Lin Tan

Department of Computer Science
201 N Goodwin Ave, Urbana, IL 61801

Email: lintan2@cs.uiuc.edu
http://opera.cs.uiuc.edu/~ lintan2

Gender: Female
Visa: F-1

RESEARCH INTERESTS

Systems; Software reliability; Software engineering; Security; Computer architecture; Focusing on leveraging comments or applying natural language processing techniques to address systems reliability problems

EDUCATION

University of Illinois at Urbana-Champaign Ph.D. Candidate in Computer Science, 01/2005–08/2009
Advisor: Professor Yuanyuan Zhou.

University of California, Santa Barbara Ph.D. Student in Computer Science, 09/2003–12/2004

Zhejiang University, China B.S. in Computer Science, 09/1999–06/2003

SELECTED AWARDS AND HONORS

- IBM Ph.D. Fellowship, 2008 - 2009
- Verizon Fellowship, 2006 - 2008
- Paper selected for the IEEE Micro's Top Picks from Computer Architecture Conferences, 2006
- Outstanding Graduate, Zhejiang Province, 2003 – Awarded to top 1% of graduating seniors

PUBLICATIONS

JOURNAL ARTICLES

1. IEEE Micro'06 **Lin Tan** and Timothy Sherwood. "Architectures for Bit-Split String Scanning in Intrusion Detection." *IEEE Micro: Micro's Top Picks from Computer Architecture Conferences*, January-February 2006. Acceptance Rate: 16% (13/80). *Cited by 13 according to Google Scholar*.

2. ACM TACO'06 **Lin Tan**, Brett Brotherton and Timothy Sherwood. "Bit-Split String Matching Engines for Intrusion Detection and Prevention." *ACM Transactions on Architecture and Code Generation (ACM-TACO)*, 2006.

CONFERENCE PAPERS

3. ICSE'09 (**Alphabetic order**) Yoann Padioleau, **Lin Tan** and Yuanyuan Zhou. "Listening to Programmers - Taxonomies and Characteristics of Comments in Operating System Code." *To appear in the proceedings of the 31st International Conference on Software Engineering*, May 2009. Acceptance Rate: 12.3% (50/405).

4. USENIXSec'08 **Lin Tan**, Xiaolan (Catherine) Zhang, Xiao Ma, Weiwei Xiong and Yuanyuan Zhou. "AutoISES: Automatically Inferring Security Specifications and Detecting Violations." *In the Proceedings of the 17th USENIX Security Symposium*, July-August, 2008. Acceptance Rate: 16% (27/170).

5. SOSP'07 **Lin Tan**, Ding Yuan, Gopal Krishna and Yuanyuan Zhou. "/* iComment: Bugs or Bad Comments? */" *In the Proceedings of the 21st ACM Symposium on Operating Systems Principles*, October 2007. Acceptance Rate: 19% (25/131).

PUBLICATIONS (continued)

6. DASC'07 **Lin Tan**, Ellick M. Chan, Reza Farivar, Nevedita Mallick, Jeffrey C. Carlyle, Francis M. David and Roy H. Campbell. "iKernel: Isolating Buggy and Malicious Device Drivers Using Hardware Virtualization Support." *In the Proceedings of the 3rd IEEE International Symposium on Dependable, Autonomic and Secure Computing*, September 2007.

7. ISCA'05 **Lin Tan** and Timothy Sherwood. "A High Throughput String Matching Architecture for Intrusion Detection and Prevention." *In the proceedings of the 32nd (100,000two-th) Annual International Symposium on Computer Architecture*, June 2005. Acceptance Rate: 23% (45/194). *Cited by 62 according to Google Scholar*.

8. SOSP'05 Qingbo Zhu, Zhifeng Chen, **Lin Tan**, Yuanyuan Zhou, Kimberly Keeton and John Wilkes. "Hibernator: Helping Disk Arrays Sleep Through the Winter." *In the proceedings of the 20th ACM Symposium on Operating Systems Principles*, October 2005. Acceptance Rate: 13% (20/155). *Cited by 70 according to Google Scholar*.

WORKSHOP PAPERS

9. HotOS'07 **Lin Tan**, Ding Yuan and Yuanyuan Zhou. "HotComments: How to Make Program Comments More Useful?" *In the Proceedings of the 11th Workshop on Hot Topics in Operating Systems*, May 2007. Acceptance Rate: 20% (21/105).

10. ASID'06 (**First 2 authors in alphabetic order**) Zhenmin Li, **Lin Tan**, Xuanhui Wang, Shan Lu, Yuanyuan Zhou and Chengxiang Zhai, "Have Things Changed Now? – An Empirical Study of Bug Characteristics in Modern Open Source Software", *In the proceedings of the 1st Workshop on Architectural and System Support for Improving Software Dependability* held together with *ASPLOS*, October 2006. *Cited by 11 according to Google Scholar*.

11. Bugs'05 Shan Lu, Zhenmin Li, Feng Qin, **Lin Tan**, Pin Zhou and Yuanyuan Zhou, "BugBench: A Benchmark for Evaluating Bug Detection Tools", *In the Workshop on the Evaluation of Software Defect Detection Tools*, June 2005. *Cited by 23 according to Google Scholar*.

12. WARFP'05 Timothy Sherwood, Ryan Kastner, Yan Meng, **Lin Tan** and Shreyas Prasad. "Supporting Interdisciplinary Domain Specific Architecture Research with Reconfigurable Devices." *In the Workshop on Architecture Research using FPGA Platforms*, February 2005.

PATENTS AND PENDING PATENTS

- Lin Tan, Xiaolan (Catherine) Zhang, and Yuanyuan Zhou. "Extraction of Code Level Security Specification."
- Madanlal Musuvathi and Lin Tan. "Automatically Generating Test Cases for Binary Code."
- Timothy Sherwood and Lin Tan. "Pattern Matching Technique for High Throughput Network Processing." U.S. Patent Application Ser. No. 60/779,734, filed on March 7, 2006. (**Already licensed**)

RESEARCH EXPERIENCE

University of Illinois at Urbana-Champaign, Research Assistant, January 2005 – Present

- *Leveraging code comments for improving software reliability*: I proposed and designed the first work to use code comments for improving software reliability. This is also one of the first studies to apply *natural language processing* techniques to improve software reliability and to systems research.
  My work includes: (1) automatically extracting specifications from comments written in natural language to detect 60 new bugs and bad/wrong comments in 4 large software projects (Linux, Mozilla, Apache and Wine), with many of these bugs and bad comments already confirmed and fixed by developers; (2) calling for research effort to alleviate the damage caused by bad comments, which have already caused developers to introduce new bugs; and (3) conducting a comprehensive comment characteristics study by examining 2100 comments randomly sampled from 6 pieces of large and popular software (Linux, FreeBSD, OpenSolaris, MySQL, Apache and Eclipse).

RESEARCH EXPERIENCE (continued)

- *Understanding software bugs*: More than 20 research groups used the bugs and statistics from my studies. My work includes: (1) analyzing bug characteristics of two large open source software with up to 4 million lines of code and approximately 90 releases; (2) applying machine learning techniques to automatically classify around 29,000 software bugs; and (3) helping my colleagues build the BugBench benchmark from real world bugs.

IBM Research Watson Center, Research Intern, June 2007 – August 2007

- Combined different techniques including *data mining* with static analysis to automatically extract code-level security specifications.
- Detected exploitable security violations to code-level specification in the Linux kernel and Xen.
- Explored the space of different granularity for security specification extraction.

Microsoft Research Redmond, Research Intern, May 2006 – August 2006

- Designed an automatic and systematic test case generator, called Probit.
- Implemented and Evaluated Probit on a set of programs, including the SPEC INT 2000 benchmark and the byte extract modules in SNORT.
- Demonstrated that Probit has a higher code coverage than random testing.

University of California at Santa Barbara, Research Assistant, March 2004 – December 2004

- Designed a new bit-split multi-pattern string matching algorithm and its extension that can process multiple bytes at a time.
- Showed both the theoretical and practical optimal split parameters for the algorithm.
- *Proved* the upper bound on the total number of states in both single-byte and multi-byte bit-split state machines.
- Proposed a novel configurable string matching architecture for more efficient intrusion detection.
- Demonstrated the potential of building a string matching system that is 10 times more efficient than the currently best known approaches by combing the algorithm and the architecture.

TEACHING EXPERIENCE

Guest Lecturer for CS 527 "Advanced Topics in Software Engineering.", October 2008

- Department of Computer Science, University of Illinois at Urbana-Champaign

Teaching assistant for CS 433 "Computer System Organization.", Fall 2005 & Spring 2005

- Department of Computer Science, University of Illinois at Urbana-Champaign
- Held office hours, designed and graded projects, homework and exams.

Teaching assistant for CS 154 "Computer Architecture.", Spring 2004

- Department of Computer Science, University of California at Santa Barbara
- Gave discussion lectures every week, held office hours, graded projects, homework and exams.

Teaching assistant for CS 40 "Foundation of Computer Science.", Winter 2004

- Department of Computer Science, University of California at Santa Barbara
- Gave discussion lectures every week, held office hours, designed and graded homework and exams.

Teaching assistant for CS 60 "Introduction to C, C++, and UNIX.", Fall 2003

- Department of Computer Science, University of California at Santa Barbara
- Gave two discussion lectures every week, held office hours, graded projects, homework and exams.

CONFERENCE AND WORKSHOP TALKS

- "Listening to Programmers - Taxonomies and Characteristics of Comments in Operating System Code." ICSE'09. Vancouver, Canada. May 2009 (Forthcoming).
- "AutoISES: Automatically Inferring Security Specifications and Detecting Violations." USENIX Security'08. San Jose, CA. August 2008.
- "/* iComment: Bugs or Bad Comments? */" SOSP'07, Stevenson, WA. October, 2007.
- "HotComments: How to Make Program Comments More Useful?" HotOS'07. San Diego, CA. May 2007.
- "Have Things Changed Now? – An Empirical Study of Bug Characteristics in Modern Open Source Software." ASID'06. San Jose, CA. October 2006.
- "A High Throughput String Matching Architecture for Intrusion Detection and Prevention." ISCA'05. Madison, WI. June 2005.

INVITED TALKS

- "Making the Best of an Internship in Systems." Diversity '08 collocated with OSDI'08, San Diego, CA. December, 2008.
- "/* iComment: Bugs or Bad Comments? */" Cisco Systems, Inc., Milpitas, CA. August, 2008.
- "AutoISES: Automatically Inferring Security Specifications and Detecting Violations." ITI Trust & Security Seminar, University of Illinois at Urbana Champaign, IL. July, 2008.
- "/* iComment: Bugs or Bad Comments? */" DAIS seminar, University of Illinois at Urbana Champaign, IL. October, 2007.
- "/* iComment: Bugs or Bad Comments? */" Systems seminar, University of Illinois at Urbana Champaign, IL. September, 2007.

PROFESSIONAL SERVICES

- *Invited speaker* for the workshop on Supporting Diversity in Systems Research (Diversity'08) collocated with OSDI'08.
- External reviewer for SOSP, ASPLOS, PLDI, HPCA, USENIX ATC, NSDI, PACT, SIGMETRICS, HotOS, HotDep, DSN, Computer Networks Journal, and IEEE Journal JSAC.

REFERENCES

Available upon request.