Arjun Nitin Bhagoji

Ph.D. Candidate
Princeton University
Princeton, NJ 08544

(+1)-609-558-2325
abhagoji@princeton.edu
http://www.princeton.edu/∼abhagoji/

Research Interests

My research has mainly focused on the robustness of machine learning systems, based on analyses of the underlying algorithms. I am interested in understanding how intelligent, and possibly malicious agents, can influence machine learning algorithms to achieve their own goals. I aim to use this understanding to guide the development of robust algorithms and the theory of learning in presence of adversaries. I am also interested in the privacy and ethics concerns raised by systems that learn from publicly gathered data.

Education

Program	**Institution**	**Years**
Ph.D., Electrical Engineering	Princeton University Princeton, NJ	Fall 2015 – 2020 *(expected)*
B.Tech. (Honors) and M.Tech., Electrical Engineering (minor in Physics)	Indian Institute of Technology Madras Chennai, India	2010 – 2015

Publications

Papers

- **A. N. Bhagoji***, D. Cullina* and P. Mittal, "Lower Bounds on Adversarial Robustness from Optimal Transport", *33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019 (*Acceptance rate: 21.2%*)
- V. Sehwag*, **A. N. Bhagoji***, L. Song*, C. Sitawarin, D. Cullina, A. Mosenia, P. Mittal and M. Chiang, "Analyzing the Robustness of Open-World Machine Learning", *12th ACM Workshop on Artificial Intelligence and Security (AISec)*, 2019, arXiv:1905.01726 (*Acceptance rate: 23.8%*)
- **A. N. Bhagoji**, S. Chakraborty, P. Mittal and S. Calo, "Analyzing Federated Learning through an Adversarial Lens", *36th International Conference on Machine Learning (ICML)*, 2019 arXiv:1811.12470 (*Acceptance rate: 22.6%*)
- D. Cullina, **A. N. Bhagoji** and P. Mittal , "PAC-learning in the presence of evasion adversaries", *32nd Conference on Neural Information Processing Systems (NeurIPS)*, 2018, arXiv:1806.01471 (*Acceptance rate: 20.8%*)
- C. Sitawarin, **A. N. Bhagoji**, A. Mosenia, M. Chiang and P. Mittal, "Rogue Signs: Deceiving Traffic Sign Recognition with Malicious Ads and Logos", *1st Deep Learning and Security Workshop (co-located with IEEE S&P)*, 2018, arXiv:1801.02780
- **A. N. Bhagoji**, W. He, B. Li and D. Song, "Practical Black-box Attacks on Deep Neural Networks using Efficient Query Mechanisms", *European Conference on Computer Vision (ECCV) 2018*, pp. 154-169, CVF Open Access (*Acceptance rate: 31.8%*)
- **A. N. Bhagoji**, D. Cullina, C. Sitawarin and P. Mittal , "Enhancing robustness of machine learning systems via data transformations" (Invited Paper), *52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018, pp. 1-5, doi
- **A. N. Bhagoji** and P. Sarvepalli, "Equivalence of 2D color codes (without translational symmetry) to surface codes", *International Symposium on Information Theory (ISIT) 2015*, pp. 1109–1114, doi

Workshop presentations and abstracts

- **A. N. Bhagoji**, S. Chakraborty, P. Mittal and S. Calo, "Model Poisoning Attacks in Federated Learning", *NeurIPS 2018 Workshop on Security in Machine Learning* (*Oral, Acceptance rate: 4.6%*)
- V. Sehwag, **A. N. Bhagoji**, C. Sitawarin, A. Mosenia, M. Chiang and P. Mittal, "Not all pixels are born equal: An analysis of evasion attacks under locality constraints", *ACM Computer and Communications Security (CCS) 2018*
- **A. N. Bhagoji**, W. He, B. Li and D. Song, "Black-box Attacks on Deep Neural Networks via Gradient Estimation", *International Conference on Learning Representations (ICLR) 2018 Workshop*, OpenReview
- **A. N. Bhagoji**, "Optimus Prime: Linear Transformations to Secure Machine Learning Systems", *USENIX HotSec 2017*
- **A. N. Bhagoji** and P. Kumar, "Genre classification using centrality measures on a word network", *International Workshop on Complex Networks (CompleNet) 2014*
- N. Sivadas, **A. N. Bhagoji** et. al., "A Nanosatellite Mission to Study Charged Particle Precipitation from the Van Allen Radiation Belts caused due to Seismo-Electromagnetic Emissions", *5th Nano-Satellite Symposium*, 2013, arXiv:1411.6034

Pre-prints and papers under submission

- A.B. Aloshious, **A. N. Bhagoji** and P. Sarvepalli, "On the Local Equivalence of 2D Color Codes and Surface Codes with Applications", arXiv:1804.00866
- C. Sitawarin, **A. N. Bhagoji**, A. Mosenia, M. Chiang and P. Mittal, "DARTS: Deceiving Autonomous Cars with Toxic Signs", arXiv:1802.06430
- **A. N. Bhagoji**, D. Cullina, C. Sitawarin and P. Mittal, "Enhancing Robustness of Machine Learning Systems via Data Transformations", arXiv:1704.02654

Awards & Achievements

- **Recipient of the Yan Huo *94 Graduate Fellowship in Electrical Engineering 2019**
- Travel Grant from the ICML Board to present at ICML 2019 (Long Beach, U.S.A.)
- **Recipient of the Siemens FutureMakers Fellowship in Machine Learning 2018**
- Travel Grant from NIPS Foundation to present at NIPS 2018 (Montréal, Canada)
- **Award for Excellence** (2018) from School of Engineering and Applied Sciences (SEAS) (Princeton University)
- Travel Grant from SEAS (Princeton University) to present at ECCV 2018 (Munich, Germany)
- **Finalist for the Bell Labs Prize 2017 (Top 9 of over 300 participating teams)**
- Travel Grant for USENIX '17 (Vancouver, Canada)
- Travel Grant from SEAS (Princeton University) to present at HotSec '17 (Vancouver, Canada)
- Recipient of the First Year Fellowship (2015-2016) for graduate studies at Princeton University
- Awarded the DAAD WISE Scholarship 2013 for a summer internship in Germany

Projects & Experience

- Graduate Research Assistant at Princeton University [Feb. 2016 - Present]
  advised by Prof. Prateek Mittal working on the security of machine learning systems
- Summer Research Intern at the I.B.M. T.J. Watson Research Center [May - Sept. 2018]
  with Dr. Supriyo Chakraborty studying the robustness of distributed learning systems
- Visiting student at the University of California, Berkeley with Prof. Dawn Song [June - Sept. 2017]
  working on practical and theoretical limits of black-box attacks on machine learning systems
- Undergraduate researcher at I.I.T. Madras with Prof. Pradeep Sarvepalli [May 2014 - July 2015]
  working on equivalences between color codes and surface codes

- Summer research internship at the University of Rostock with Dr. Jan Sperling [May - August 2013] on "Entanglement quasi-probabilities of randomized states"
- Member of the High Energy Particle Detector team of the IIT Madras Student Satellite Project (iitmsat) [March 2011 - Dec. 2012]

Teaching & Mentoring

At Princeton University

- Teaching Assistant for *ELE535: Machine Learning and Pattern Recognition* [Fall 2017]
  - Helped design computational exercises
  - Held office hours and taught material on wide range of topics in machine learning
- **Mentoring**
  - Matteo Russo (B.S.E, Computer Science 2020)
  - Chawin Sitawarin (B.S.E, Electrical Engineering 2019; currently Ph.D. student at UC Berkeley)

At the Indian Institute of Technology, Madras

- Teaching Assistant for *EE5121: Convex Optimization* [Spring 2015]
- Teaching Assistant for *EE5701: Advanced Communications Lab* [Fall 2014]

Professional Service & Development

- Reviewer: CVPR 2020, AAAI 2019, NeurIPS MLITS Workshop 2019, NeurIPS 2019, ICCV 2019, NeurIPS MLITS Workshop 2018, IEEE Transactions on Information Forensics
- Selected for 1$^{\text{st}}$ *Ethics of AI* Professional Development Learning Cohort at Princeton University
- Sub-Reviewer: ACM CCS 2016, 2017, 2018; USENIX Security 2016, 2017, 2018; NDSS 2016, 2017, 2018

Relevant Graduate Coursework

- Information Theory
- Random Graphs and Coding Theory
- Statistical Theory and Methods
- Information Security
- Statistical Learning
- Cryptography
- Privacy and Security for Computing and Communications
- Theoretical Machine Learning
- Sparsity, Structure and Inference
- Fairness in Machine Learning
- Convex Optimization (I.I.T Madras)
- Probability Foundations (I.I.T Madras)

References[1]

- Prateek Mittal,
  Associate Professor,
  Princeton University
- Dawn Song,
  Professor,
  University of California, Berkeley
- Supriyo Chakraborty,
  Research Staff Member,
  IBM T. J. Watson Research Center
- Daniel Cullina,
  Assistant Professor,
  Pennsylvania State University

[1] Contact details available on request