

Arifa Nisar

Phone: (312) 375-7157

Email: arifa@soe.ucsc.edu

<http://users.soe.ucsc.edu/~arifa>

Research Interests

Parallel and distributed file systems, high performance computing, parallel storage

Education

Ph.D. Computer Engineering, Northwestern University, Evanston, IL

September 2006 - August 2010

Thesis Title: Parallel I/O Optimizations Through Request Delegation for High-Performance Computing Systems.

Advisor: Prof. Alok Choudhary

Bachelor of Science Electrical Engineering, University of Engineering and Technology Lahore, Pakistan

January 1999 - December 2002

Experience

Storage Systems Research Center, University of California Santa Cruz, Santa Cruz, CA

NSF/CRA Computing Innovation Fellow

October 2010 to Present

Designing and Implementing Adaptive Heterogenous Storage for High Performance Computing.

Center for Ultra-Scale Computing and Information Security, Northwestern University, Evanston, IL USA

Research Assistant

September 2006 to August 2010

I have been involved in many high performance storage and computing related projects which broadly fall into following categories.

- Design and implementation of portable solutions for parallel storage which seamlessly improve the overall performance of computational intensive applications manifold. My research work demonstrates the capability of improving real data intensive applications like combustion, molecular dynamics, fusion, and climate prediction simulations.
- Benchmarking of parallel file systems to determine the best I/O access strategies for a given parallel storage configuration. One such I/O benchmark software has been successfully tested with many Lustre file system deployments to suggest the most optimal I/O access patterns. This I/O benchmark software package is under preparation for public release.
- Design and Implementation of a novel approach to solving the biclustering problem efficiently by reformulating it as a graph drawing problem, This reformulation was also adopted to propose a highly scalable parallel solution for biclustering.

IBM Almaden Research, Almaden, CA USA

Graduate Summer Intern

June-September 2008

Investigation and evaluation of relaxed semantics support for pNFS in ROMIO, an MPI-IO implementation. We evaluated different methods like forced invalidation and revalidation of cache, read ahead avoidance, selective locking, and user controlled flush for providing user controlled semantics strictness.

Mathematics and Science Division, Argonne National Laboratory, Argonne, IL USA

Givens Associate

June-September 2007

Design and implementation of a portable plugin for petascale systems that allocates a small subset of compute nodes to allow user level cache and process collaboration for I/O performance improvement.

Mentor Graphics, Lahore, Pakistan

Senior Software Engineer

September 2003 - July 2006

Porting Mentor Graphic's Flagship Real Time Operating System, Nucleus PLUS, and its product line on several processor architectures including ARM, SAMSUNG, PowerPC, Coldfire, Blackfin, Hitachi, and C167.

Al-Khwarizmi Institute of Computer Science, Lahore, Pakistan

Research Associate

Jan 2003 - August 2003

Design and preparation of a course of Computer Architecture for Virtual University of Pakistan

Selected Publications

Refereed Publications

- **Arifa Nisar**, Wei-keng Liao and Alok Choudhary, Delegation-based I/O Mechanism for High Performance Computing Systems. IEEE Transactions on Parallel and Distributed Systems (To Appear TPDS 2011, PrePrints available)
- **Arifa Nisar**, Waseem Ahmad, Wei-keng Liao and Alok Choudhary, High Performance Parallel/Distributed Biclustering Using Barycenter Heuristic. SIAM International Conference on Data Mining (SDM09).
- **Arifa Nisar**, Wei-keng Liao and Alok Choudhary, Scaling parallel I/O performance through I/O delegate and caching system. Proceedings of the 2008 ACM/IEEE conference on Supercomputing.
- Dean Hildebrand, **Arifa Nisar**, and Roger Haskin. pNFS, POSIX, and MPI-IO: A Tale of Three Semantics. Proceedings of the 4th International Petascale Data Storage Workshop (PDSW '09), November 2009, Portland, Oregon, USA.
- Kui Gao, Wei-keng Liao, **Arifa Nisar**, Alok Choudhary, Robert Ross, and Robert Latham. Using Subfiling to Improve Programming Flexibility and Performance of Parallel Shared-file I/O. In the Proceedings of the International Conference on Parallel Processing, Vienna, Austria, September 2009.
- Wei-keng Liao , Avery Ching, Kenin Coloma, **Arifa Nisar**, Alok Choudhary, Jacky Chen, Ramanan Sankaran, and Scott Klasky, Using MPI File caching to improve parallel write performance for large-scale scientific applications, in SC. The ACM/IEEE Conference on Supercomputing, November 2007

Book Chapters

- Avery Ching, Kenin Coloma, **Arifa Nisar**, Wei-keng Liao, and Alok Choudhary, Book Chapter: Distributed File systems. In Wiley Encyclopedia of Computer Science and Engineering. John Wiley & Sons, Inc., 2007

Invited Publications

- Alok Choudhary, Wei-keng Liao, Kui Gao, **Arifa Nisar**, Robert Ross, Rajeev Thakur, and Robert Latham. Scalable I/O and Analytics. In the Journal of Physics: Conference Series, Volume 180, No. 012048 (10pp), August 2009. (Proceedings of SciDAC conference, 14-18 June 2009, San Diego, California, USA).

Conference and Invited Talks

- Optimizing I/O for Large-Scale Scientific Applications using I/O Delegation, October 2009, MCS Division, Argonne National Lab.
- Eliminating lock contention for parallel I/O using static file domain based I/O delegation, September 2009, CUCIS Seminar, Northwestern University.
- High Performance Parallel/Distributed Biclustering Using Barycenter Heuristic, 2009 SIAM International Conference on Data Mining, Reno, NV.
- Scaling parallel I/O performance through I/O delegate and caching system, 2008 ACM/IEEE conference on Supercomputing, Austin TX.
- Scaling MPI-IO performance with pNFS, August 2008 Intern Talk IBM Almaden Research Center.

Significant Courses Work

Distributed Systems, Design and Analysis of Algorithms, Natural Language Processing, Advanced Computer Architecture, Parallel and Distributed Data Systems, Advanced Computer Networks, Introduction to Parallel Computing

Significant Class Projects:

Kernel Memory Allocation Algorithms: A number of algorithms for kernel memory allocations were implemented. I implemented “Resource Map Allocator”, “Buddy Systems” and “Lazy Buddy System”. My implementation of Resource Map Allocator stood first in the performance contest for efficient memory utilization and performance.

Level of Detail/Scope for Search Engine: A tool was developed on the top of Google Search Engine, which allowed user to adjust required “level of detail and scope” in search results. This implementation exploits the concept of “chaining” to retrieve more/less specific keywords based on selected detail level. Filter refines the results based on features like size of document, length of title etc.

Parallel Biclustering Algorithm: A parallel biclustering algorithm was designed and implemented that minimizes the crossing between rows and columns and then identifies the cluster locally. Data was partitioned horizontally to attain accurate solution in a parallel computing framework. This work has been published in 2009 SIAM International Conference on Data Mining.

Technical Skills

Programming: C, C++, Fortran, MPI, MPI-IO

Large-scale Systems: Cray Systems, Bluegene Systems, Teragrid Resources

Proficient in commercial parallel file systems protocols: Lustre, GPFS, pNFS, PVFS

Programming in Class Setup: JAVA, LISP, MATLAB

Service Activities

- Active research member of Scientific Data Management Community, A SciDAC Center of the Department of Energy (Dec 2006 - present)
- Volunteer at High Performance Distributed Computing (HPDC) 2010
- Reviewer for:
 - The ACM/IEEE Conference on Supercomputing 2007, 2008, 2009.
 - International Conference on Parallel Processing 2009.

Teaching Assistant

- Teaching Assistant: Department of Electrical Engineering and Computer Science
 - EECS-311 “Data Structures and Data Management” Spring 2009
 - EECS-339 “Introduction to Database Systems” Fall 2009
- Searle Center for Teaching Excellence Northwestern University. **Planning an Effective Class Period:** This course systematically covered different stages from design to execution of a lecture plan. The course introduced the important tools to implement the class activities while keeping the overall course goals in vision.

Honors and Awards

Morrison Graduate Fellowship, Northwestern University, Fall 2009.

Givens Associate Fellowship, Argonne National Laboratory, Summer 2007.

Merit Scholarship, Lahore Board of Education, Pakistan 1999-2002.

Employment Authorization

US Permanent Resident