# Jiahao Yu

+86 18621806870 • yujiahao@sjtu.edu.cn
www.jiahaoyu1999.com

## Education

**Shanghai Jiao Tong University** **Shanghai, China**
*B.S., School of Electronic Information and Electrical Engineering Sept. 2017 – Jun. 2021 (Expected)*

- Zhiyuan Honors Program
- Information Engineering
- GPA: overall 85.3/100; junior 87.7/100

## Research Interests

I'm generally interested in security and privacy issues of deep learning and federated learning. Specifically, my work focuses on adversarial attack, backdoor attack, attribute inference and differential privacy.

## Publications

**Voiceprint Mimicry Attack TowardsSpeaker Verification System in Smart Home**

- Lei Zhang, Yan Meng, **Jiahao Yu**, ChongXiang, Brandon Folk, Haojin Zhu
- In Proceedings of 2020 IEEE InternationalConference on Computer Communications (INFOCOM 2020)

**Invisible Backdoor Attacks Against Deep Neural Networks**

- Shaofeng Li, Benjamin Zi Hao Zhao, **Jiahao Yu**, Minhui Xue, Dali Kaafar, Haojin Zhu
- arXiv preprint arXiv:1909.02742v1, 2019

**MSTrojan: Backdoor Attacks in Federated Learning with First-Order Triggers**

- **Jiahao Yu**, Jungang Yang, Liyao Xiang, Weiting Li, Quanshi Zhang
- Under review by IEEE Transactions on Mobile Computing (TMC)

**Deep Model Privacy Leakage through Malicious Adversarial Training**

- **Jiahao Yu**, Liyao Xiang, Shunchen Cai, Hongxu Li
- Under review by 2021 AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI 2021)

**Matrix Gaussian Mechanism for Differentially-Private Learning**

- Jungang Yang, **Jiahao Yu**, Ruidong Chen, Weiting Li, Liyao Xiang, Xinbing Wang, Baochun Li
- Under review by Proceedings of 2020 IEEE International Conference on Computer Communications (INFOCOM 2021)

# Research Projects

**Multi-Step Trojan Attack in Federated Learning** **Sept. 2019 – Mar. 2020**
*Advisor: Prof. Liyao Xiang* *Shanghai Jiao Tong University*

- Designed a new backdoor attack in federated learning by training the local model and trojan triggers simultaneously
- Gived the proof of convergence of the attack algorithm and the bound of attack performance
- Compared the attack with previous works on 4 image datasets: CIFAR-10/100, GTSRB and Caltech256; the attack success rates were much higher on all four datasets (5%-20% improvement)

**Differentially-Private mechanism in Deep Learning** **May. 2020 – Jul. 2020**
*Advisor: Prof. Liyao Xiang* *Shanghai Jiao Tong University*

- Designed a differentially-private mechanism called *Matrix Gaussian Mechanism (MGM)* utilizing the matrix Gaussian distribution to guarantee $(\epsilon, \delta)$-differential privacy
- Rigorously proved that *MGM* meets differential privacy and has a tighter noise bound, in light of which higher utility than previous works can be achieved.
- Implemented *MGM* in horizontal and vertical federated learning and compared with other 3 mechanisms, which showed better performance with same noise

**Adversarial Attack Towards Speaker Verification System** **May. 2018 – Aug. 2018**
*Advisor: Prof. Haojin Zhu* *Shanghai Jiao Tong University*

- Proposed *VMask*, the first practical attack towards automatic speaker verification systems; *VMask* used grey-box and black-box to generate adversarial audio
- Implemented *VMask* on popular automatic speaker verification systems, and the attack success rates of grey-box and black-box scenarios are nearly 100% and 70%
- Implemented *VMask* on real-world devices; Apple HomeKit based on Siri speaker verification system can be attacked by *VMask*

**Model Inversion Attack in Robust Models** **Feb. 2019 – Aug. 2019**
*Advisor: Prof. Liyao Xiang* *Shanghai Jiao Tong University*

- Provided a new metric to evaluate privacy leakage of model inversion attack, which is consistent with human evaluation results
- Analyzed the privacy leakage of robust models with the new metric
- Proposed a new attack in federated learning to steal the class representation information from the global model

**Invisible Backdoor Attacks Against Deep Neural Networks** **Jun. 2018 – Sept. 2018**
*Advisor: Prof. Haojin Zhu* *Shanghai Jiao Tong University*

- Provided an optimization framework for the creation of invisible backdoor attacks, which addresses the challenges that backdoor patterns are obvious towards humans
- Proposed $L_0$ and $L_2$ optimizations to generate backdoor patterns
- Evaluated the attack on 3 image datasets: CIFAR-10/100, GTSRB, and got much higher invisibility score by SSIM (nearly 1)

**Adversarial Examples Towards NLP Models** **Jul. 2020 – Present**
*Advisor: Prof. Bo Li* *University of Illinois at Urbana-Champaign*

- Generated adversarial example towards sentiment classification models via style-transfer Variational AutoEncoder while maintaining attributes from original instances
- Evaluated proposed attack towards RNN and BERT models on 2 text datasets: Yelp and datasets collected from Microsoft Research

**Adversarial Robustness for Malware Detectors** **Sept. 2020 – Present**
*Advisor: Dr. Bin Zhu and Dr. Shay Kels* *Microsoft Research Asia*

- Exploited adversarial training to enhance the robustness of malware detectors based on deep learning
- Generated adversarial examples via obfusaction; the sequence of obfusaction actions were generated from Sequence GAN

## Honors and Awards

Zhiyuan Honors Awards
Zhiyuan Honors Scholar, Shanghai Jiao Tong University

- Research Project: Adversarial Deep Learning and its Applications in Internet of Things
- Outstanding Achievement Award (best project of the year)

## Activities Experiences

**Data and System Security Workshop** **Aug. 2019, Zhejiang University**

- Attended this workshop and listened to the report about security issues both in academy and industry

**Beijing Academy of Artificial Intelligence Conference** **Jun. 2020, BAAI**

- Attended this online conference and listened to report of artificial intelligence and its applications

**2020 Workshop on Federated Learning and Analytics** **Jul. 2020, Google**

- Attended this online workshop and listened to the report about open problems in federated learning

**Federated Learning Workshop using TensorFlow Federated** **Jul. 2020, Google**

- Attended this online workshop and learned to build own federated learning algorithm with Tensorflow

**Baidu Artificial Intelligence Student Club** **Oct. 2017 - present, Baidu**

- Taught new members to build artificial intelligence applications
- Cooperated with club members to attend Dian Shi global classification competition (top 10%)

## Programming Skills

Python > C++/C# >= Matlab