

Yan Li

University of California, Santa Cruz
Department of Computer Science
Storage Systems Research Center
Santa Cruz, CA 95064

Phone: (831) 713-6616
Email: yanli@ucsc.edu
Homepage: <http://elliott.li>

RESEARCH INTERESTS

Yan Li is a Ph.D. student working with Professor Darrell D. E. Long. He focuses on computer storage performance tuning, especially using machine learning and heuristics. He has an Erdős number 3.

RESEARCH PROJECTS

- Jun '13 - now ASCAR: Automatic performance tuning
He has designed the ASCAR system, which can discover optimal values for configuration parameters for complex storage systems through automated trial runs. ASCAR adopts many heuristics, optimization, and statistics methods to make it possible to evaluate a large number of candidate values in a short time. Source code and more project information can be found at <http://ascar.io>
- Oct '15 - now Pilot: Automatic performance measurement
Carrying out even the simplest form of performance benchmark requires considerable knowledge of statistics and computer systems, and painstakingly following many error-prone steps in order to meet all the requirements in accuracy, precision, comparability, repeatability, and control of overhead. We propose a collection of algorithms and heuristics to automate the steps that are necessary to produce statistically sound performance measurement results. They cover the collection, storing, analyses, and comparing of performance measurements and can help to reduce human effort and error. We implement these methods as a readily-usable software framework, Pilot. Please let us know if you'd like to try it before we open source it.
- Jun '14 - now Automation for data center and other complex systems
Based on the observation that most performance issues in the real world are caused by misconfiguration, this study explores ideas to automate the troubleshooting and tuning of complex systems by applying artificial intelligence and machine learning methods. The goal is to reduce human's involvement and reduce costs.
- Jun '15 - Sep '15 Designing and prove complex parallel storage systems
Complex and parallel systems are hard to design, verify, and debug – some rare race conditions and data corruption bugs may exist for years after the systems went into production before being discovered – yet they must undergo rigorous examination and test before people can trust their data to them. For this project he uses +Cal/TLA+ for formally proving the correctness of complex object storage system that involves complex lockless data structures and pipelines.

- Nov '12 - Sep '14 Energy-aware storage for HPC
In the following decades, energy will be a gating issue in large scale storage systems, from high-performance computing (HPC) to large data centers. In this research, He works on (1) developing profiling and simulation tools for optimizing large distributed programs, (2) energy-efficient scheduling and allocation algorithms for future storage systems.
- 2011-2012 Horus, Fine-Grained Encryption-Based Security for Large-Scale Storage
Horus encrypts large datasets using keyed hash trees to generate different keys for each region of the dataset, providing fine-grained security with very low overhead. In this research, he developed the Horus prototype (with others) and wrote the paper that has been accepted to FAST'13.
- 2011 Understanding Data Survivability in Archival Storage Systems
This research pointed out the importance of understanding data survivability in archival storage systems and developed a method to calculate it for data stored on heterogeneous storage devices, taking rare but disastrous events into account.

EMPLOYMENT HISTORY

- Summer '15 Research Intern, TurboStor
He worked on designing, developing, and performance tuning future generation high-performance storage systems. Complex and highly parallel storage systems are hard to design, verify, and debug – some rare race conditions and data corruption bugs may exist for years after the systems went into production before being discovered. For this project, he developed a complete model based on +Cal/TLA+ and formally proved the correctness of a complex object storage system that involves complex lockless data structures and pipelines.
- Winter '13 - now Administrator, ACM Transactions on Storage
ACM Transactions on Storage (<http://tos.acm.org/>) is the preeminent journal for computer storage research.
- Summer '14 Research Intern, Symantec Research Labs
System management and optimization automation, reducing complexity and costs.
- Summer '13 Research Intern, IBM Almaden Research Center
He worked on the Quality of Service (QoS) of GPFS for HPC applications.
- Summer '12 Engineering Intern, Google
He worked in the Linux kernel performance team and developed a new scheduler that improved the server response latency.

- 2012 Teaching Assistant of CMPS 109 Advanced Programming Language with Professor Ira Pohl. This course is an upper-division computer science course on C++ OOP and template.
- 2008-2011 Senior Software Engineer, MeeGo core system developer, Intel
MeeGo was hitherto the next-generation research operating system sponsored by Intel and Linux Foundation for emerging mobile platforms such as handset, tablet and in-vehicle infotainment systems. His worked covered:
- Linux kernel maintainer: maintained Intel's patchsets on the latest Linux kernel, oversaw the kernel release process, be responsible for the overall kernel quality.
 - Ported Linux kernels and Xorg to latest Intel hardware.
 - Developed the secure file-system for MeeGo (eCryptfs based).
 - Developed persistent on-disk caching of KBD compilation of Xorg, increased the Xorg start-up speed by 1 to 2 seconds on all platforms.
 - Designed the process for ensuring legal compliance of using GPLv3 licensed software in MeeGo.
 - Developed infrastructure and components (VPN, Microsoft Exchange email client, WPA2 EAP, etc.) for using MeeGo in enterprise environment.
- 2010 Co-admin and mentor of Google Summer of Code 2010 for Intel's MeeGo Platform
Co-administered the MeeGo/Maemo project for Google Summer of Code 2010. Mentored one student who worked on cloud storage support for MeeGo.
- 2009-2012 GNOME Foundation Member
GNOME Foundation (<http://www.gnome.org/about/>) is a non-profit desktop software development organization based in Cambridge, Massachusetts, United States, coordinating the efforts in the GNOME project, which aims at developing a free and open source graphics user interface for desktop computers. He worked as a contributor to Evolution, the personal information management application.
- 2006-2008 Staff Software Engineer, IBM
QA Project Manager of the Asianux on POWER team. He led a 4-person QA team for ensuring the quality of Asianux, which is a Chinese local Linux distro for enterprise market running on IBM POWER platform. His worked covered overseeing the designing and running of QA tasks, as well as technical support of local market.

HONORS AND AWARDS

- 2014, Symantec Research Labs Graduate Fellowship
- 2012, UCSC Graduate Research Symposium Alumni Association Award.
- 2009, Special Recognition Award, Intel.
- 2008, First Patent Plateau Award of IBM for filing 4 patent applications.
- 2007, IBM First Patent Award.
- 2004, China National Third Prize for the Electric Grid Line Loss Online Management System, by Chinese Society for Electrical Engineering. Yan Li was the software architect and one of the main developers.

- 2003, National Fellowship for Graduate Study during Master program.
- 2002, Shandong Province Undergraduate Electronic Design Contest, China, First Prize.
- 2001, China National Undergraduate Electronic Design Contest, First Prize.
- 1997, China National Olympiad in Informatics, Shandong District, Gold Medal.
- 1995, Shandong Olympiad in Informatics, China, First Prize.

PUBLICATIONS

Refereed Conference Papers

1. **Yan Li**, Yash Gupta, Ethan L. Miller, Darrell D. E. Long. "Pilot: A Framework that Understands How to Do Performance Benchmarks The Right Way." *IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2016)*, London, UK, September 19–21, 2016.
2. **Yan Li**, Xiaoyuan Lu, Ethan L. Miller, Darrell D. E. Long. "ASCAR: Automating Contention Management for High-Performance Storage Systems." *31st International Conference on Massive Storage Systems and Technologies (MSST 2015)*, Santa Clara, CA, USA: May 30–June 5, 2015.
3. **Yan Li**, Darrell D. E. Long. "Which Storage Device is the Greenest? Modeling the Energy Cost of I/O Workloads." *IEEE 22nd International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2014)*, Paris, France, September 9–11, 2014.
4. **Yan Li**, Nakul Sanjay Dhotre, Yasuhiro Ohara, Thomas M. Kroeger, Ethan L. Miller, Darrell D. E. Long. "Horus: Fine-Grained Encryption-Based Security for Large-Scale Storage," *Proceedings of the 11th USENIX Conference on File and Storage Technologies (FAST '13)*, San Jose, CA, USA: Usenix Association, February 12–15, 2013.
5. Yulai Xie, Kiran-Kumar Muniswamy-Reddy, Dan Feng, **Yan Li**, Darrell D. E. Long, Zhipeng Tan, Lei Chen. "A Hybrid Approach for Efficient Provenance Storage," *The 21st ACM Conference on Information and Knowledge Management (CIKM '12)*, Maui, HI, USA: ACM, October 29–November 2, 2012.
6. **Yan Li**, Ethan L. Miller, Darrell D. E. Long. "Understanding Data Survivability in Archival Storage Systems," *The 5th Annual International Systems and Storage Conference (SYSTOR 2012)*, Haifa, Israel: ACM, June 4–6, 2012.

Invited Lectures

1. "Horus: Fine-Grained Encryption-Based Security for Large-Scale Storage," Sandia National Laboratories, 2012.

Industrial Conference Talks

1. **Yan Li**. "ASCAR: Increasing Performance Through Automated Contention Management," (video: <https://goo.gl/1RKcpn>, slides: <http://goo.gl/8BbjCa>), *Lustre User Group 2016*, Portland, Oregon, U.S.A., April, 2016.
2. **Yan Li**. "MeeGo Enterprise Goes Beyond Desktop," (<http://sf2011.mee-go.com/program/sessions/mee-go-enterprise-goes-beyond-desktop>), *MeeGo Conference 2011*, San Francisco, U.S.A., May, 2011.

3. **Yan Li**. "MeeGo and GPLv3," (<http://sf2011.meego.com/program/sessions/meego-and-gplv3>), *MeeGo Conference 2011*, San Francisco, U.S.A., May, 2011.
4. **Yan Li**. "Using MeeGo as Enterprise Desktop," (<http://conference2010.meego.com/session/using-meego-enterprise-desktop>). *MeeGo Conference 2010*, Dublin, Ireland, November, 2010.

Patents

1. Yao Qi, **Yan Li**, Wei Ying Yu, Yong Zheng. 2007. Providing customizable, process-specific just-in-time debugging in an operating system. US9128837B2.
2. Yao Qi, **Yan Li**, Wei Ying Yu, Yong Zheng. 2007. Method and System for Debugging a Program in a Multi-thread Environment. US8201152B2.

Journals

1. Yulai Xie, Kiran-Kumar Muniswamy-Reddy, Dan Feng, **Yan Li**, Darrell D. E. Long. Evaluation of A Hybrid Approach for Efficient Provenance Storage. *ACM Transactions on Storage*. 9(4). Nov. 2013. <http://dx.doi.org/10.1145/2501986>
2. **Yan Li**. 2005. Use WANT to Build Delphi Project. *Programmer Magazine*, Feb 2005.
3. Jin Li, **Yan Li**. 2002. Information Collect System in Power Department. *Journal of Qingdao Institute of Chemical Technology* Vol. 23: 65-67.

Book

1. **Yan Li**, Jingjing Liu, and Yiming Yu. 2006. *J2ME Development and Application*. China Machine Press, Beijing, China. ISBN: 7111188349

Technical Report

1. **Yan Li**. 2006. Research on Refactorable Software Platform: A Framework Reuse Solution. Qingdao: Ocean University of China.

EDUCATION

2011-	Ph.D. student	University of California, Santa Cruz, Computer Science
2006	M.E.	Ocean University of China, Computer Applied Technology
2003	B.E.	Ocean University of China, Electronic Engineering

Last updated: June 16, 2016

<http://elliott.li/yanli-cv.pdf>