Publications — Blaine Nelson

Theses

- **Behavior of Machine Learning Algorithms in Adversarial Environments (PhD dissertation)**. Blaine Nelson. UC Berkeley, Department of EECS technical report UCB/EECS-2010-140. November 23 2010.
- **Designing, Implementing, and Analyzing a System for Virus Detection. (Master's Thesis)**. Blaine Nelson. UC Berkeley, Department of EECS technical report UCB/EECS-2006-27, March 19 2006.

Journal Papers/Book Chapters

- **Query Strategies for Evading Convex-Inducing Classifiers**. Blaine Nelson, Benjamin I. P. Rubinstein, Ling Huang, Anthony D. Joseph, Steven J. Lee, Satish Rao, and J. D. Tygar. In *Journal of Machine Learning Research* 13(May), pages 1293–1332, 2012. Also accessible as arXiv report arXiv:1007.0484.
- **The Security of Machine Learning**. Marco Barreno, Blaine Nelson, Anthony D. Joseph, and J. D. Tygar. In *Machine Learning Journal* 81(2), pages 121–148, 2010.
- **Misleading learners: Co-opting your spam filter**. Blaine Nelson, Marco Barreno, Fuching Jack Chi, Anthony D. Joseph, Benjamin I. P. Rubinstein, Udam Saini, Charles Sutton, J. D. Tygar, and Kai Xia. Book chapter in Jeffrey J. P. Tsai and Philip S. Yu (eds.) *Machine Learning in Cyber Trust: Security, Privacy, and Reliability*, pages 17–51, 2009.

Conference Papers

- **Poisoning Attacks against Support Vector Machines**. Battista Biggio, Blaine Nelson, and Pavel Laskov. In *Proceedings of the 29th Annual International Conference on Machine Learning (ICML)*, 2012. Also accessible as arXiv report arXiv:1206.6389.
- **Microbagging Estimators: An Ensemble Approach to Distance-weighted Classifiers**. Blaine Nelson, Battista Biggio, and Pavel Laskov. In *Proceedings of the 3rd Asian Conference on Machine Learning (ACML)*, pages 63–79, 2011.
- **Support Vector Machines Under Adversarial Label Noise**. Battista Biggio, Blaine Nelson, and Pavel Laskov. In *Proceedings of the 3rd Asian Conference on Machine Learning (ACML)*, pages 97–112, 2011.
- **Near-Optimal Evasion of Convex-Inducing Classifiers**. Blaine Nelson, Benjamin I. P. Rubinstein, Ling Huang, Anthony D. Joseph, Shing-hon Lau, Steven Lee, Satish Rao, Anthony Tran and J. D. Tygar. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 549–556, 2010. Also accessible as arXiv report arXiv:1003.2751.
- **ANTIDOTE: Understanding and Defending against Poisoning of Anomaly Detectors**. Benjamin I. P. Rubinstein, Blaine Nelson, Ling Huang, Anthony D. Joseph, Shing-hon Lau, Satish Rao, Nina Taft, and J. D. Tygar. In *Proceedings of the 9th Internet Measurement Conference (IMC)*, pages 1–14, 2009.
- **CircuitTSAT: A Solver for Large Instances of the Disjunctive Temporal Problem**. Blaine Nelson and T. K. Satish Kumar. In *Proceedings of the 18th International Conference on Automated Planning and Scheduling (ICAPS)*, pages 232–239, 2008.
- **Revisiting Probabilistic Models for Clustering with Constraints**. Blaine Nelson and Ira Cohen. In *Proceedings of the 24th Annual International Conference on Machine Learning (ICML)*, pages 673–680, 2007.
- **Can Machine Learning Be Secure? (Invited paper)**. Marco Barreno, Blaine Nelson, Russell Sears, Anthony D. Joseph, and J. D. Tygar. In *Proceedings of the ACM Symposium on InformAtion, Computer and Communications Security (ASIACCS)*, pages 16–25, 2006.
- **Analyzing Behavioral Features for Email Classification**. Steven Martin, Anil Sewani, Blaine Nelson, Karl Chen, and Anthony D. Joseph. In *Proceedings of the IEEE 2nd Conference on Email and Anti-Spam (CEAS)*, 2005.
- **A Comparison of Neural Networks and Subspace Detectors for the Discrimination of Low-metal-content Landmines**. Blaine Nelson, Debbie Schofield, Leslie Collins. In *Proceedings of the 8th Detection and Remediation Technologies for Mines and Minelike Targets Conference*, 2003.

Reports / Workshop Papers / Extended Abstracts

- **Machine Learning Methods for Computer Security (Dagstuhl Perspectives Workshop 12371)**. Anthony D. Joseph and Pavel Laskov and Fabio Roli and J. Doug Tygar and Blaine Nelson. In *Dagstuhl Reports* 2(9), ISSN 2192-5283, pages 109–130, 2013.
- **Understanding the Risk Factors of Learning in Adversarial Environments**. Blaine Nelson, Battista Biggio, and Pavel Laskov. In *Proceedings of the 4th ACM Workshop on Artificial Intelligence and Security (AISec)*, 2011.
- **Adversarial Machine Learning (Invited paper)**. Ling Huang, Anthony D. Joseph, Blaine Nelson, Benjamin I. P. Rubinstein, and J. D. Tygar. In *Proceedings of the 4th ACM Workshop on Artificial Intelligence and Security (AISec)*, 2011.
- **Classifier Evasion: Models and Open Problems** Blaine Nelson, Benjamin I. P. Rubinstein, Ling Huang, Anthony D. Joseph, and J. D. Tygar. In *Privacy and Security Issues in Data Mining and Machine Learning*, volume 6549 of LNCS, pages 92–98, 2011.
- **Stealthy Poisoning Attacks on PCA-based Anomaly Detectors**. Benjamin I. P. Rubinstein, Blaine Nelson, Ling Huang, Anthony D. Joseph, Shing-hon Lau, Satish Rao, Nina Taft, and J. D. Tygar. In *ACM SIGMETRICS Performance Evaluation Review* 37(2), pages 73–74, 2009.
- **Open Problems in the Security of Learning**. Marco Barreno, Peter L. Bartlett, Fuching Jack Chi, Anthony D. Joseph, Blaine Nelson, Benjamin I. P. Rubinstein, Udam Saini, and J. D. Tygar. In *Proceedings of the 1st ACM Workshop on Artificial Intelligence and Security (AISec)*, pages 19–26, 2008.
- **Evading Anomaly Detection through Variance Injection Attacks on PCA (Extended Abstract)**. Benjamin I. P. Rubinstein, Blaine Nelson, Ling Huang, Anthony D. Joseph, Shing-hon Lau, Nina Taft, and J. D. Tygar. In *Proceedings of the 11th International Symposium on Recent Advances in Intrusion Detection (RAID)*, pages 394–395, 2008. Winner of the RAID'08 Best Poster Award.
- **Exploiting Machine Learning to Subvert Your Spam Filter**. Blaine Nelson, Marco Barreno, Fuching Jack Chi, Anthony D. Joseph, Benjamin I.P. Rubinstein, Udam Saini, Charles Sutton, J. D. Tygar, and Kai Xia. In *Proceedings of the 1st USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, pages 1–9, 2008.
- **Bounding an Attack's Complexity for a Simple Learning Model**. Blaine Nelson, and Anthony D. Joseph. In *Proceedings of the 1st Workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML)*, 2006
- **User Model Transfer for Email Virus Detection**. Marco Barreno, Blaine Nelson, Russell Sears, and Anthony D. Joseph. In *Proceedings of the 1st Workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML)*, 2006.