

# David Lion

---

University of Toronto  
6 King's College Road, Room 372  
Toronto, Ontario · M5S 3H5 · Canada

Email: david.lion@mail.utoronto.ca  
URL: <http://dsrg.utoronto.ca/people/david>  
Phone: 416 725 8625

## Education

- |  |                       |
|--|-----------------------|
| <b>Ph.D. in Computer Engineering</b><br>University of Toronto<br><i>Advisor:</i> Prof. Ding Yuan   | Sept 2016 - Present   |
| <b>M.A.Sc. in Computer Engineering</b><br>University of Toronto<br><i>Advisor:</i> Prof. Ding Yuan | Sept 2014 - June 2017 |
| <b>B.A.Sc. in Computer Engineering</b><br>University of Toronto                                    | Sept 2009 - June 2014 |

## Research Interests

The design and implementation of software systems with an emphasis on performance issues and failure debugging of distributed systems.

## Research Projects

### **M3: End-to-End Memory Management in Elastic System Software Stacks** Summer 2016 - Fall 2020

- Designed a system that bridges memory abstractions between all layers in a system software stack, allowing applications to continuously adapt to current system memory availability.
- The system improves performance and maximizes memory utilization by removing static memory settings that are fundamentally incapable of reacting to workload or memory usage changes.
- The system uses a global monitor written in C++ and was implemented and evaluated on Spark, the JVM, a Go caching application, and the Go runtime using Java, C++, and Go.
  - Open sourced: <https://github.com/dsrg-uoft>.

### **Eliminate JVM Warm-up Overhead in Data-Parallel Systems** Summer 2015 - Fall 2016

- Studied the effects and extent of JVM Warm-up overhead on popular data-parallel distributed systems, such as HDFS, Hive on Tez, and Spark through comprehensive instrumentation written in Java and C++.
- Found that warm-up overhead is frequently the bottleneck, even in I/O intensive work, that it reveals a contradiction between the principle of parallelization and the JVM, and that multi-layer systems aggravate the problem.
- Designed HotTub, a new JVM which amortizes the warm-up overhead over the lifetime of a node instead of a single job, by reusing a pool of already warm JVMs.
  - Written using C++, C, Java, and x86 assembly. Open sourced: <https://github.com/dsrg-uoft/hottub>.

### **Automated Reproduction of Distributed System Failures** Summer 2014 - Summer 2015

- Generates a unit test which will result in the same error log seen in production by the developer, providing a reproduction to aid debugging and testing.
- Static analysis, written in C++ and Java, uses production logs to find past events relating to the chosen error, building constraints required to trigger the error.
- SAT solver is used to generate inputs to the observed events to create a unit test of the error.

- Provides non-intrusive profiling of distributed systems by using existing developer logs aiding debugging and understanding with no penalty.
- Static analysis of JVM bytecode, written in Java, is used to generate a model of requests by inferring related log messages to understand the behaviour of a request.
- Dynamic analysis, written in Java, uses the model to gather request information from production logs and store it into a database which users can query.

## Publications

1. David Lion, Adrian Chiu, and Ding Yuan. M3: End-to-End Memory Management in Elastic System Software Stacks. In *Proceedings of the Sixteenth European Conference on Computer Systems*, EuroSys '21, pages 507–522. ACM, 2021
2. Yongle Zhang, Serguei Makarov, Xiang Ren, David Lion, and Ding Yuan. Pensieve: Non-Intrusive Failure Reproduction for Distributed Systems Using the Event Chaining Approach. In *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, pages 19–33. ACM, 2017
3. Naif Tarafdar, Thomas Lin, Nariman Eskandari, David Lion, Alberto Leon-Garcia, and Paul Chow. Heterogeneous Virtualized Network Function Framework for the Data Center. In *27th International Conference on Field Programmable Logic and Applications*, FPL '17, pages 1–8. IEEE, 2017
4. David Lion, Adrian Chiu, Hailong Sun, Xin Zhuang, Nikola Grcevski, and Ding Yuan. Don't Get Caught in the Cold, Warm-up Your JVM: Understand and Eliminate JVM Warm-up Overhead in Data-Parallel Systems. In *12th USENIX Symposium on Operating Systems Design and Implementation*, OSDI '16, pages 383–400. USENIX Association, 2016
5. Xu Zhao, Yongle Zhang, David Lion, Muhammad Faizan Ullah, Yu Luo, Ding Yuan, and Michael Stumm. lprof: A Non-intrusive Request Flow Profiler for Distributed Systems. In *11th USENIX Symposium on Operating Systems Design and Implementation*, OSDI '14, pages 629–644. USENIX Association, 2014

## Teaching Experience

### Course Instructor, UofT CSC369, Operating Systems

- Summer 2021

### Teaching Assistant, UofT ECE344, Operating Systems

- Every semester from Spring 2015 to Spring 2021 (12 semesters in total)

### Teaching Assistant, UofT ECE454, Computer Systems Programming

- Fall 2014

## Course Projects

### Hardening F2FS

Fall 2016

- Modified F2FS file system in the Linux kernel, written in C, to be more robust through metadata checksums and replication.

### C to MPI-C Source-to-Source Translation Through ROSE

Winter 2015

- Modified ROSE Compiler, written in C and C++, to automatically parallelize C source code through pragmas into Message Passing Interface (MPI) C source code.

### Virtual Machine Public and Private Cloud Migration

Fall 2013 - Winter 2014

- Used multiple hypervisor layers and a scheduler, written in Java, to implement automatic and abstracted virtual machine migration between public and private clouds.

## Professional Experiences

### Software Engineer, Apple Platform Advanced Micro Devices (AMD)

May 2012 - September 2013

- Placed second out of 120 interns for the Intern of the Year Award.
- Added features to and maintained kernel-side device drivers, OpenCL drivers and tools for AMD graphics hardware on the Apple OSX platform, written in C++ and C.
- Migrated the OpenCL driver to a future framework, maintaining correctness of the features.
- Collaborated directly with Apple employees to fix bugs not isolated in the AMD driver.
- Analyzed core dumps and GPU hang logs to determine the cause of panics, crashes, and hangs.

### Web Developer, Gamebot

Summer 2011

- Worked with C#, ASP.NET MVC (model-view-controller) framework, Powershell, SVN.

## Technical Expertise

**Languages:** C++ (*5 years*), C (*4 years*), Java (*4 years*), Bash (*4 years*), Python (*2 years*), Perl (*<1 year*), x86 Assembly (*<1 year*)

**Version Control:** Git (*4 years*), Subversion(SVN) (*2 years*)