

Fan Long

Massachusetts Institute of Technology
32 Vassar Street, 32-G730
Cambridge, MA, 02139
☎ +1 (617) 253 7768
✉ fanl@csail.mit.edu
🏠 people.csail.mit.edu/fanl

Research Interests

My areas of interest are software engineering, programming language, and security. My research to date focuses on developing automated programming systems that combine program analysis, runtime systems, and machine learning to improve software reliability and security.

Education

- Summer 2017 **Ph.D., Electrical Engineering and Computer Science,**
(Expected) *Massachusetts Institute of Technology*, Cambridge, MA.
Advisor: Prof. Martin Rinard
- September 2012 **S.M., Electrical Engineering and Computer Science,**
Massachusetts Institute of Technology, Cambridge, MA.
Advisor: Prof. Martin Rinard
- July 2010 **B.S., Computer Science, Tsinghua University**, Beijing, China.
Enrolled in Tsinghua Xuetaang Special Pilot CS Class (Directed by Prof. Andrew Yao)

Publication Overview

I have published 16 peer-reviewed conference papers (7 as the first author) in venues such as PLDI, ICSE, POPL, FSE, CCS, ASPLOS, ISSTA, NDSS, and NSDI. Three selected papers are listed below (full list of papers is available on page 4).

- POPL'16 Fan Long and Martin Rinard. Automatic Patch Generation by Learning Correct Code. In *Symposium on Principle of Programming Languages*, 2016.
- PLDI'14 Fan Long, Stelios Sidiroglou, and Martin Rinard. Automatic Runtime Error Repair and Containment via Recovery Shepherding. In *Conference on Programming Language Design and Implementation*, 2014.
- ICSE'12 Fan Long, Vijay Ganesh, Michael Carbin, Stelios Sidiroglou, and Martin Rinard. Automatic Input Rectification. In *International Conference on Software Engineering*, 2012.

Research Experience

Projects as the Lead Contributor at MIT

- 2014-present **Automatic Patch Generation via Learning from Human Patches:**
Designed and implemented a series of automatic patch generation systems [2, 5, 6] including Prophet [2], the state-of-the-art patch generation system for C. Prophet learns from past successful human patches to recognize correct patches during the patch generation process.
- 2014-present **Quantitative Analysis of Patch Generation Techniques:**
Systematically analyzed previous patch generation techniques and identified several critical issues of the previous techniques, i.e., the weak test suite and search space explosion [1, 6]. The analysis results inspired the development of many new patch generation systems [2, 5].

2013-2014 **Lightweight Error Recovery via Recovery Shepherding:**
Designed and implemented RCV [10], a lightweight dynamic program analysis and runtime system for enabling software applications to recover from fatal null-dereference and divide-by-zero errors. It works with off-the-shelf binary programs and incurs negligible overhead.

2012-2014 **Sound Input Filtering for Integer Overflow Errors:**
Designed and implemented SIFT [9], the first sound input filter system for critical integer overflow errors. SIFT guarantees that if an input passes the filter, it will not trigger integer overflow errors at any analyzed critical sites.

2010-2012 **Automatic Input Rectification:**
Designed and implemented SOAP [12], the first automatic input rectification system that changes atypical inputs, instead of discarding them, to preserve desirable data in atypical but benign inputs.

Projects with Major Contributions at MIT

2014-2015 **Automatic Code Transfer:**
Helped design and implement CodePhage [7], the first patch generation tool that transfers correct code across applications.

2014-2015 **Attacks on the Weakness of Control Flow Integrity:**
Analyzed the imprecision of the control flow graphs (CFG) constructed by pointer analysis algorithms. The analysis enables control-jujutsu attacks on fine-grained control flow integrity (CFI) techniques that rely on such imprecise CFGs [3].

2012-2015 **Automatic Overflow Bug Finding:**
Helped design and implement DIODE [8], a concolic execution tool that finds critical integer overflow errors at memory allocation sites.

2010-2013 **Automatic Program Generation from Natural Language:**
Helped design and implement NL2P [11], a novel synthesis tool that automatically generates input parser programs from natural language descriptions of the input format.

Projects before MIT

2008-2009 **Automatic API Documentation Generation:**
Designed and implemented Altair [15], a novel API documentation generation tool. It uses a combination of machine learning and program analysis techniques to generate cross-reference sections for API documents.

2008-2010 **Targeted Record and Replay for Debugging:**
Designed and implemented iTarget [14], a lightweight record and replay system that enables targeted replay of specific components of a software application.

2008 **Model Checking of Distributed Systems**
Helped implement MODIST [16], a model checking tool that automatically finds critical errors in real world large distributed systems.

Press

Automatic Patch Generation via Learning from Human Patches [2].

- **PCWorld:** <https://goo.gl/9J8QPV>
- **MITNews:** <https://goo.gl/vpddXv>

Automatic Integer Overflow Bug Finding [8].

- **MITNews:** <https://goo.gl/hxwrTk>

Automatic Code Transfer [7].

- **Fortune:** <https://goo.gl/WsTkdz>
- **Slashdot:** <https://goo.gl/CGyY6U>
- **MITNews:** <https://goo.gl/KIsm1W>

Parser Generation from Natural Language Description [11].

- **MITNews:** <https://goo.gl/NwGePp>
- **Upstart Business Journals:** <https://goo.gl/aXiRiX>

Awards

- 2008 Silver Medal (6th place) and Asia Champion in ACM-ICPC World Finals.
- 2007 Gold Medal (1st place) in ACM-ICPC Changchun Regional.
- 2006 Gold Medal (3rd place) in 18th International Olympiad in Informatics.
- 2005 Gold Medal in 17th International Olympiad in Informatics.

Teaching, Mentoring, and Grant

- Spring 2016 **Teaching Assistant for MIT undergraduate course 6.035: Computer Language Engineering.** I prepared projects materials, run project info session, held office hours, prepared and graded two midterm and final exams, and gave two lectures: *Foundations of Data-flow Analysis* and *Instruction Scheduling*.
- Fall 2015 **Substitute Lecturer for MIT undergraduate course 6.033: Computer Systems Engineering.** I presented two lectures: *Why Cryptosystems Fail* and *Bitcoin*.
- Fall 2006 **Teaching Assistant for Tsinghua undergraduate course: Introduction to Programming.** I prepared programming problem set and helped students to debug their submitted programs.
- 2012-present **Student Mentor.** I advised multiple high school, undergraduate, master, and junior graduate students while at MIT. I help advised one master thesis, “An Analysis of Patch Plausibility and Correctness for Generate-And-Validate Patch Generation Systems” by Zichao Qi.
- 2007-2008 **National Programming Contest Team Coach.** I prepared training programs and organized selection contests for Chinese National Team for the International Olympiad of Informatics (IOI) 2007 and 2008
- 2010-present **DARPA Funding: MRC and MUSE projects.** Worked on the grant proposal, attended PI meetings, and presented results on PI meetings and site visits.

Service

- Reviewer
 - **PLDI:** Programming Language Design and Implementation (2016, ERC)
 - **POPL:** Principle of Programming Languages (2017)
 - **ESE:** Empirical Software Engineering, Springer
 - **RGCHK:** Research Grants Council of Hong Kong
- Member **MIT EECS Graduate Admission Student Committee,** Reviewed application materials of new candidate graduate students, interviewed the candidate students, and rendered opinions on the candidates.
- Volunteer **HackMIT,** Served as a graduate student supervisor for hackathon (Fall 2013).

Publications

- [1] Fan Long and Martin Rinard. An analysis of the search spaces for generate and validate patch generation systems. In *Proceedings of the 38th International Conference on Software Engineering, ICSE 2016*, pages 702–713, New York, NY, USA, 2016. ACM.
- [2] Fan Long and Martin Rinard. Automatic patch generation by learning correct code. In *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2016*, pages 298–312, New York, NY, USA, 2016. ACM.
- [3] Isaac Evans, Fan Long, Ulziibayar Otgonbaatar, Howard Shrobe, Martin Rinard, Hamed Okhravi, and Stelios Sidiroglou-Douskos. Control jujutsu: On the weaknesses of fine-grained control flow integrity. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security, CCS 2015*, pages 901–913, New York, NY, USA, 2015. ACM.
- [4] Brendan Juba, Christopher Musco, Fan Long, Stelios Sidiroglou-Douskos, and Martin C. Rinard. Principled sampling for anomaly detection. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2015*, 2015.
- [5] Fan Long and Martin Rinard. Staged program repair with condition synthesis. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering, ESEC/FSE 2015*, pages 166–178, New York, NY, USA, 2015. ACM.
- [6] Zichao Qi, Fan Long, Sara Achour, and Martin Rinard. An analysis of patch plausibility and correctness for generate-and-validate patch generation systems. In *Proceedings of the 2015 International Symposium on Software Testing and Analysis, ISSTA 2015*, pages 24–36, New York, NY, USA, 2015. ACM.
- [7] Stelios Sidiroglou-Douskos, Eric Lahtinen, Fan Long, and Martin Rinard. Automatic error elimination by horizontal code transfer across multiple applications. In *Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2015*, pages 43–54, New York, NY, USA, 2015. ACM.
- [8] Stelios Sidiroglou-Douskos, Eric Lahtinen, Nathan Rittenhouse, Paolo Piselli, Fan Long, Deokhwan Kim, and Martin Rinard. Targeted automatic integer overflow discovery using goal-directed conditional branch enforcement. In *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS 2015*, pages 473–486, New York, NY, USA, 2015. ACM.
- [9] Fan Long, Stelios Sidiroglou-Douskos, Deokhwan Kim, and Martin Rinard. Sound input filter generation for integer overflow errors. In *Proceedings of the 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2014*, pages 439–452, New York, NY, USA, 2014. ACM.
- [10] Fan Long, Stelios Sidiroglou-Douskos, and Martin Rinard. Automatic runtime error repair and containment via recovery shepherding. In *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2014*, pages 227–238, New York, NY, USA, 2014. ACM.
- [11] Tao Lei, Fan Long, Regina Barzilay, and Martin C. Rinard. From natural language specifications to program input parsers. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*, pages 1294–1303, 2013.
- [12] Fan Long, Vijay Ganesh, Michael Carbin, Stelios Sidiroglou, and Martin Rinard. Automatic input rectification. In *Proceedings of the 2012 International Conference on Software Engineering, ICSE 2012*, pages 80–90. IEEE Press, 2012.
- [13] Zhenyu Guo, Dong Zhou, Haoxiang Lin, Mao Yang, Fan Long, Chaoqiang Deng, Changshu Liu, and Lidong Zhou. G2: A graph processing system for diagnosing distributed systems. In *Proceedings of the 2011 USENIX Conference on USENIX Annual Technical Conference, USENIXATC 2011*, pages 27–27, Berkeley, CA, USA, 2011. USENIX Association.

- [14] Ming Wu, Fan Long, Xi Wang, Zhilei Xu, Haoxiang Lin, Xuezheng Liu, Zhenyu Guo, Huayang Guo, Lidong Zhou, and Zheng Zhang. Language-based replay via data flow cut. In *Proceedings of the Eighteenth ACM SIGSOFT International Symposium on Foundations of Software Engineering*, FSE 2010, pages 197–206, New York, NY, USA, 2010. ACM.
- [15] Fan Long, Xi Wang, and Yang Cai. Api hyperlinking via structural overlap. In *Proceedings of the 7th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering*, ESEC/FSE 2009, pages 203–212, New York, NY, USA, 2009. ACM.
- [16] Junfeng Yang, Tisheng Chen, Ming Wu, Zhilei Xu, Xuezheng Liu, Haoxiang Lin, Mao Yang, Fan Long, Lintao Zhang, and Lidong Zhou. Modist: Transparent model checking of unmodified distributed systems. In *Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation*, NSDI 2009, pages 213–228, Berkeley, CA, USA, 2009. USENIX Association.

Talks

Automatic Patch Generation by Learning Correct Code [2]

- Sep. 2016 DARPA MUSE PI Meeting, Santa Cruz, CA
- July 2016 Invited talk, Peking University, Beijing
- Mar. 2016 Invited talk, Georgia Tech, Atlanta, GA
- Jan. 2016 CREST Open Workshop, University College London, London
- Jan. 2016 Principle of Programming Languages (POPL), St. Petersburg, FL
- July 2015 DARPA MUSE PI Meeting, Menlo Park, CA

An Analysis of the Search Spaces for Patch Generation Systems [1]

- May 2016 International Conference on Software Engineering (ICSE), Austin, TA

The Weaknesses of Fine-Grained Control Flow Integrity [3]

- Oct. 2015 Conference on Computer and Communications Security (CCS), Denver, CO

Staged Program Repair [5]

- Sep. 2015 Foundations of Software Engineering (FSE), Bergamo
- July 2015 Invited talk, Tsinghua University, Beijing

Runtime Error Repair and Containment via Recovery Shepherdng [10]

- Apr. 2015 DARPA MRC PI Meeting, Melo Park, CA

Sound Input Filter Generation for Integer Overflow Errors [9]

- Jan. 2014 Principle of Programming Languages (POPL)
- Jan. 2014 DARPA MRC, PI Meeting, San Deigo, CA
- May 2013 DARPA MRC PI Meeting, Park Ridge, NJ (Poster)

Automatic Input Rectification [12]

- Dec. 2012 Invited talk, Peking University, Beijing
- June 2012 International Conference on Software Engineering (ICSE), Zurich, Switzerland
- Nov. 2011 DARPA MRC Kickoff Meeting, Crystal City, VA (Poster)

API Hyperlinking via Structural Overlap [15]

- Aug. 2009 Foundations of Software Engineering (FSE), Amsterdam
- July 2009 Microsoft Research Asia, Beijing

References

Martin Rinard
Professor
EECS and CSAIL
Massachusetts Institute of Technology
32-G828, 32 Vassar Street
Cambridge, MA, 02139
✉ rinard@csail.mit.edu
☎ 617-258-6922

Regina Barzilay
Professor
EECS and CSAIL
Massachusetts Institute of Technology
32-G468, 32 Vassar Street
Cambridge, MA, 02139
✉ regina@csail.mit.edu
☎ 617-258-5706

Stelios Sidiroglou-Douskos
Research Scientist
CSAIL
Massachusetts Institute of Technology
32-G728, 32 Vassar Street
Cambridge, MA, 02139
✉ stelios@csail.mit.edu
☎ 917-664-4965

Suresh Jagannathan
Professor
Department of Computer Science
Purdue University
305 North University Street
West Lafayette, IN, 47907
✉ suresh@cs.purdue.edu
☎ 765-494-0971

Armando Solar-Lezama
Associate Professor
EECS and CSAIL
Massachusetts Institute of Technology
32-G840, 32 Vassar Street
Cambridge, MA, 02139
✉ asolar@csail.mit.edu
☎ 617-258-9727