

Chao Liu

CONTACT INFORMATION

Siebel Center for Computer Science, Room 2119B *Work Phone:* (217) 244-3570 *Cell Phone:* (217) 721-7884
University of Illinois at Urbana-Champaign *E-mail:* chaoliu@cs.uiuc.edu
201 N. Goodwin, Urbana, IL 61801 USA *Web:* <http://www.ews.uiuc.edu/~chaoliu/index.htm>

RESEARCH INTERESTS

Inter-disciplinary research between **data mining**, **software engineering** and **computer systems**.

- In particular, I have been working on automated fault localization and program failure analysis, which demonstrate the power of data mining and statistical analysis in improving software reliability.
- In general, I believe that system reliability, security and performance can be leveraged through data mining and statistical analysis of system trace data.

EDUCATION

12/05 - 08/07 Ph.D. Department of Computer Science, University of Illinois at Urbana-Champaign
(expected) Advisor: Jiawei Han.
Expected to graduate in 2007. GPA: 3.97/4.00.

08/03 - 12/05 M.S. Department of Computer Science, University of Illinois at Urbana-Champaign
Thesis: "SOBER: Statistical Model-based Bug Localization."
Advisor: Jiawei Han.
GPA: 3.97/4.00

09/99 - 07/03 B.S. Department of Computer Science, Peking University, Beijing, China
Major GPA: 3.84/4.00

RESEARCH EXPERIENCE

Department of Computer Science, University of Illinois at Urbana-Champaign. **Aug. 2003 - Present**

- Research assistant working on the project "Production-Run On-the-fly Bug Elimination", which is supported by NSF-ITR (Funding #: CCR-0325603).
- Published papers: IEEE-TSE, ICDM'06, FSE'06, SDM'06, FSE'05, SDM'05.

IBM Almaden Research Center, San Jose, CA, US **May. 2005 - Aug. 2005**

- Summer intern working on the design and implementation of a next-generation Content Management System.
- Co-devised a scan-free archiving algorithm that improves efficiency by orders of magnitude (patent pending).

Security Research Group, National Center for Supercomputing Applications **Jun. 2004 - Aug. 2004**

- Summer intern working on cluster security-related issues, published a paper at CCGRID'05.
- Developed an SVM-based algorithm to detect identity masqueraders using process accounting data.

Microsoft Research Asia, Beijing, China **Feb. 2003 - May 2003**

- Designed and implemented an efficient neural network for handwriting recognition.
- The implementation was integrated into to the *Digital Ink* project for Microsoft Windows XP on Tablet PC.

JOURNAL PUBLICATIONS

- IEEE-TSE **Chao Liu**, Long Fei, Xifeng Yan, Jiawei Han and Samuel Midkiff, “Statistical Debugging: A Hypothesis Testing-based Approach,” *IEEE Transaction on Software Engineering*, Vol. 32, No. 10, pp. 831-848, Oct., 2006.
- IEEE-TSE **Chao Liu** and Jiawei Han, “R-Proximity: Failure Proximity Defined via Statistical Debugging,” *IEEE Transaction on Software Engineering*, Sept. 2006. (*under review*)

CONFERENCE PUBLICATIONS

- ICDM’06 **Chao Liu**, Zeng Lian and Jiawei Han, “How Bayesians Debug,” in *Proceeding of the 6th IEEE International Conference on Data Mining*, pp. 382-393, Hong Kong, China, Dec., 2006.
FULL PAPER
- FSE’06 **Chao Liu** and Jiawei Han, “Failure Proximity: A Fault Localization-Based Approach,” in *Proceeding of the 14th ACM SIGSOFT Symposium on the Foundations of Software Engineering*, pp. 286-295, Portland, USA, Nov., 2006.
FULL PAPER
- FSE’06 **Chao Liu**, “Fault-aware Fingerprinting: Towards Mutualism between Failure Investigation and Statistical Debugging,” in *Proceeding of the 14th ACM SIGSOFT Symposium on the Foundations of Software Engineering*, Portland, USA, Nov., 2006.
DOCTORAL SYMPOSIUM
- KDD’06 **Chao Liu**, Chen Chen, Jiawei Han and Philip S. Yu, “GPLAG: Detection of Software Plagiarism by Program Dependence Graph Analysis,” in *Proceeding of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 872-881, Philadelphia, USA, Aug., 2006.
FULL PAPER
- WWW’06 Qiaozhu Mei, **Chao Liu**, Hang Su and Chengxiang Zhai, “A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs,” in *Proceeding of the 15th International Conference on World Wide Web*, pp. 533-542, Edinburgh, Scotland, May, 2006.
FULL PAPER
- SDM’06 **Chao Liu**, Xifeng Yan and Jiawei Han, “Mining Control Flow Abnormality for Logic Error Isolation,” in *Proceeding of 2006 SIAM International Conference on Data Mining*, pp. 106-117, Bethesda, US, April, 2006.
FULL PAPER
- ICN’06 Bo Zhao and **Chao Liu** “Efficient SIP-Specific Event Notification,” in *Proceeding of the 5th IEEE International Conference on Networking*, pp. 1-8, Mauritius, April, 2006.
FULL PAPER
- FSE’05 **Chao Liu**, Xifeng Yan, Long Fei, Jiawei Han and Samuel Midkiff, “SOBER: Statistical Model-based Bug Localization,” in *Proceeding of the 5th joint meeting of the European Software Engineering Conference and ACM SIGSOFT Symposium on the Foundations of Software Engineering*, pp. 286-295, Lisbon, Portugal, Sept., 2005.
FULL PAPER
- CCGRID’05 William Yurcik and **Chao Liu**. “A First Step Toward Detecting SSH Identity Theft on HPC Clusters: Discriminating Cluster Masqueraders Based on Command Behavior,” in *Proceeding of the 5th International Symposium on Cluster Computing and the Grid*, pp. 111-120, Cardiff, UK, May, 2005.
FULL PAPER
- SDM’05 **Chao Liu**, Xifeng Yan, Hwanjo Yu, Jiawei Han and Philip S. Yu, “Mining Behavior Graphs for “Backtrace” of Noncrashing Bugs,” in *Proceeding of 2005 SIAM International Conference on Data Mining*, pp. 286-297, Newport Beach, US, April, 2005.
FULL PAPER
- PAKDD’03 **Chao Liu**, Ming Zhang, Minrui Zheng and Yixin Chen. “Step-by-step Regression: A More Efficient Alternative for Polynomial Multiple Linear Regression in Stream Cube,” in *Proceeding of the 7th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’03)*, pp. 437-448, Seoul, Korea, April, 2003.
FULL PAPER

TECHNICAL REPORTS

- TR587, IU Chris Giannella, Jiawei Han, Edward Robertson and **Chao Liu**, “Mining Frequent Itemsets Over Arbitrary Time Intervals in Data Streams,” *Technical Report TR587*, Indiana University, Nov, 2003.

HONORS AND AWARDS

- 2006 **IBM Travel Grant** for Attendance of KDD'06.
- 2005 **ACM SIGSOFT Travel Grant** for Attendance of ESEC/FSE'05.
- 2005 **IBM Travel Grant** for Attendance of SDM'05.
- 2003 - 2006 **ILLIAC Fellowship**, Department of Computer Science, UIUC.
- 1999 - 2003 **Mingde Scholarship**, Peking University, China.
- 1999 **First Place** in National College Entrance Examination, Henan Province, China.
- 1998 **First Prize in CMO'98**, Chinese Mathematics Olympiad.
- 1998 **First Prize in CPhO'98**, Chinese Physics Olympiad.

TECHNICAL TALKS

- “Failure Proximity: A Fault Localization-Based Approach,” the 14th ACM SIGSOFT Symposium on the Foundations of Software Engineering, Portland, USA, Nov., 2006.
- “Fault-aware Fingerprinting: Towards Mutualism between Failure Investigation and Statistical Debugging,” the 14th ACM SIGSOFT Symposium on the Foundations of Software Engineering, Portland, USA, Nov., 2006.
- “Statistical Debugging and Automated Program Failure Analysis,” invited talk, the Database and Information System Laboratory, University of Illinois at Urbana-Champaign, Oct., 2006.
- “GPLAG: Detection of Software Plagiarism by Program Dependence Graph Analysis,” the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, USA, Aug., 2006.
- “A Lightweight Survey of Fault Localization Techniques,” invited lecture in CS598DM, University of Illinois at Urbana-Champaign, Nov., 2005.
- “SOBER: Statistical Model-based Fault Localization,” the 5th joint meeting of the European Software Engineering Conference and ACM SIGSOFT Symposium on the Foundations of Software Engineering, Lisbon, Portugal, Sept., 2005.
- “SOBER: Statistical Model-based Fault Localization,” IBM Almaden Research Center, San Jose, CA, July, 2005.
- “Mining Behavior Graphs for ‘Backtrace’ of Noncrashing Bugs,” 2005 SIAM International Conference on Data Mining, Newport Beach, CA, April, 2005.
- “Step-by-step Regression: A More Efficient Alternative for Polynomial Multiple Linear Regression in Stream Cube,” the 7th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'03), Seoul, South Korea, April, 2003.

PROFESSIONAL SERVICES

- Student Panelist in the SIGSOFT 2006 Inspiration Program, which “aims at undergraduate and M.S. students who are seriously contemplating obtaining a Ph.D. and will provide them with a ‘taste of software engineering’.”
- Referee for IEEE Transaction on Knowledge and Data Engineering (IEEE-TKDE), PAKDD'07, ICDM'06, KDD'06, KDD'05, ICDE'05, KDD'04.
- Student Member of the Society for Industrial and Applied Mathematics (SIAM).
- IEEE Member, ACM Member and ACM SIGSOFT Member.
- Volunteer for the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

COMPUTER SKILLS

Statistical Packages: R environments, S-Plus and Matlab Statistics Toolbox.

Scientific Computing: Matlab, Maple and Mathematica.

Programming Languages: C, C++, Perl and csh/tcsh/bash shell script.

Database Systems: Working experience with DB2, Microsoft SQL Server and MySQL.

Development Platforms: Windows, Unix and Linux.

RELEASED SOFTWARE

SOBER: An Automated Debugging Tool via Statistical Analysis. (<http://www.ews.uiuc.edu/~chaoliu/sober.htm>).

PATENTS

Ying Chen, **Chao Liu** and Lesley Mbogo, "Scan-free Archiving," *ARC920060007US1*, IBM Almaden Research Center, US. (pending)

REFERENCES

Prof. Jiawei Han (Thesis Advisor)

Department of Computer Science,
University of Illinois at Urbana-Champaign
Email: hanj@cs.uiuc.edu
WWW: www.cs.uiuc.edu/~hanj
Tel: (217) 333-6903

Prof. Samuel P. Midkiff

School of Electrical and Computer Engineering
Purdue University
Email: smidkiff@purdue.edu
WWW: cobweb.ecn.purdue.edu/~smidkiff
Tel: (765) 494-3440

Dr. Philip S. Yu

Manager, Software Tools and Techniques group
IBM Thomas J. Watson Research Center
Email: psyu@us.ibm.com
WWW: www.research.ibm.com/people/p/psyu/index.html
Tel: (914) 784-7141

Prof. Darko Marinov

Department of Computer Science,
University of Illinois at Urbana-Champaign
Email: marinov@cs.uiuc.edu
WWW: www-faculty.cs.uiuc.edu/~marinov
Tel: (217) 265-6117

Dr. Ying Chen (Internship Mentor)

IBM Almaden Research Center
San Jose, CA
Email: yingchen@us.ibm.com
Tel: (408) 927-1858