Osbert Bastani

10 Comstock Circle, Apt. 304, Stanford, CA 94305 • 972-822-6868 • obastani@stanford.edu

Objective

Pursue research on program analysis, including applications of machine learning and data mining to program analysis, practical applications of static analysis, and techniques for scaling static analyses.

Education

Stanford University
Ph.D. in Computer Science, expected May 2017
Recipient: Stanford School of Engineering Fellowship, 2012-2013
NSF Graduate Research Fellowship Program Honorable Mention, Summer 2011

Harvard University
A.B., *cum laude*, in Mathematics, May, 2012 GPA: 3.81
Recipient: William Lowell Putnam Competition rank 265, Fall 2008

Texas Academy of Mathematics and Science
Class of 2012, National Merit GPA: 4.0
Recipient: Intel Science Talent Search Semifinalist, Spring 2008
Siemens Competition Semifinalist, Fall 2007

Research Experience

STAMP Project (Stanford University), with Saswat Anand & Prof. Alex Aiken, 2012-Present
The goal of this project is to apply static analysis to the detection of Android malware. We use taint analysis to determine whether an Android app leaks sensitive information to untrusted sources. The major technical challenges we address include: (i) writing specifications that describe how data flows through Android library methods, (ii) scaling the static analysis to large apps, and (iii) finding meaningful interpretations of the results. I have designed an algorithm for solving the *missing specifications problem*, i.e. to infer potentially missing library specifications in a sound manner. By using a shortest path context-free language reachability algorithm, my algorithm interactively proposes specifications to further guarantee results that are precise. My tool has generated a large number of useful specifications in the case of taint analysis of Android apps. I have also constructed a novel GUI tool to facilitate the interactive analysis of Android apps. Other projects I am working on include inferring specifications from tests, and exploring various techniques for modularizing the static analyses to improve scalability.

DeepDive Project (Google), with Domagoj Babic & Monirul Sharif, 2014
The goal of this project is to create a scalable framework for performing static analysis of Android apps. I have designed approaches for handling various Android specification functionality including concurrent and asynchronous language features, Android application life cycle, and inter-component calls. I have also implemented single-static assignment construction, including dominator tree, dominance frontier computations, and live variables analysis. Finally, I have designed and implemented the flow insensitive interprocedural analysis framework, as well as points-to analysis, string analysis, taint analysis, and reachability analysis. I have designed approaches to detecting specific families of malicious apps based on this framework.

Interactive Recommendations (Technicolor Labs), with Branislav Kveton & Udi Weinsberg, 2013-Present
Current recommender systems are largely static, at best giving users predetermined filters for narrowing down results. Generalized binary search algorithms can do so by interactively querying the user's preferences over a set of binary features. However, in practice different users have different beliefs about the features of a given item. I have modeled this variation by treating entries in the binary feature matrix as independent Bernoulli random variables with known priors, and I have designed an efficient algorithm to perform the generalized binary search in the stochastic setting. Finally, I have built a system that extracts natural language features from Yelp reviews and performs the resulting interactive recommendation algorithm.

Zero-Shot Learning (Stanford University), with Richard Socher & Prof. Christopher Manning, 2013
I have worked on approaches to zero-shot learning, specifically using cross-modal transfer to allow classification of objects in unobserved categories. Specifically, I have helped design an approach to the problem that detects "novel" objects and classifies these based on the closest word vector corresponding to a previously unobserved class.

Reduced Precision Fast Fourier Transforms (Caltech), with Vladimir Dergachev, 2010
I have designed an algorithm for computing reduced precision fast Fourier transforms (FFTs) by approximating complex multiplication to finite quotients of number fields and then performing number theoretic transforms. I have demonstrated that my technique allows speedups over performing the full precision FFT without sacrificing ability to detect signals.

Homomorphic Encryptions (University of Texas at Dallas), with Liangliang Xiao, 2005-2009
Homomorphic encryption is a form of encryption that allows clients to perform computations on encrypted data without decrypting the data, which has important applications to cloud computing, power grid management, sensor data aggregation, and more. Existing schemes require a high degree polynomial computational overhead. We develop a homomorphic encryption

scheme that sacrifices semantic security for the advantage of being practically implementable, introducing only a small constant factor of computational overhead. We prove that the encryption scheme is secure against bounded plaintext attacks.

Real Root Counting (Texas A&M University), with Dimitar Popov, Chris Hillar, & Prof. Maurice Rojas, 2008
Characterizing and finding solutions to systems of polynomial equations is an important problem with applications to phylogenetics, control theory, rational drug design, and numerous other problems. The goal of this project was to understand the set of solutions to *sparse* polynomial systems, where each polynomial has few coefficients (but possibly large degree). We discovered a parametrization of the amoeba of the reduced discriminant variety of a sparse polynomial system, which allowed us to develop efficient algorithms for determining their zeros loci.

Publications

O. Bastani, S. Anand and A. Aiken, "Specification Inference Using Context-Free Language Reachability", Proceedings of the Symposium on the Principles of Programming Languages, 2015, to appear.

R. Socher, M. Ganjoo, H. Sridhar, O. Bastani, C. D. Manning, A. Y. Ng: Zero-Shot Learning Through Cross-Modal Transfer. CoRR abs/1301.3666 (2013).

O. Bastani, C. Hillar, D. Popov, M. Rojas, "Randomization, Sums of Squares, Near-Circuits, and Faster Real Root Counting", *Randomization, Relaxation, and Complexity in Polynomial Equation Solving*, Contemporary Mathematics, AMS Press, 2011.