

Baishakhi Ray

*Assistant Professor
Department of Computer Science
University of Virginia
Charlottesville, VA - 22904*

*Voice: (303) 748-2958
Email: rayb@virginia.edu
Homepage: <http://rayb.info>*

RESEARCH INTEREST

I am primarily interested in Software Engineering research with a focus on improving software reliability and security. In particular, I devise novel program analysis techniques to analyze existing code properties and apply advanced machine learning and NLP models to learn from those properties. Such models help me building tools that automate program development, bug detection, and program repair for real-world large scale software.

ACADEMIC POSITIONS

University of Virginia, VA, USA
Assistant Professor

October 2015 - present

University of California Davis, CA, USA
Postdoctoral Research Fellow
Advisor: Prem Devanbu

October 2013 - September 2015

The University of Texas at Austin, TX, USA
Graduate Research Assistant
Adviser: Miryung Kim

January 2011 - May 2013

EDUCATION

- PostDoc, Computer Science, University of California, Davis. 2013 - 2015
Advisor: Prem Devanbu
- PhD, Electrical & Computer Engineering, University of Texas, Austin. 2010-2013 (GPA: 3.97)
Thesis: Analysis of Cross-System Porting and Porting Errors in Software Projects
Advisor: Miryung Kim
- MS, Computer Science, University of Colorado, Boulder. 2007-2009 (GPA: 4.0)
Thesis: SecureWear: Securing Wearable Mobile Social Networks
Advisor: Richard Han
- B.Tech., Computer Science, Calcutta University, India. 2001-2004
B.Sc., Physics, Presidency College, Kolkata, India. 1998-2001 (GPA: 3.84)
Top 1% in University

HONORS & AWARDS

- NSF CHS: Small: Translating Compilers for Visual Computing in Dynamic Languages
- NSF TWC: Small: Automated Detection and Repair of Error Handling Bugs in SSL/TLS Implementations
- Best Paper Award, SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE 2017)
- Best Paper Award, International Conference on Mining Software Repositories, (MSR 2017)
- CACM Research Highlight, 2017
- Best Practical Paper Award, IEEE Symposium on Security and Privacy (S&P Oakland), 2014
- Nominated for Distinguished Paper Award, Automated Software Engineering (ASE, 2013).

- Google Summer of Code, 2012
- Ranked 6th out of 15,000 students in B.Sc. Physics (Hons.) examination.
- Jawaharlal Nehru Summer Scholarship for Advanced Research, India 2001 (100 students are selected nationwide).

OTHER PROFESSIONAL EXPERIENCE

Microsoft Research, Redmond, USA Research Intern Mentors: Christian Bird, Nachiappan Nagappan, Thomas Zimmermann	May 2013–August 2013
Google Summer of Code, Google Inc. Research Intern Mentors: Suzette Person, Neha Rungta, NASA	May 2012 – August 2012
Avaya Research Lab, Westminster, CO, USA Research Intern	May 2008 – Aug 2008
Ericsson Pvt. Ltd. Boulder, CO, USA Software Engineer	February 2009–June 2010
Ixia, Sasken, and Texas Instruments, India Software Engineer	August 2004–July 2007

CONFERENCE PUBLICATIONS

(Top Tier conferences are marked in **Blue**)

2018

1. DeepTest: Automated Testing of Deep-Neural-Network-driven Autonomous Cars. 10 pages, Y. [Tian](#), K. Pei, S. Jana, **B. Ray**. In 40th International Conference on Software Engineering (**ICSE'18**), May'16, 10 pages, acceptance rate: 20.9%. (advisor and lead faculty author)

2017

2. Automatically Diagnosing and Repairing Error Handling Bugs in C, [Y. Tian](#), **B. Ray**, 10 pages. In 11th joint meeting of the European Software Engineering conference and the ACM Sigsoft Symposium on the Foundations of Software (**ESEC-FSE'17**), September'17, acceptance rate: 24.4%. **Best Paper Award**. (advisor and lead faculty author)
3. GitcProc: A Tool for Processing and Classifying GitHub Commits, C. Casalnuovo, Y. Suchak, **B. Ray**, C. Rubio-Gonzalez, 4 pages, July'17, In International Symposium on Software Testing and Analysis (ISSTA'17). (collaborator)
4. Some From Here, Some From There: Cross-Project Code Reuse in GitHub, M. Gharehyazie, **B. Ray**, V. Filkov, 10 pages. The 14th International Conference on Mining Software Repositories (MSR'17), May'17, acceptance rate: 27%. **Best Paper Award**. (co-supervisor and 2nd lead author)
5. A Large Scale Study of Programming Languages and Code Quality in Github. **B. Ray**, D. Posnett, P. T. Devanbu, V. Filkov. **CACM Research Highlights**. October'17. (lead author)

Ongoing Work

6. Which Similarity Metric to Use for Software Documents? A study on Information Retrieval based Software Engineering Tasks. Md M. Rahman, S. Chakraborty, **B. Ray** (advisor and lead faculty author). Nominated for ICSE, 2018 poster track.
7. NNLRank: A Recommender System for Developer Onboarding, C. Liu, D. Yang, X. Zhang, H. Hu, J. Barson, **B. Ray** (lead faculty author)
8. Searching for Performance: Metaheuristic Methods for Improving Software Configurations. C. Tang, K. Sullivan, B. Ray (co-advisor). Nominated for ICSE, 2018 poster track.
9. Deobfuscating Android Applications through Deep Learning. F. Su, J. Bell, G. Kaiser, **B. Ray** (lead faculty author)
10. Entropy Guided Spectrum Based Bug Localization Using Statistical Language Model. S. Chakraborty, Y. Li, M. Irvine, R. Saha, **B. Ray** (advisor and lead faculty author)

2016

11. On the Naturalness of Buggy Code. **B. Ray**, V. Hellendoorn, S. Godhane, Z. Tu, A. Bacchelli, P. Devanbu, 10 pages. In 38th International Conference on Software Engineering (**ICSE'16**), May'16, 10 pages, acceptance rate: 19%. (lead author)
12. Automatically Detecting Error Handling Bugs using Error Specifications. S. Jana, Y. J. Kang, S. Roth, **B. Ray**. In **USENIX Security '16**, August'16, 18 pages, acceptance rate: 15.5%. (2nd lead author)
13. APEx: Automated Inference of Error Specifications for C APIs. Y. J. Kang, **B. Ray**, S. Jana. In 31st IEEE/ACM International Conference on Automated Software Engineering (**ASE'16**), November'16, 10 pages, acceptance rate: 19.1%. (lead faculty author)

2015

14. Assert Use in GitHub Projects. C. Casalnuovo, P. Devanbu, A. Oliveira, V. Filkov, **B. Ray**, 10 pages. In 37th International Conference on Software Engineering (**ICSE'15**), May'15, 10 pages, acceptance rate: 18.5%. (lead author)
15. Gender and Tenure Diversity in GitHub Teams. B. Vasilescu, D. Posnett, **B. Ray**, M. Brand, A. Serebrenik, P. Devanbu, V. Filkov, In International Conference on Human Factors in Computing Systems (**CHI'15**), April'15, 10 pages, acceptance rate: 23%. (collaborator)

2014

16. A Large Scale Study of Programming Languages and Code Quality in Github. **B. Ray**, D. Posnett, V. Filkov, P. T. Devanbu. In ACM SIGSOFT, 22nd International Symposium on the Foundations of Software Engineering (**FSE'14**), November'14, pages: 155-165, acceptance rate: 22%, **Selected for publication in the "Research Highlights" section of the CACM.** (lead author)
17. Using Frankencerts for Automated Adversarial Testing of Certificate Validation in SSL/TLS Implementations. C. Brubaker, S. Jana, B. Ray, S. Khurshid, and V. Shmatikov. In 35th IEEE Symposium on Security and Privacy, 2014 (**S&P Oakland'14**), May'14, pages: 114-129, acceptance rate: 13%, **Best Practical Paper Award.** (collaborator)

2013

18. Detecting and Characterizing Semantic Inconsistencies in Ported Code. B. Ray, M. Kim, S. Person, N. Rungta. In 28th IEEE/ACM International Conference on Automated Software Engineering, 2013 ([ASE'13](#)), pages: 367-377, acceptance rate: 23%, November'13, Nominated for Distinguished Paper Award, invited for ASE journal special issue. (lead author)
19. An Empirical Study of API Stability and Adoption in the Android Ecosystem. T. McDonnell, B. Ray, M. Kim. In 29th IEEE International Conference on Software Maintenance, 2013 (ICSM'13), September'13, pages: 70-79, acceptance rate: 22%. (co-supervisor)

2012

20. Case Study of Cross-System Porting in Forked Projects. B. Ray, M. Kim. In ACM SIGSOFT, the 20th International Symposium on the Foundations of Software Engineering ([FSE'12](#)), November'12, 10 pages, acceptance rate:17%. (lead author)
21. Repertoire: A Cross-System Porting Analysis Tool for Forked Software Projects. B. Ray, C. Wiley, M. Kim. In ACM SIGSOFT the 20th International Symposium on the Foundations of Software Engineering, Formal Research Tool Demonstration (FSE'12), November'12, pages: 8:1-8:4. (lead author)
22. An Empirical Study of Supplementary Bug Fixes. J. Park, M. Kim, B. Ray, D. Bae. In The 9th IEEE Working Conference on Mining Software Repositories (MSR'12), June'12, pages: 40-49, acceptance rate: 28% Invited to the Special Issue of Journal of Empirical Software Engineering (EMSE). (co-author)

2011

23. PTask: Operating System Abstractions To Manage GPUs as Compute Devices. CJ. Rossbach, J. Currey, M. Silberstein, B. Ray, E. Witchel. In Proceedings of the 23rd ACM Symposium on Operating System Principles ([SOSP'11](#)), October'11, pages: 233-248, acceptance rate: 17%. (co-author)

OLDER

24. Touch Me [wE@r: Getting Physical with Social Networks](#), A. Beach, B. Ray, L. Buechley. In 2009 Workshop on Sensor-based Models and Feedback Systems for Social Computing. Associated with SocialCom 2009, pages: 960-965. A Protocol for Building Secure and Reliable Covert Channel. B. Ray and S. Mishra. In 6th Annual Conference on Privacy, Security and Trust, 2008. (PST'08), pages: 246-253
25. WhozThat? Evolving an Ecosystem for Context-Aware Mobile Social Networks. A. Beach, B. Ray, et al., In IEEE Network Magazine Special Issue on Composable context aware services, 2008, pages: 50-55.

BOOK CHAPTER

1. A Large Ecosystem Study to Understand the Effect of Programming Languages on Code Quality. B. Ray, D. Posnett. Perspectives on Data Science for Software Engineering, Morgan Kaufmann, 2016. (lead author)
2. SecureWear: A Framework for Securing Mobile Social Networks. B. Ray, R. Han. Advances in Computer Science and Information Technology. Computer Science and Engineering, Vol. 85, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer Berlin Heidelberg, 2012, pages: 515-524. (lead author)

CITATION COUNT: 949

H-Index: 13

STUDENTS

PhD students

- Chong Tang (co-advised with Kevin Sullivan): passed proposal
- Yuchi Tian: 3rd year (passed qualification exam in Fall 2017)
- Md Masudur Rahman: 3rd year (will appear in qualification exam in Fall 2017)
- Saikat Chakraborty: 2nd year

Masters students

- Chengjun Yuan: graduated (1st job Facebook)

Undergraduate students

- Matt Irvine: graduated (1st job Microsoft)
- Ketao Yin: graduated (now graduate student at Georgia Tech)
- Yujian Li
- Jed Barson

FUNDING

Awarded

1. NSF TWC: Small: Collaborative: Automated Detection and Repair of Error Handling Bugs in SSL/TLS Implementations.
 - \$250,000.00
 - Effective date of 09/01/16.
 - Award Duration: 36 (months) amounts
 - I as a lead PI was responsible for all the aims, in particular for bug detection using symbolic execution and its repair.
 - Co-PI: Suman Jana, Columbia University
2. NSF CHS: Small: Translating Compilers for Visual Computing in Dynamic Languages.
 - \$450,000.00
 - Effective date of 09/01/16.
 - Award Duration: 36 (months) amounts
 - I as a Co-PI was responsible to learn program transformation rule from existing code.
 - PI: Connelly Barnes (UVa), Co-PI: Westley Weimer (formerly at UVa)

Pending

1. SHF: Small: Search-based Tuning of Big Data Software Systems
2. SaTC: CORE: Medium: Collaborative: Finding Semantic Security Bugs with Pseudo-Oracle Testing

INVITED TALKS:

- NYU, 2017: Leveraging Big Code to Improve Software Reliability
- Rutgers, 2017: Leveraging Big Code to Improve Software Reliability
- College of William & Mary, 2017: Leveraging Big Code to Improve Software Reliability
- Fujitsu Laboratories America, Sunnyvale, CA, May 2016: On the Naturalness of Bug.
- Fujitsu Laboratories America, Sunnyvale, CA, January 2014: Analysis of Cross-System Porting and Porting Errors in Software Projects.
- NASA Ames Research Center, Mountain View, CA, November 2013: Detecting and Characterizing Semantic Inconsistencies in Ported Code.

- IBM Research, March 2013, Delhi, India: A Case Study of Cross-Systems Porting in Forked Projects.

TEACHING EXPERIENCE

COURSES

- Spring 2018: Program Analysis for Robust Software (CS 4501)
- Spring 2017 : [Empirical Software Engineering](#) (CS 6501)
- Fall 2016 : [Software Engineering](#) (CS 6240)
- Spring 2016 : [Data Science In Software Engineering](#) (CS 6501)

GUEST LECTURER

- Introduction to Programming and Problem Solving, Fall 2014, University of California, Davis
- Software Engineering, Fall 2013, University of California, Davis
- Software Engineering and Design Laboratory, Fall 2013, The University of Texas at Austin

SERVICE PERFORMANCE

DEPARTMENT SERVICE:

- General faculty recruitment committee, 2016 & 2017
- Computer Science Chair Reappointment Committee, 2017
- IT manager search committee
- CS Systems Engineer search committee
- Served in 2 qualifications, 3 proposals, and 1 PhD defense committee.

CONFERENCE PROGRAM COMMITTEE:

- 2020: Co-chair, FSE tool demo track
- 2018: ASIACCS, ASE, ICPC
- 2017: ASE (ERP), ICSE NIER, OOPSLA Onward, ICPC, MSR, ICSME.
- 2016: ASE (ERP), MSR, ICSE Visions of 2025 and Beyond (V2025), ICSE Tool Demonstrations, FSE SRC, FSE NL+SE Workshop, ICSME Era, ISEC, APSEC.
- 2015: MSR, MSR Challenge, ISEC.
- 2014: FSE Artifact.
- 2013: OOPSLA Artifact.

JOURNAL REVIEWER:

- Transactions on Software Engineering (TSE)
- Empirical Software Engineering (EMSE)
- ACM Transactions on Software Engineering and Methodology (TOSEM)
- IEEE Transactions on Dependable and Secure Computing (TDSC)

OTHER:

- Serve NSF panel.
- Serve as NDSEG reviewer.
- Organized NL+SE 2016 Workshop, co-located with FSE 2017.

MEDIA COVERAGE

- Language Study: [SlashDot](#), [The Register](#), [Reddit](#), [InfoWorld](#), [Hacker News](#).
- Frankencerts : [Reddit](#), [Golem](#), [Heise](#).