

Avani P. Wildani née Gadani

CONTACT INFORMATION

703 Solano Ave
Albany, CA, 94706

Voice: +1 909.437.8626
E-mail: avani@soe.ucsc.edu

Objective: Tenure-track faculty position for 2012/2013

CURRENT RESEARCH

UCSC Storage Systems Research Center

My current research revolves around using machine-learning based prediction models to dynamically re-arrange data between storage devices for applications including power management, availability increase through fault isolation, and data compression. The focus is on exascale and cloud storage systems though we have applications in archival, high performance, and general enterprise storage. A complementary project I am working on involves isolating features from a large set of distinct workload types to do a wide reaching workload characterization study. Much research in systems depends generalized assumptions about workload types that we hope to formalize build from.

I previously worked in power-aware coding to explore the trade-offs between power consumption, performance, and availability when reconstructing data using irregular erasure codes. I also have a side project that is attempting to build a characteristic database of workloads using learned features.

EDUCATION

University of California, Santa Cruz, CA USA

PhD Computer Science (ongoing: started September 2007)

University of New Mexico, Albuquerque, NM USA

MS Computer Science : Machine Learning (May 2007)

Harvey Mudd College, Claremont, CA USA

B.S Joint Math / Computer Science (May 2003)

Relevant Classes: Advanced Math and Computer Science Courses including -

Advanced Algorithms, Scientific Computing, Graph Theory, Computational Geometry, Probability, Differential Geometry, Machine Learning, Artificial Intelligence, Software Engineering, Computational Biology, Programming Languages Theory, Computational Linguistics, Networks, Storage Systems, Adv. Operating Systems, Information Theory, and Databases

PROJECTS

MIND Institute

This research involved using fMRI images to detect interconnections within the brains of both healthy and schizophrenic patients. From this, we worked on deriving an efficient technique to perform clustering and hidden variable **Bayesian Structure Search** analyses. I have also used **Support Vector Machines** to obtain a baseline for our other classification methods, and worked extensively with **Spectral Graph Clustering** methods to handle the patients' categorical data.

SLC Project

Research Assistant in the proposal stage of a large scale learning project involving gene expression data from rodents. I modeled the data using **Bayesian Networks** after doing basic clustering for manual parameter selection.

Sandia National Laboratories Clinic

Program Manager of a team working with Dr. Kevin Boyack at Sandia to cluster discrete data points and produce a visualization tool to help with large data-set analysis. I **implemented density clustering** and also researched validity metrics to integrate with the tool we wrote in MATLAB.

WORK
EXPERIENCE

IBM Almaden Research Center: *Research Intern* **Summer 2011 (continuing)**
Worked with a small team to look at addressing the disk bottleneck problem in data de-duplication for primary storage by using my thesis work in data grouping. Involved analyzing traces of 100GB+ in short amounts of time as well as writing Python libraries to simulate de-duplication. This is continuing work that will be submitted to USENIX2012.

Sandia Labs - Computer Science Research Institute: *Research Intern* **Summer 2010**
Analyzed block I/O data from a multi-use production system in order to define and identify similarity between particular data blocks. Looked into creating access groupings based on the similarity data. Research on similarity was published in SYSTOR 2011.

Google: *Intern* **Summer 2009**
Implemented a PID-based control system in Python to automatically balance customers for Bigtable (a key/value store). Extensively researched reward functions and SLA/SLO management techniques and wrote out a detailed experiment plan to investigate the tradeoffs between economic and standard scheduling techniques. Worked with the Census team to expand their SLO framework to include SLAs.

IBM Almaden Research Center: *Research Intern* **Summer 2008 (continued through fall)**
Worked with a team of 4 researchers to develop a highly accurate method to automatically classify log excerpts from commercially deployed storage systems to speed up problem isolation. This involved coding in Java, particularly the Lucene search engine, and Python for the backend and data-manipulation respectively. I also did a significant portion of the system design and analysis. This work was published in SRDS 2009.

Google: *Intern* **Summer 2006, Summer 2007**
Documented Internal Build System Components using HTML/CSS and Wiki. Studied and documented Code Dependencies in Python, UNIX Shell, and C++ scripts. Used Python to implement statistical techniques to analyze web data.

CENIC: *Systems Administrator* **Oct 2003 - May 2004**
Configured and supported a variety of **internal solutions** for CENIC's network management. This included limited exposure to **Cisco optical equipment** (15808 and 15540), and the deployment of a complete Cisco VoIP solution for the office. Also **administered 5-7 OS X and Solaris servers** including mail, news, Request Tracker, and Ciscoworks.

University of California, Santa Cruz, CA USA:

Research Assistant **March 2008 - present**
Storage Systems research under Dr. Ethan Miller. (see Projects)

University of New Mexico, Albuquerque, NM USA: *Research Assistant* **Jan 2005 - May 2007**
Machine Learning research under Dr. Terran Lane. (see Projects)

Teaching Assistant **Aug 2004 - Dec 2004**
Lead teaching assistant for **Introduction to Java Programming**. Wrote and graded assignments, Taught Labs, Managed and coordinated with 5 TAs, and maintained the class web site.

Harvey Mudd College, Claremont, CA USA *Research Assistant* **May 2002 - Jan 2008**
IDXP/BEEP Intrusion Detection project at Harvey Mudd College, with Prof. Mike Erlinger, Head

of the IDWG. Over Summer 2002, we **implemented a functional IDS** using Perl and Java with help from Aerospace and Silicon Defense. We also monitored traffic using a modification of the snort tool and NFR and send these alerts into a **MySQL database**.

PUBLICATIONS

Avani Wildani, Lee Ward, Ethan L. Miller, *Efficiently Identifying Working Sets in Block I/O Streams*, Proceedings of the 4th Annual International Systems and Storage Conference (SYSTOR 2011), May 2011.

Avani Wildani, Ethan L. Miller, *Semantic Data Placement for Power Management in Archival Storage*, Proceedings of the 5th International Workshop on Petascale Data Storage (PDSW10), held in conjunction with SC2010, November 2010.

Wendy Belluomini, Binny Gill, **Avani Wildani**, Pin Zhou, *GAUL: Gestalt Analysis of Unstructured Logs for Diagnosing Recurring Problem in Large Enterprise Storage Systems*, 29th IEEE International Symposium on Reliable Distributed Systems, November 2010.

Ari Rabkin, Wei Xu, **Avani Wildani**, Armando Fox, Dave Patterson, Randy Katz, *A Graphical Representation for Identifier Structure in Logs*, Workshop on Managing Systems via Log Analysis and Machine Learning Techniques (SLAML 2010), October 2010.

Avani Wildani, Thomas Schwarz, Ethan L. Miller, Darrell D. E. Long, *Protecting Against Rare Event Failures in Archival Systems*, Proceedings of the 17th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2009), September 2009.

Kevin Greenan, Darrell D. E. Long, Ethan L. Miller, Thomas Schwarz, **Avani Wildani**, *Building Flexible, Fault-Tolerant Flash-based Storage Systems*, Proceedings of the Fifth Workshop on Hot Topics in System Dependability (HotDep 2009), June 2009.

TALKS

Avani Wildani, Ethan L. Miller, *Probabilistic Reputation for Personal Trust Networks*, FAST 2009 WIP

SERVICE

FAST 2011 Summary Writer
FAST 2009 Summary Writer
CS Graduate Student Association Treasurer, 2006-2007

STRENGTHS

Programming and Markup Languages I've used for large projects:

Python (including NumPy/Numeric and SimPy), Java, Prolog, MATLAB, L^AT_EX, Shell Scripting, XML, HTML.

Programming Languages I have used in a classroom setting:

C, C++ (STL), SML, Lisp, Ciao, Rex, ns2 simulator Tcl

Software and Operating Systems I can use proficiently:

UNIX Utilities, MATLAB, Microsoft Office
Solaris 7-9, Mac OS X, Windows 3.1 through XP, Linux (Ubuntu, Redhat, Debian, etc.)