Nils Lukas

nils.h.lukas@gmail.com • nilslukas.github.io

Updated on March 26, 2024

Research Interests

Design safe and reliable Machine Learning systems in the presence of untrustworthy

- **Providers**: Confidential computing via Homomorphic Encryption & Secret Sharing.
- **Data**: Mitigate data poisoning during training & prompt injection during inference.
- **Models**: Protect training data privacy through PII scrubbing & differential privacy.
- **Users**: Control misuse by detecting generated (mis)information with watermarking.

Education

University of Waterloo, Canada — 2019 - 02/2024
Ph.D. in Computer Science
- Advisor: Florian Kerschbaum
- Thesis: Analyzing Threats of Large-Scale Machine Learning Systems

RWTH-Aachen, Germany
M.Sc. in Computer Science *(w/Distinction)* — 2016 - 2018
B.Sc. in Computer Science — 10/2012 - 2016

Honors & Awards

Best Poster Award, David R. Cheriton [300 CAD]	2023
Distinguished Contribution Award, Microsoft MLADS conference	2023
David R. Cheriton Scholarship, University of Waterloo [20,000 CAD]	2022 - 2024
Outstanding Reviewer (Top 10%), ICML'22	2022
Best Poster Award, Rogers [1,000 CAD]	2020
Graduation with Excellence, RWTH-Aachen	2018
KU Global Scholarship, Korea University [1.2 million KRW]	2016
MOGAM Scholarship, RWTH-Aachen [3,000 EUR]	2014

Conference Publications

[USENIX'24] Fast and Private Inference of Deep Neural Networks by Co-designing Activation Functions
Abdulrahman Diaa, Lucas Fenaux, Thomas Humphries, Marian Dietz, Faezeh Ebrahimianghazani, Bailey Kacsmar, Xinda Li, **Nils Lukas**, Rasoul Akhavan Mahdavi, Simon Oya, Ehsan Amjadian, Florian Kerschbaum. In the 33rd USENIX Security Symposium, 2024.

[ICLR'24] AR: 30.8% (2 250/7 262)
Leveraging Optimization for Adaptive Attacks on Image Watermarks
Nils Lukas, Abdulrahman Diaa, Lucas Fenaux, Florian Kerschbaum. In the Twelfth International Conference on Learning Representations, 2024.

[ICLR'24] AR: 30.8% (2 250/7 262)
Media Coverage
Universal Backdoor Attacks
Benjamin Schneider, **Nils Lukas**, Florian Kerschbaum. In the Twelfth International Conference on Learning Representations, 2024.

[USENIX'23] AR: 29.2% (422/1 444)
PTW: Pivotal Tuning Watermarking for Pre-Trained Image Generators
Nils Lukas and Florian Kerschbaum. In the 32nd USENIX Security Symposium, 2023.

[S&P'23] AR: 17.0% (195/1 147)
Distinguished Contribution Award at Microsoft MLADS
Analyzing Leakage of Personally Identifiable Information in Language Models
Nils Lukas, Ahmed Salem, Robert Sim, Shruti Tople, Lukas Wutschitz, Santiago Zanella-Béguelin. In the 44th IEEE Symposium on Security and Privacy, 2023.

[S&P'22] AR: 14.5% (147/1 012)
SoK: How Robust is Image Classification Deep Neural Network Watermarking?
Nils Lukas, Edward Jiang, Xinda Li, Florian Kerschbaum. In the 43rd IEEE Symposium on Security and Privacy, 2022.

[ICLR'21] AR: 28.7% (860/2 997)
Spotlight (Top 5%)
Deep Neural Network Fingerprinting by Conferrable Adversarial Examples
Nils Lukas, Yuxuan Zhang, Florian Kerschbaum. The Ninth International Conference on Learning Representations, 2021.

[IH&MMSEC'21] AR: 40.3% (128/318)
On the Robustness of Backdoor-based Watermarking in Deep Neural Networks
Masoumeh Shafieinejad, **Nils Lukas**, Jiaqi Wang, Xinda Li, Florian Kerschbaum. Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security, 2021.

[ACSAC'20] AR: 20.9% (104/497)
Practical Over-Threshold Multi-Party Private Set Intersection
Rasoul Mahdavi, Thomas Humphries, Bailey Kacsmar, Simeon Krastnikov, **Nils Lukas**, John Premkumar, Masoumeh Shafieinejad, Simon Oya, Florian Kerschbaum, Erik-Oliver Blass. Annual Computer Security Applications Conference (ACSAC), 2020.

[EuroS&P'20] AR: 20.9% (39/187)
Differentially Private Two-Party Set Operations
Bailey Kacsmar, Basit Khurram, **Nils Lukas**, Alexander Norton, Masoumeh Shafieinejad, Zhiwei Shang, Yaser Baseri, Maryam Sepehri, Simon Oya, Florian Kerschbaum. IEEE European Symposium on Security and Privacy (EuroS&P), 2020.

Journal Publications

[AIP'18]
SunFlower: A new Solar Tower Simulation Method for use in Field Layout Optimization,
Pascal Richter, Gregor Heiming, **Nils Lukas**, Martin Frank. AIP Conference Proceedings, Volume 2033, Issue 1, 2018.

Working Papers

Pick your Poison: Undetectability versus Robustness in Data Poisoning Attacks against Deep Image Classifiers
Nils Lukas and Florian Kerschbaum.

PEPSI: Practically Efficient Private Set Intersection in the Unbalanced Setting
Rasoul Mahdavi, **Nils Lukas**, Faezeh Ebrahimianghazani, Thomas Humphries, Bailey Kacsmar, John Premkumar, Xinda Li, Simon Oya, Ehsan Amjadian, Florian Kerschbaum.

Work Experience

Research Intern, Royal Bank of Canada (BorealisAI), Toronto — 2024
- Host: Kevin Wilson

Research Intern, Microsoft Research, Cambridge — 2022
- Hosts: Shruti Tople, Lukas Wutschitz

Research Assistant, RWTH-Aachen — 2014-2018
Student Researcher, DSA Daten- und Systemtechnik GmbH, Aachen — 2016
Software Engineer Intern, A.R. Bayer DSP Systeme GmbH, Düsseldorf — 2012

Teaching

Teaching Assistant, University of Waterloo
- CS458/658: Computer Security and Privacy — 2020,2021
- CS246 - Object Oriented Programming — 2021

Co-Instructor, RWTH-Aachen
- Course: Data-driven Medicine — 2018

Research Talks

Analyzing Privacy Leakage in Language Models, Microsoft Research — 2024
- Host: Robert Sim

Analyzing Privacy Leakage in Language Models, Meta — 2023
Watermarking Generative Models, Google — 2023
- Host: Somesh Jha

Watermarking Generative Models, University of California, Berkeley — 2023
- Host: Dawn Song

Analyzing Privacy Leakage in Language Models, MongoDB — 2023
- Hosts: Marilyn George, Archita Agarwal

Service

Program Committee

- Recent Advances in Intrusion Detection (RAID) 2024

Artifact Evaluation Reviewer

- The ACM Conference on Computer and Communications Security (CCS) 2023

Reviewer

- International Conference on Learning Representations (ICLR) 2024
- International World Wide Web Conference (TheWebConf) 2024
- Recent Advances in Intrusion Detection (RAID) 2023
- Neural Information Processing Systems (NeurIPS) 2022,2023
- International Conference on Machine Learning (ICML) 2022
- The Conference on Information and Knowledge Management (CIKM) 2020

Other

- **Sub-Reviewer**, Proceedings on Privacy Enhancing Technologies (PETS) 2021,2022,2023
- **Session Chair**, IEEE Symposium on Security and Privacy (S&P) 2023
- **Organizing Hackathon**, Workshop on Semantic Web Solutions for Large-Scale Biomedical Data Analytics (SeWeBMeDA) 2018

Student Board Member, Cybersecurity and Privacy Institute 2022-2024

School Advisory Committee on Appointments Liaison, CrySP Lab 2022