NEAL MANGAOKAR

nealmgkr@umich.edu ◇ https://nealmangaokar.com ◇ 571-435-1407

EDUCATION

University of Michigan, Ann Arbor, MI August 2020 - Present
Ph.D. in Computer Science
Funding: NSF Graduate Research Fellowship
Advised by Dr. Atul Prakash

Virginia Tech, Blacksburg, VA August 2016 - May 2020
B.S. in Computer Science - *summa cum laude*
(GPA: 4.0/4.0, Ranked: 1/6789)
Advised by Dr. Bimal Viswanath

RESEARCH INTERESTS

Security, Privacy, and Trustworthiness of ML Systems, Adversarial Robustness, Robustness of Deepfake Detection and Large Language Models.

PUBLICATIONS

PRP: Propagating Universal Perturbations to Attack Large Language Model Guard-Rails
Neal Mangaokar (co-lead), Ashish Hooda (co-lead), Jihye Choi, Shreyas Chandrashekaran, Kassem Fawaz, Somesh Jha, and Atul Prakash
ACL 2024, Bangkok, Thailand, August 2024

D4: Detection of Adversarial Diffusion Deepfakes Using Disjoint Ensembles
Ashish Hooda (co-lead), Neal Mangaokar (co-lead), Ryan Feng, Kassem Fawaz, Somesh Jha, and Atul Prakash
IEEE/CVF WACV, Waikoloa, Hawaii, January 2024

Stateful Defenses for Machine Learning Models Are Not Yet Secure Against Black-box Attacks
Ryan Feng (co-lead), Ashish Hooda (co-lead), Neal Mangaokar (co-lead), Kassem Fawaz, Somesh Jha, and Atul Prakash
ACM CCS, Copenhagen, Denmark, November 2023

Theoretically Principled Trade-off for Stateful Defenses against Query-Based Black-Box Attacks
Ashish Hooda (co-lead), Neal Mangaokar (co-lead), Ryan Feng, Kassem Fawaz, Somesh Jha, and Atul Prakash
ICML AdvML Frontiers Workshop, Honolulu, Hawaii, July 2023

GRAPHITE: Generating Automatic Physical Examples for Machine-Learning Attacks on Computer Vision Systems
Ryan Feng, Neal Mangaokar, Jiefeng Chen, Earlence Fernandes, Somesh Jha, and Atul Prakash
IEEE EuroS&P, Genova, Italy, June 2022

Dispelling Misconceptions and Characterizing the Failings of Deepfake Detection
Neal Mangaokar and Atul Prakash
IEEE S&P Magazine, October 2021

Deepfake Videos in the Wild: Analysis and Detection
Jiameng Pu (co-lead), Neal Mangaokar (co-lead), Lauren Kelly, Parantapa Bhattacharya, Kavya Sundaram, Mobin Javed, Bolun Wang, and Bimal Viswanath
ACM WWW, Ljubljana, Slovenia, August 2021

T-Miner: A Generative Approach to Defend Against Trojan Attacks on DNN-based Text Classification
Ahmadreza Azizi, Ibrahim Asadullah Tahmid, Asim Waheed, Neal Mangaokar, Jiameng Pu, Mobin Javed, Chandan K. Reddy, and Bimal Viswanath
USENIX Security Symposium, Vancouver, CA, August 2021

NoiseScope: Detecting Deepfake Images in a Blind Setting
Jiameng Pu, Neal Mangaokar, Bolun Wang, Chandan K. Reddy, and Bimal Viswanath
ACM ACSAC, Austin, USA, December 2020

Jekyll: Attacking Medical Image Diagnostics using Deep Generative Models
Neal Mangaokar, Jiameng Pu, Parantapa Bhattacharya, Chandan K. Reddy, and Bimal Viswanath
IEEE EuroS&P, Genova, Italy, September 2020

EXPERIENCE

Amazon June 2024 - September 2024
Applied Scientist Intern *Boston, MA*

- Security research on the ESS-Detective team.

University of Michigan Security Lab August 2020 - Present
Graduate Research Assistant *Ann Arbor, MI*

- Developed the first end-to-end jailbreak for LLMs with auxiliary safety guardrails.
- Identified a zero-day in MLaaS stateful defenses, enabling black-box adversarial examples by rejection-sampling from reverse-engineered query distributions.
- Developed a framework for deepfake image detection that is robust to black-box adversarial examples.

Virginia Tech Corporate Research Center August 2018 - August 2020
Undergraduate Research Assistant *Blacksburg, VA*

- Developed GAN-based algorithms for generating medical deepfakes that misled expert physicians from real-world hospitals.
- Conducted large-scale study on the distributional robustness of existing deepfake video detectors.
- Developed one of the first defenses for poisoned text classifiers, using generative seq2seq techniques to detect and extract trigger phrases.

Reinventing Geospatial, Inc. May 2018 - August 2018
Software Engineering Intern *Fairfax, VA*

- Architected and developed scaling geo-data retrieval and visualization system for field infantry.

Virginia Tech October 2017 - May 2018
Undergraduate Teaching Assistant *Blacksburg, VA*

- Hosted programming labs and office hours, and created grading rubrics for course projects.

HONORS

Keynotes and Invited Talks

- Google AI Red Team: Deepfake Detection in an Adversarial Setting (August 2023)
- Google AI Red Team: Adaptive Black-Box Attacks against Stateful Defense Models (October 2023)

Grants, Fellowships, and Scholarships

- ACM CCS Travel Grant: 2023

- National Science Foundation Graduate Research Fellowship: 2022-2027
- University of Michigan Fellowships (Gerstacker Foundation, Michael P. Wellman, Wurman Family, GSS): 2020-2021
- Virginia Tech CS Research Scholarship: 2018-2019
- Virginia Tech CS Advisor Scholarship: 2018-2019
- Virginia Tech's Computer Science Resources Consortium Scholarship: 2017-2018, 2018-2019
- Virginia Tech Investment in Excellence Scholarship: 2017-2018
- Pratt Engineering Scholarship: 2016-2017

Awards

- David Heilman Research Award (Outstanding Undergraduate Research): 2020
- Phi Beta Kappa Academic Honor Society: 2020
- Virginia Tech CS Senior Scholar Award: 2020
- Virginia Tech CS Junior Scholar Award: 2019
- Block.One Blockchain Hackathon Champion Team: 2019
- Virginia Tech CS Sophomore Scholar Award: 2018

TECHNICAL

Languages	Python, Java, C, Javascript, MATLAB, Julia.
Frameworks/Libraries	PyTorch, Tensorflow, Keras, React, JQuery, Dojo, Node.js, Express.js, Electron.js, Bootstrap, Spring.
Technologies	Git, Markdown, LaTeX, Maven, Docker, Webpack/Babel.
Graduate Coursework	Machine Learning, Probability and Random Processes, Algorithms, Computer and Network Security, Distributed Systems