Russell Sears

Status

I am a PhD student working with Eric Brewer at the University of California at Berkeley. My current research provides extensible transactional storage primitives to applications that are a poor fit for general purpose transaction systems. I am implementing these ideas as an open source library named Stasis.

As an undergraduate and in industry 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)
In progress.

Intel Research Berkeley (Fall 2007)
Added simple 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. It is based on log structured merge trees and targets traditional, but write-intensive, relational workloads. It makes use of lightweight compression schemes from the column store literature, giving it the potential to exceed hardware storage bandwidth during replication.

Intel Research Berkeley (2006)
Investigated the use of data flow technologies in Stasis, implemented storage allocation, and developed LSN free pages.

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.

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

- Russell Sears, Mark Callaghan, Eric Brewer. *Rose: Compressed, log-structured replication*, To appear VLDB 2008.
- Russell Sears, Catharine van Ingen. *Fragmentation in Large Object Repositories,* CIDR 2007.
- Russell Sears, Eric Brewer. *Stasis: Flexible Transactional Storage*, OSDI 2006.
- Marco Barreno, Blaine Nelson, Russell Sears, Anthony D. Joseph. *User Model Transfer for Email Virus Detection*, SysML 2006.
- Blaine Nelson, Marco Barreno, Russell Sears, Anthony D. Joseph, J. D. Tygar. *Can Machine Learning be Secure? (Invited paper)* ASIACCS 2006.
- Russell 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, multithreaded development and testing, XML messaging and translation, source-to-source translation, static analysis and binary instrumentation
- Proficient in Java, Perl, C, C++, OCaml, Linux administration; some experience with C#, Python, make, bash, R, Octave, and Lisp
- Web development: SQL, CGI, Java Applets/Servlets, WebDAV, CORBA, Object Relational Mapping

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.

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