Jing Xu

Curriculum Vitae

CISPA Helmholtz Center for Information Security
Stuhlsatzenhaus 5, 66123 Saarbrücken
Germany
✱ *Born: 4 April, 1994*
+49 1747735811
✉ *jingxu.buaa@gmail.com*
xujing1994.github.io

About Me

I am a researcher with a strong academic background in computer science, focusing on machine learning, artificial intelligence, and security. Proven ability to develop and deploy scalable, secure, and robust machine learning models, including graph neural networks and large language models. Passionate about developing AI-driven solutions and leveraging cutting-edge AI technologies to address real-world challenges.

Education

2019-2024 **PhD, Computer Science**, ***Delft University of Technology***, **The Netherlands**

- Research Focus: Machine Learning, Graph Neural Networks, Security.
- Thesis: *Exploring backdoor attacks on graph neural networks.*
- Supervisors: Prof. Inald Lagendijk, Prof. Frans A. Oliehoek and Dr. Stjepan Picek.
- Achievements: Published 10+ papers at top-tier conferences & journals. Successfully defended in May 2024.

2016-2019 **MSc, Optical Engineering**, ***Beihang University***, **China**

- Specialization: Electrical Engineering, Signal Processing, Computer Vision
- Thesis: Research on Multi-Source Information Fusion in the All-Source Navigation and Positioning System based on the Factor Graph
- GPA: **3.857/4.0**-RANK: **top 5%**

2012-2016 **BSc, Electrical Engineering**, ***Shanghai University***, **China**

- Specialization: Computer Vision, Signal Processing, Automata
- Thesis: Coin Automatic Recognition System based on Computer Vision
- GPA:**3.86/4.0**-RANK: **1/931**

Work Experience

2023.11-present **Researcher**, ***CISPA, SprintML Lab***, **Germany**

- Developed privacy-preserving machine learning mechanisms to protect sensitive data in Large Language Models (LLMs) during training, fine-tuning, or soft prompt transferring.
- Implemented ML pipelines using tools such as Git, Docker, and GitLab, ensuring efficient deployment and scaling of production models.
- Collaborated with cross-functional teams to design and deploy machine learning solutions that address security risks.
- Mentored junior researchers in developing machine learning models and contributed to multiple research publications.

2019 **Researcher Intern**, ***Momo Technology Company, Deep Learning Lab***, **China**

- Developed data pipelines and automated workflows for processing and curating large datasets used in model training and evaluation.
- Developed GAN-based methods for face recognition and object detection.
- Developed deep learning models for object detection against spoofing, ensuring model robustness and applicability in high-performance environments.

Publications

2024 **“POST: A Framework for Privacy of Soft-prompt Transfer."** Xun Wang*, **Jing Xu**, Franziska Boenisch, Michael Backes, Adam Dziedzic. *ICML 2024 Next Generation of AI Safety Workshop.*

2024 **“Tabdoor: Backdoor Vulnerabilities in Transformer-based Neural Networks for Tabular Data."** Bart Pleiter*, Behrad Tajalli, Stefanos Koffas, Gorka Abad, **Jing Xu**, Martha Larson, Stjepan Picek. *arXiv.*

2023 **“Poster: Multi-target & Multi-trigger Backdoor Attacks on Graph Neural Networks." Jing Xu***, Stjepan Picek. *The ACM Conference on Computer and Communications Security (CCS).*

2023 **“On Exploring Backdoor Attacks in Federated Graph Neural Networks." Jing Xu***, Stefanos Koffas, Stjepan Picek. *The Learning from Authoritative Security Experiment Results (LASER) workshop.*

2023 **“Watermarking Graph Neural Networks based on Backdoor Attacks." Jing Xu***, Stefanos Koffas, Oguzhan Ersoy, Stjepan Picek. *IEEE European Symposium on Security and Privacy (Euro S&P).*

2023 **“Rethinking the Trigger-injecting Position in Graph Backdoor Attack." Jing Xu***, Gorka Abad, Stjepan Picek. *IJCNN.*

2023 **“BlindSage: Label Inference Attacks against Node-level Vertical Federated Graph Neural Networks."** Marco Arazzi*, Mauro Conti, Stefanos Koffas, Marina Krcek, Antonino Nocera, Stjepan Picek, **Jing Xu**. *arXiv.*

2023 **“SoK: A Systematic Evaluation of Backdoor Trigger Characteristics in Image Classification."** Gorka Abad, **Jing Xu**, Stefanos Koffas, Behrad Tajalli, Stjepan Picek, Mauro Conti. *arXiv.*

2023 **“Unveiling the Threat: Investigating Distributed and Centralized Backdoor Attacks in Federated Graph Neural Networks."** Gorka Abad, **Jing Xu***, Stefanos Koffas, Stjepan Picek. *Digital Threats: Research and Practice (DTRAP).*

2023 **“A Systematic Evaluation of Backdoor Attacks in Various Domains."** Stefanos Koffas, Behrad Tajalli, **Jing Xu**, Mauro Conti and Stjepan Picek. *Embedded Machine Learning for Cyber-Physical, IoT, and Edge Computing: Use Cases and Emerging Challenges, 2023, pages 519 - 552.*

2022 **“More is Better (Mostly): On the Backdoor Attacks in Federated Graph Neural Networks." Jing Xu***, Rui Wang, Kaitai Liang, Stjepan Picek. *Annual Computer Security Applications Conference (ACSAC).*

2022 **“Poster: Clean-label Backdoor Attack on Graph Neural Networks." Jing Xu***, Stjepan Picek. *The ACM Conference on Computer and Communications Security (CCS).*

2022 **“Label-Only Membership Inference Attack against Node-Level Graph Neural Networks."** Mauro Conti, Jiaxin Li*, Stjepan Picek, **Jing Xu**. *AISec, CCS Workshop.*

2022 **“Can You Hear It? Backdoor Attacks via Ultrasonic Triggers."** Stefanos Koffas*, **Jing Xu**, Mauro Conti, Stjepan Picek. *The ACM Workshop on Wireless Security and Machine Learning (WiseML).*

2021 **“Explainability-based backdoor attacks against graph neural networks." Jing Xu***, Minhui(Jason) Xue, Stjepan Picek. *The ACM Workshop on Wireless Security and Machine Learning (WiseML).*

Skills

Developer Tools: Linux, Slurm, Docker, VS Code, Git, GitLab, tmux, SSH, Jupyter

Libraries: Python, C++, PyTorch, TensorFlow, Hugging Face Transformers, Scikit-learn

Data Processing & Analysis: Pandas, NumPy, Matplotlib, Data Pipelines

Model Evaluation & Optimization: Hyperparameter Tuning, Accuracy Metrics

Algorithms: LLMs, vision-language models, GNNs, GANs, Fine-tuning models

Language: Mandarin–Native, English–Fluent, German–Beginner

Selected Projects

2024 **Differentially Private Graph Prompt Learning**

- First study to demonstrate private information can leak from graph prompts.
- Developed privacy-preserving machine learning models for secure data handling in production environments.

2024 **Private Soft-prompt Transfer**

- Explored secure soft-prompt transfer techniques for privacy-preserving LLMs.
- Proposed a novel method to transfer private prompts between LLMs using only public data.

2023 **Protect Ownership of Graph Neural Networks**

- Developed a watermarking framework to verify ownership of graph neural networks, ensuring model integrity and security.
- Conducted hypothesis testing to provide statistical analysis for verifying model ownership in practice.

2020-2023 **Exploring Security of Graph Neural Networks**

- Designed explainability-based backdoor attacks against GNNs, where the performance of our attack can be better explained and visualized.
- Applied federated learning to train GNNs over isolated private graph data.
- Designed multiple novel backdoor attacks to enhance the development of more secure and robust GNN models.

Honors

2018	BUAA Outstanding Graduate Student, Outstanding Member
2017	BUAA First Prize Scholarship (two consecutive years)
2016	SHU Outstanding Graduate Student, Outstanding Student, Outstanding Member
2015	SHU First Prize Scholarship (three consecutive years), GuangHua Scholarship

Teaching and Supervision

Course

2022 Security and Privacy of Machine Learning, Radboud University, The Netherlands

Supervision

Master Student Alex Simonov, 2022–, Delft University of Technology.

List of Referees

Referee 1 **Dr. Stjepan Picek**, *Radboud University*, The Netherlands, stjepan.picek@ru.nl

Referee 2 **Prof. Inald Lagendijk**, *Delft University of Technology*, The Netherlands, R.L.Lagendijk@tudelft.nl

Referee 3 **Prof. George Smaragdakis**, *Delft University of Technology*, The Netherlands, G.Smaragdakis@tudelft.nl