Xinlei He — CV

xinlei.he@cispa.de • xinleihe.github.io • Xinlei He

Education

CISPA Helmholtz Center for Information Security — **Saarbrücken, Germany**
Ph.D. in Computer Science — *February 2020 –*
Advisor: Dr. Yang Zhang

Fudan University — **Shanghai, China**
Master in Computer Science — *September 2017 – January 2020*
Advisor: Prof. Yang Chen

Fudan University — **Shanghai, China**
Bachelor in Computer Science — *September 2013 – June 2017*
Advisor: Prof. Yang Chen

Research Interests

Trustworthy Machine Learning

Service

- PC member
  - 2024: IEEE S&P
  - 2022: ESORICS
  - 2021: ESORICS (Poster Session)
  - 2020: SocInfo
- External reviewer
  - 2022: USENIX Security, CCS, NDSS, IEEE S&P, Euro S&P, NeurIPS
  - 2021: USENIX Security, CCS, NDSS, Euro S&P, ICLR, WWW, ICWSM, CHI, AAAI, AISACCS, PETS
  - 2020: Usenix Security, CCS, CSCW, ESORICS, PETS

Awards

- The Norton Labs Graduate Fellowship 2022

Publication

Conference

[1] Yiting Qu, Xinyue Shen, **Xinlei He**, Michael Backes, Savvas Zannettou, and Yang Zhang. Unsafe Diffusion: On the Generation of Unsafe Images and Hateful Memes From Text-To-Image Models. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, 2023.

[2] Ziqing Yang, **Xinlei He**, Zheng Li, Michael Backes, Mathias Humbert, Pascal Berrang, and Yang Zhang. Data Poisoning Attacks Against Multimodal Encoders. In *International Conference on Machine Learning (ICML)*. PMLR, 2023.

[3] Yihan Ma, Zhikun Zhang, Ning Yu, **Xinlei He**, Michael Backes, Yun Shen, and Yang Zhang. Generated Graph Detection. In *International Conference on Machine Learning (ICML)*. PMLR, 2023.

[4] Zeyang Sha, **Xinlei He**, Ning Yu, Michael Backes, and Yang Zhang. Can't Steal? Cont-Steal! Contrastive Stealing Attacks Against Image Encoders. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2023.

[5] Boyang Zhang, **Xinlei He**, Yun Shen, Tianhao Wang, and Yang Zhang. A Plot is Worth a Thousand Words: Model Information Stealing Attacks via Scientific Plots. In *USENIX Security Symposium (USENIX Security)*. USENIX, 2023.

[6] Yiting Qu, **Xinlei He**, Shannon Pierson, Michael Backes, Yang Zhang, and Savvas Zannettou. On the Evolution of (Hateful) Memes by Means of Multimodal Contrastive Learning. In *IEEE Symposium on Security and Privacy (S&P)*. IEEE, 2023.

[7] **Xinlei He**, Hongbin Liu, Neil Zhenqiang Gong, and Yang Zhang. Semi-Leak: Membership Inference Attacks Against Semi-supervised Learning. In *European Conference on Computer Vision (ECCV)*. Springer, 2022.

[8] Tianshuo Cong, **Xinlei He**, and Yang Zhang. SSLGuard: A Watermarking Scheme for Self-supervised Learning Pre-trained Encoders. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, 2022.

[9] Zheng Li, Yiyong Liu, **Xinlei He**, Ning Yu, Michael Backes, and Yang Zhang. Auditing Membership Leakages of Multi-Exit Networks. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, 2022.

[10] Xinyue Shen, **Xinlei He**, Michael Backes, Jeremy Blackburn, Savvas Zannettou, and Yang Zhang. On Xing Tian and the Perseverance of Anti-China Sentiment Online. In *International Conference on Weblogs and Social Media (ICWSM)*, pages 944–955. AAAI, 2022.

[11] Yun Shen*, **Xinlei He***, Yufei Han, and Yang Zhang. Model Stealing Attacks Against Inductive Graph Neural Networks. In *IEEE Symposium on Security and Privacy (S&P)*. IEEE, 2022 (* Equal contribution).

[12] Yugeng Liu, Rui Wen, **Xinlei He**, Ahmed Salem, Zhikun Zhang, Michael Backes, Emiliano De Cristofaro, Mario Fritz, and Yang Zhang. ML-Doctor: Holistic Risk Assessment of Inference Attacks Against Machine Learning Models. In *USENIX Security Symposium (USENIX Security)*. USENIX, 2022.

[13] **Xinlei He**, Jinyuan Jia, Michael Backes, Neil Zhenqiang Gong, and Yang Zhang. Stealing Links from Graph Neural Networks. In *USENIX Security Symposium (USENIX Security)*. USENIX, 2021.

[14] **Xinlei He** and Yang Zhang. Quantifying and Mitigating Privacy Risks of Contrastive Learning. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, 2021.

Journal

[15] Qinge Xie, Qingyuan Gong, **Xinlei He**, Yang Chen, Xin Wang, Haitao Zheng, and Ben Y. Zhao. Trimming mobile applications for bandwidth-challenged networks in developing regions. *IEEE Transactions on Mobile Computing (TMC)*, 2021.

[16] **Xinlei He**, Qingyuan Gong, Yang Chen, Yang Zhang, Xin Wang, and Xiaoming Fu. Datingsec: Detecting malicious accounts in dating apps using a content-based attention network. *IEEE Transactions on Dependable and Secure Computing (TDSC)*, 2020.

[17] Qingyuan Gong, Yang Chen, **Xinlei He**, Yu Xiao, Pan Hui, Xin Wang, and Xiaoming Fu. Cross-site prediction on social influence for cold-start users in online social networks. *ACM Transactions on the Web (TWEB)*, 2020.

[18] Qingyuan Gong, Yang Chen, **Xinlei He**, Zhou Zhuang, Tianyi Wang, Hong Huang, Xin Wang, and Xiaoming Fu. Deepscan: Exploiting deep learning for malicious account detection in location-based social networks. *IEEE Communications Magazine*, 2018.

Teaching

Lectures

Teaching Assistant — **Advanced Lecture: Privacy Enhancing Technologies**
May 2020 - September 2020, Saarland University

Teaching Assistant — **Seminar: Data-driven Approaches on Understanding Disinformation**
May 2020 - September 2020, Saarland University