

# Xu Zhao

372 D.L.Pratt Building  
University of Toronto  
Toronto, Ontario, Canada  
☎ +1 (647) 879 3928  
✉ [i@xuzhao.net](mailto:i@xuzhao.net)  
🌐 [www.xuzhao.net](http://www.xuzhao.net)

---

## Research Interests

Distributed systems, failure diagnosis, and automated software logging

My research focuses on building tools to automate the diagnosis of complex distributed software system failures.

---

## Education

Summer 2019 **Ph.D., Electrical and Computer Engineering**

(Expected) *University of Toronto, Toronto, ON*  
Advisor: Prof. Ding Yuan

2015 **M.A.Sc in Electrical and Computer Engineering**

*University of Toronto, Toronto, ON*  
Advisor: Prof. Ding Yuan

2013 **B.E. in Computer Science**

*Tsinghua University, Beijing, China*

---

## Publications

- [1] Xu Zhao, Kirk Rodrigues, Yu Luo, Michael Stumm, Ding Yuan, and Yuanyuan Zhou. Log20: Fully Automated Optimal Placement of Log Printing Statements under Specified Overhead Threshold. In *Proceedings of the 26th ACM Symposium on Operating Systems Principles (SOSP'17)*, Shanghai, China, October 2017.
- [2] Xu Zhao, Kirk Rodrigues, Yu Luo, Michael Stumm, Ding Yuan, and Yuanyuan Zhou. The Game of Twenty Questions: Do You Know Where to Log? In *Proceedings of the 16th Workshop on Hot Topics in Operating Systems (HotOS'17)*, Whistler, BC, Canada, May 2017.
- [3] Xu Zhao, Kirk Rodrigues, Yu Luo, Ding Yuan, and Michael Stumm. Non-intrusive Performance Profiling of Entire Software Stacks based on the Flow Reconstruction Principle. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)*, Savannah, GA, November 2016.
- [4] Xu Zhao\*, Yongle Zhang\*, David Lion, Muhammad FaizanUllah, Yu Luo, Ding Yuan, and Michael Stumm. Lprof: A Non-intrusive Request Flow Profiler for Distributed Systems. In *Proceedings of the 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI'14)*, Broomfield, CO, October 2014.  
\*: Joint first authors.
- [5] Ding Yuan, Yu Luo, Xin Zhuang, Guilherme Rodrigues, Xu Zhao, Yongle Zhang, Pranay U. Jain, and Michael Stumm. Simple Testing Can Prevent Most Critical Failures: An Analysis of Production Failures in Distributed Data-intensive Systems. In *Proceedings of the 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI'14)*, Broomfield, CO, October 2014.

- [6] Tian Xiao, Zhenyu Guo, Hucheng Zhou, Jiaying Zhang, Xu Zhao, Chencheng Ye, Xi Wang, Wei Lin, Wenguang Chen, Lidong Zhou. Cybertron: Pushing the limit on I/O reduction in data-parallel programs. In *Proceedings of the 2014 ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA'14)*, Portland, OR, October 2014.

---

## Research Experience

2013 – current **Graduate Student, University of Toronto.**

- *Log20*, a tool that determines an optimal placement of logging statements under a specified amount of performance overhead. Guided by information theory, *Log20* proposes a metric, *entropy*, to measure how effective each placement of log printing statement is in disambiguating program execution paths. *Log20* provides a curve showing the trade-off between the informativeness of the logs and the performance slowdown, allowing developers to choose the desired balance.  
**Impact:** Developers of AngularJS, the Opera web browser, and Amplifinity Inc. have started projects to use *Log20* as their logging system. *Log20* is licensed by a world-class IT company.
- *Stitch*, a log-based flow reconstruction tool for distributed software stacks. *Stitch* tackles the problem of distributed software failure diagnosis from a complete different perspective by discovering the human intuition in debugging these systems. *Stitch* relies on the *Flow Reconstruction Principle*, which states that developers will output sufficient information in logs so as to reconstruct runtime execution flows posteriori.
- *lprof*, a non-intrusive request flow profiler for distributed systems. *lprof* uses static analysis technique to reconstruct the execution flow in distributed systems. It builds the system execution model by recognizing log-printing statements that belong to the same the execution and separating different executions by the identifier values that remain invariant within the same request.  
**Impact:** *lprof* is now used by developers in a world-class IT company to monitor a production order processing system that is used by 80 million users and processes up to billions of transactions a day.
- A detailed study on 198 real world failures from five distributed systems. The study shows 92% of the catastrophic failures on production systems are the result of incorrect handling of errors explicitly signalled in software. Furthermore, 33% of these catastrophic failures could have been detected by simple static checks on empty or dummy error handlers. We further provide a tool, named Aspirator, that is capable of locating these simple bugs. It has detected over 200 bugs in Hadoop and Spark that have been fixed after reporting them. Aspirator is used in Google's error-prone bug checker and HBase.

2012 – 2013 **Research Intern, Microsoft Research Asia.**

*Cybertron*, an I/O cost reduction plugin for data-parallel programs.

*Cybertron* analyzes byte-level data dependencies on different stages of a data-parallel program and reduces the program I/O cost by filtering never-used data in early stage.

2011 – 2012 **Research Assistant, Tsinghua University.**

Implemented a verification tool for a deterministic thread scheduler in QEMU emulator.

---

## Press

- 2017 **Log20: Fully Automated Optimal Placement of Log Printing Statements under Specified Overhead Threshold.**  
Featured by *the morning paper* and *SRE weekly*.

---

## Awards

- 2018 **Facebook Fellowship**, *Facebook Inc.*  
*One of the 17 winners out of over 800 applicants.*
- 2017 **ECE Fellowship**, *University of Toronto.*
- 2016 **Edward Rogers Graduate Scholarship**, *University of Toronto.*
- 2013 **Computer Science Department Outstanding Graduates**, *Tsinghua University.*

---

## Recent Teaching Experience

- Fall 2018 **Course Instructor**, *CSC467: Compilers and Interpreters.*  
Served as the sole instructor for the course. Presented 3 lectures weekly, redesigned the course website, lecture materials, and exams from the ground up.
- Winter 2018 **Teaching Assistant**, *ECE344: Operating Systems.*  
Held lab sessions, handled questions on Piazza, and graded the midterm and final exam.
- Fall 2017 **Teaching Assistant**, *ECE454: Computer Systems Programming.*  
Redesigned lab materials, provided in-lab help, and graded exams and the hardest lab of the course (accounts for 9% of the total mark).
- Winter 2017 **Teaching Assistant**, *ECE344: Operating Systems.*  
Held lab sessions, handled questions on Piazza, and graded the midterm and final exam.

---

## Presentations

- Nov. 2017 **Log20: Fully Automated Optimal Placement of Log Printing Statements under Specified Overhead Threshold**, *Department of Computer Science, Tsinghua University, Beijing, China.*
- Oct. 2017 **Log20: Fully Automated Optimal Placement of Log Printing Statements under Specified Overhead Threshold**, *SOSP'17, Shanghai, China.*
- May. 2017 **The Game of Twenty Questions: Do You Know Where to Log?**, *HotOS'17, Whistler, BC, Canada.*
- Nov. 2016 **Non-intrusive Performance Profiling of Entire Software Stacks based on the Flow Reconstruction Principle**, *OSDI'16, Savannah, GA.*

---

## Industry Experience

- 2014 **Intern**, *Twitter Inc.*
- 2013 **Research intern**, *Microsoft Research Asia.*