Russell Sears

Status

I am a PhD student working with Eric Brewer at the University of California at Berkeley. I plan to complete my PhD in the Summer of 2009, and am seeking a research position in the San Francisco Bay Area.

My PhD research was the development of Stasis, a library of extensible transactional storage primitives for applications that are a poor fit for general purpose transaction systems. I am now working on applying Stasis to a wider range of storage applications.

As an undergraduate I specialized in the development of bioinformatics algorithms and software for large data sets.

Contact Information

Electronic:

sears@cs.berkeley.edu

http://www.cs.berkeley.edu/~sears

Current Research

Stasis: An extensible transactional storage library

Stasis provides concurrent, high-performance transactional storage with ARIES-style recovery and application-specific on-disk data layout. Unlike past implementations, the library provides fine-grained, layered access to its components.

Industry Experience

Riverbed (Summer 2008): Riverbed provides wide area network acceleration systems that perform protocol optimizations and cross-protocol, distributed deduplication of redundant network traffic.

Intel Research Berkeley (Fall 2007): Added C bindings to P2, a distributed datalog implementation, in order to provide Stasis' transactional storage primitives in networked environments.

Google (Summer 2007): Implemented a prototype database replication engine, Rose, which I presented at VLDB 2008. It is based on log structured merge trees and targets low-latency, write-intensive relational workloads. It uses lightweight compression schemes from the column store literature, and can exceed hardware storage bandwidth during replication.

Intel Research Berkeley (2006): Investigated the integration of data flow technologies into Stasis, and developed the LSN-free recovery algorithm and Stasis' allocation protocols.

Microsoft Research, San Francisco (2005): Studied fragmentation properties of SQL Server and NTFS in the context of large scale, reliable storage for web applications.

LabBook (now Rescentris) (2002)

Genomics Viewer: An XML-based web service for the production and visualization of genome annotations. This is a commercialization of the techniques we developed at OSU. I contributed to the design and implementation of the data analysis pipeline, query interface and client-server communication protocols.

Document management: An extension to WebDAV with security properties appropriate for regulated research environments.

LeadScope (2001): Recast the Gene Ontology data set into categories that are relevant to pharmaceutical research. This facilitates the classification and selection of candidate drug compounds.

Network Security Product Prototype (2001): Built a prototype VPN that avoided the use of centralized server infrastructure and used mix networks to prevent outside observers from inferring the contents of messages and pair-wise communication patterns.

Lincoln Electric (1998) : Collected market data that influenced Lincoln Electric's corporate acquisition strategy.

Teaching

Taught sections and developed projects for *CS161: Security* for two semesters, including the first time it was offered.

References

Dr. Eric Brewer
Computer Science, University of California, Berkeley
brewer@cs.berkeley.edu

Dr. Bo Yuan
Biomedical Informatics, Ohio State University
yuanbo@cs.sjtu.edu.cn

Publications

- R. Spillane, R. Sears, C. Yalamanchili, S. Gaikwad, M. Chinni, E. Zadok. *Story Book: An Efficient Extensible Provenance Framework*, to appear TAPP 2009
- R. Sears, M. Callaghan, E. Brewer. *Rose: Compressed, log-structured replication*, VLDB 2008.
- R. Sears, C. van Ingen. *Fragmentation in Large Object Repositories,* CIDR 2007.
- R. Sears, E. Brewer. *Stasis: Flexible Transactional Storage*, OSDI 2006.
- M. Barreno, B. Nelson, R. Sears, A. D. Joseph. *User Model Transfer for Email Virus Detection*, SysML 2006.
- B. Nelson, M. Barreno, R. Sears, A. D. Joseph, J. D. Tygar. *Can Machine Learning be Secure? (Invited paper)* ASIACCS 2006.
- R. Sears. *A Flexible, Extensible Transaction Framework,* HPTS 2005.
- Jun Xu, et al. *Dandruff-associated Malassezia genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens.* Proceedings of the National Academy of Sciences. 2007; 104(47).
- Fred A Wright, et al. *A Draft Annotation and Overview of the Human Genome*, Genome Biology. 2001; 2(7).
- Degen Zhuo, et al. *Assembly, Annotation, and Integration of UNIGENE Clusters into the Human Genome Draft*, Genome Research, 2001.
- Bo Yuan, et al. *Physical Mapping and Functional Annotation of 60,000 Human Genes*, Presented at the SC2000 Conference 11/04/2000 - 11/10/2000.

Technical Skills

- Systems code development: database internals; source-to-source translation and binary instrumentation; static and dynamic program analysis; multithreaded development and testing, superscalar implementation techniques
- Prefer C, Java, C, Perl, OCaml and make, but also proficient in C++, C#, Python, SQL, CGI, servlets, ant, WebDAV, ORM, XML messaging and translation

Education

2003-Present	University of California, Berkeley	PhD student; obtained an MS in Spring of 2005
1997-2002	Ohio State University	Summa Cum Laude with a BS in Computer Science and Engineering

Undergraduate Research

- **Protein Structure analysis (2003)** - Applied multimedia indexing techniques to the Protein Data Bank, allowing an all-by-all comparison of protein substructures. Existing techniques performed pair-wise structure comparison.
- **Genome assemblies / annotations (2000-2004)** - Analyzed human, mouse and proprietary genomes; focused on integration of heterogeneous data, comparative genomics and genome-wide statistical analysis. Our human genome annotation was published concurrently with the Human Genome Consortium and Celera annotations and was later commercialized by Rescentris.
- **Genotyping Automation (1998)** - Designed and implemented a web-based high-throughput genotyping system, including automated signal analysis, interfaces for data visualization and correction, and isolation of systematic error from sources such as contamination.

Academic Awards:

Honor Societies:
Tau Beta Pi**,** Upsilon Pi Epsilon, Alpha Lambda Delta/Phi Eta Sigma, National Honor Society

Financial Awards:
I received a two year ARCS fellowship as a graduate student at UC Berkeley, and my graduate research was partially supported by Sun's MySQL division.

As a undergraduate, I was a National Merit Scholar (full tuition and a stipend), and received a Columbus Dispatch scholarship.