust Encoder as a Service**  
**Wenjie Qu**, Jinyuan Jia, Neil Gong  
Submitted to NDSS 2023

## RESEARCH EXPERIENCE

### *CoLink: A Programming Framework for Decentralized Data Science*

Research Intern at University of California, Berkeley

April 2022-Present

Advisor: **Prof. Dawn Song**

- Participated in the design of CoLink, a programming framework which greatly simplifies the deployment of decentralized data science solutions.
- Designed and implemented CoLink SDK python interface, based on gRPC services.
- Designed and implemented the CoLink-crypten framework which contains protocols and backends to enable users to perform general privacy-preserving machine learning tasks supporting various data collaboration scenarios without writing code, based on several MPC libraries and python SDK.

### *jTrans: Jump-Aware Transformer for Binary Code Similarity Detection*<sup>[2]</sup>

Research Intern at Tsinghua University

July 2021-January 2022

Advisor: **Prof. Chao Zhang**

- Proposed a novel neural network architecture for binary function similarity detection, encoding control flow information into the transformer.
- Proved through attention weights how our mechanism delivered the jump target information.
- Released the currently largest binary dataset to the community as a benchmark.
- Outperformed state-of-the-art binary similarity detection methods by 30.5%.

***REaaS: Enabling Adversarially Robust Downstream Classifiers via Robust Encoder as a Service***<sup>[6]</sup>

Research Intern at Duke University

June 2021-November 2022

Advisor: **Prof. Neil Gong**

- Proposed a novel method for encoder cloud service which enables a client to build a provably robust downstream classifier and derive certified radius while reducing the number of queries.
- Proposed a novel pre-training method to enhance the robustness of the encoder based on a spectral-norm regularization term.
- Achieved much better certified robustness for the clients' downstream classifiers when the cloud server pre-trains the encoder via our spectral-norm regularized training method.

***MultiGuard: Provably Robust Multi-label Classification against Adversarial Examples***<sup>[3]</sup>

Research Intern at Duke University

February 2021-May 2021

Advisor: **Prof. Neil Gong**

- Proposed the first provable defense against adversarial examples on the task of multi-label classification.
- Showed a provable lower bound of intersection size between the set of labels predicted by our MultiGuard and ground truth labels, by a variant of Neyman-Pearson Lemma.
- Outperformed previous work by 7% on top-k precision, 15% on top-k recall.

***A Certified Radius-Guided Attack Framework to Image Segmentation Models***<sup>[5]</sup>

Research Intern at Illinois Institute of Technology

August 2020-January 2021

Advisor: **Prof. Binghui Wang**

- Designed an attack framework for image segmentation models leveraging the properties of certified radius.
- Proposed the first blackbox attack to image segmentation models via gradient estimation based on bandits.
- Outperformed state-of-the-art PGD attack by 13% relatively.

**ACADEMIC SERVICE**

External Reviewer

- International Conference on Machine Learning (ICML), 2022

**HONORS & AWARDS**

- |   |      |
|---|------|
| • Autodriving CTF, DEFCON 29, 4th/89  | 2021 |
| • China National Scholarship(Highest honor awarded by Ministry of Education, 6/350) | 2020 |
| • Outstanding Undergraduate(Highest honor awarded to HUST undergraduates, top 1%)   | 2020 |
| • Merit Student (1/30)  | 2020 |
| • Bronze Medal, National Olympiad in Informatics Winter Camp                        | 2018 |
| • First Prize, National Olympiad in Informatics in Provinces                        | 2017 |