Dantong Yu

Brookhaven National Lab, Building 463, Upton, NY 11973
Office: (631)-344-3042
E-Mail: dtyu@bnl.gov

RESEARCH ENGINEER AND MANAGER

Computing research engineer with notable success directing, managing, and implementing a broad range of initiatives in data mining, data analyze and management, and network quality of service; Demonstrate knowledge in planning, design, and implementation of solutions in support of data center objectives, and related organizational leadership in addition to cutting-edge research and technology, strategic planning, project funding management, and methodology to convert user requirements into technical solutions. Five years experience in managing and improving facility operations with service level agreement and troubleshooting complex computing systems; An excellent functional and project manager to manage, motivate and guide team members; Panel members in reviewing and recommending DOE and NSF research proposals.

AREAS OF EXPERTISE

- Data Mining
- Planning/implementation
- Peta-Byte data storage management
- Project management life cycle
- High Performance Computing

- Network monitoring, performance diagnosis and tuning and QOS
- High performance data transfer
- Leadership and interpersonal skills
- Scientific papers and technical reports

WORK EXPERIENCE

Brookhaven National Laboratory (BNL), Upton, New York,   **Oct/2009-Present**

Computer Science Lead, Computational Science Center

- Data mining for data intensive sciences.
- High-performance network research.

Brookhaven National Laboratory (BNL), Upton, New York,   **April/2001-September/2009**

Group Lead (I-9)

Analyzed the evolving computing requirements of physics experiments and designed solutions to support their computing objectives. Managed the Grid operation effort for large physics collaboration. Coordinated the Grid software development. Led Peta-scale data storage management. Coordinated network research activities. Managed the user support team.

Network Research Experiences

- Analyzed the computing and networking resource utilization and developed strategies to maximize network performance and data throughput, coordinated the data transfer between BNL and other ATLAS and RHIC collaboration institutes:
  - Let the BNL STAR WAN data transfer project, and architected and tuned the data transfer framework from BNL on-line data acquisition system and storage resource management (SRM) to Korea institute of Science and Technology Information (KiSTi): a factor of ten performance improvement over the un-tuned system, and a cost-effective solution for near real-time raw data replication from BNL STAR detector to the remote event processing site.

- o Led the BNL facility effort of using Grid Ftp tools to annually transfer 700 tera-byte proton-to-proton data from BNL PHENIX counting house to the PHENIX Computing Center in Japan (CCJ) during RHIC run 2005 to 2008. Tuned the network, configured Linux proxy server based on IP masquerading and NAT to bridge non-routable private VLAN to public routable VLAN.

- o Led the BNL series of service challenge to exercise ATLAS data transfer framework and transfer real data from CERN to BNL disk and tape based storage. Achieved and exceeded 300M Byte/second sustained data rate required by the ATLAS Computing Model. BNL's data capacity and performance exceeded all of the other eight ATLAS Tier 1 sites.

- o Familiar with iperf, tcpdump, Ethereal tools to analyze TCP stream, IP packets and diagnose packet loss problems. Discovered and fixed the CISCO firewall rewriting WSCALE bits in TCP SYN packets.

- o Led the USATLAS bandwidth initiative and tuned TCP/IP between BNL and five USATLAS tier 2 universities. Had hand-on experience to tune host level TCP/IP perimeters to improve end-to-end data transfer rate.

- o Used CISCO netflow tools for TCP flow based performance monitoring from BNL to Tier 2 sites (http://netmon.usatlas.bnl.gov/netflow/index.html).

Managing and coordinating in Grid Computing Group

- Expanded Grid Computing from two to thirteen full time employees. The main responsibility is to manage and motivate twelve members in the Group. Assigned tasks to group members, kept track the project progress, and evaluated their performance.

- Led and coordinated the world largest peta-scale Grid enabled data storage system at BNL. Architected USATLAS high reliable network infrastructure to support high performance data transfer while maintaining DOE class B network security. Supervised the dCache software deployment for the USATLAS experiment to allow ATLAS international collaborators to access data with Grid middleware tools. Architected the dCache Infrastructure to satisfy both security and high performance requirements.

- Led the deployment of physics production and analysis framework and its 24x7 operation support for thousands of ATLAS physicist users from North America, Europe, and Asian. Architected OS/hardware level solutions for high availability, job throughput, and scalability.

- Architected the Oracle Real Application Cluster (RAC) to replicate the relational DB based condition data stream from CERN Geneva to BNL in real time.

- Architected and managed the Operation Integration Projects which manages large scale hardware inventory, monitors Grid services, automates alarm generation with different severity levels, automatically opens a problem ticket at the facility ticket system to allow status tracking of the facility response. System used: Nagios and RT (Report Tracking).

- Drafted on-call procedures and service level agreement to ensure quality assurance.

- Led the Grid software deployment effort at BNL, deployed the Open Science Grid software for the experiments: USATLAS, STAR, and PHENIX.

- Led the ATLAS grid computing-facility monitoring effort, developed and designed tools for monitoring and tuning GRID systems which consist of many geographically distributed server nodes.

- Designed and developed the first version of Grid User Management System (GUMS), Used Perl. GUMS evolved to be a highly reliable software product to provide a Grid identity mapping

service for a national wide collaboration, the US open science grid.

- Published research result from the tasks listed above.

Stony Brook University, Dept. of Electrical & Computer Eng., January/2008-Present
Adjunct Faculty: teach a graduate level course ESE 550 (Network management and planning).

State University of New York at Buffalo   **May/1998-August 2000**
Research Assistant of National Center of Geographic Information Agency at Buffalo (NCGIA), Buffalo, NY. Department of Defense (DoD). This project helped to rapidly identify interesting targets from large image database snapped by satellites and airplanes.

- Designed novel clustering approaches that can detect clusters of arbitrary shape from multidimensional image datasets, used C++.
- Designed a multi-dimensional index structure to optimize the near neighbor search in large datasets, used C++.
- Developed software with GUI interface to do the content-based geographical image classification, analysis and retrieval, used JAVA language.

Brookhaven National Lab, Summer Internship   **May/20/1997-August/20/1997**

- SGI workstation programming with C and C++, developed scientific data visualization packages to take advantage of the hardware supported graphical libraries to replace the exiting software-based libraries. The new packages increased frame rate per second required by the real time operations.
- SGI system administrator. Configured the system software, responsible for system backup.
- Sun workstation consultant. Provided answers to the Unix-related questions.

EDUCATIONS

Ph.D. in Computer Science, State University of New York at Buffalo. Dated in February 2001.

Dissertation: "*Multidimensional Indexing and Management for Large-Scale Databases*".

Master of Science in Computer Science, State University of New York at Buffalo. Dated in September 1998.

Bachelor of Science in computer science. Beijing University, Dated in July 1995.

PROFESSIONAL ACTIVITIES

- Co-chair of DOE Advanced Networking for Distributed Petascale Science R&D Challenges and Opportunities, https://www.bnl.gov/DOENetwork/, April 8-9, 2008, and Chair of the workshop of "DOE office of Science High-Performance Network Research PI", September 28-30, 2005. Brookhaven National Laboratory.
- NSF Cyber-enabled Discovery and Innovation (CDI), Panelist
- Review Panel for DOE Early Career Principle Investigator for networking research
- Review Panel for DOE Small Business Innovative Research (SBIR), 2004-2009.
- Reviewer for International Journal of Computers and their applications, Special Issue on Cluster/Grid Computing, Journal of ACM Multimedia Systems: Special Issue on Content-Based Image Retrieval (CBIR), International Conference on Data Engineering, International Conference on Knowledge Discovery and Data Mining, 2002, ACM Multimedia, and the International Conference on Scientific and Statistical Database Management.

Dantong Yu

GRANTS

- **PI**, *Department of Energy (DOE), $1,200,000, 10/01/2004-12/31/*2009, "TeraPaths: A QoS Enabled Collaborative Data Sharing Infrastructure for Peta-scale Computing Research".

- **Co-PI**, *Department of Energy (DOE), $1,500,000, 10/01/2009-05/31/2012. "*End Site Control Plane Subsystem".

- **PI**, *Department of Energy (DOE), $1,600,000, 10/01/2009-09/31/2012, "*Integrating Storage Resource Management with Dynamic Network Provisioning for Automated Data Transfer".

- **PI**, *Department of Energy (DOE), $1,500,000, 12/01/2009-11/30/2012,* "100G FTP: An Ultra-High Speed Data Transfer Service Over Next Generation 100 Gigabit Per Second Network".

- **PI***, Stony Brook University, matching fund, $105,000, 12/01/2009-11/30/2012,* "100G FTP: An Ultra-High Speed Data Transfer Service Over Next Generation 100 Gigabit Per Second Network".

- **PI***, Stony Brook University, College of Engineering and Applied Sciences, matching fund, $45,000, 12/01/2009-11/30/2012*.

- **PI**, Brookhaven National Lab, Laboratory Directed Research and Development (LDRD), $492,189, *10/01/2009-09/30/2011,* "Petascale Data Mining for BNL Data Intensive Sciences.

JOURNAL PUBLICATIONS

1. M.A. Moges,_D. Yu, T.G. Robertazzi, "Grid scheduling divisible loads from two sources", Computers and Mathematics with Applications, 1081-1092, 58, 2009.

2. D. Yu and A. Zhang, "*ClusterTree*: Integration of Cluster Representation and Nearest Neighbor Search for Large Datasets with High Dimensionality", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, Volume 15, Number 5, September, 2003.

3. D. Yu, G. Sheikholeslami, and A. Zhang, "*FindOut*: Finding Outliers in Very Large Datasets", *Knowledge and Information Systems (KAIS), An International Journal*, Volume 4, Number 4, pp.387-412, October 2002.

REFEREED CONFERENCE PUBLICATIONS

1. Katramatos, D., Shroff, K., Yu, D., Mckee, S., and Robertazzi, T., "Establishment and Management of Virtual End-to-End QoS Paths through Modern Hybrid WANs with TeraPaths", The First International Conference on Evolving Internet (Internet 2009), Cannes/La Bocca, French Riviera, France, August, 2009, Best Paper Award.

2. Yu, X, Qiao, C., and Yu, Dantong, "Online Job Provisioning for Large Scale Science Experiments over an Optical Grid Infrastructure", IEEE INFOCOM 2009 High-Speed Networks Workshop (HSN 2009), Rio de Janeiro, Brazil, April 24, 2009.

3. H. Xu, D. Yu, D. Xu, and A. Zhang, "SS-ClusterTree: a subspace clustering based indexing algorithm over high-dimensional image features", Proceedings of the 2008 international conference on Content based image and video retrieval, Niagara Falls, Canada, 2008.

4. D. Katramatos, B. Gibbard, D. Yu, S. Mckee, "*The TeraPaths Testbed: Exploring End-to-End Network QoS*", 2007 3rd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCom).

5. B. Gibbard, D. Katramatos, D. Yu, and S. Mckee, "*TeraPaths: End-to-End Network Path QoS Configuration Using Cross-Domain Reservation Negotiation*", Third International Workshop

on Networks for Grid Applications (GridNets), October, 2006, San Jose, California, USA.

6. D. Yu and T.G., Robertazzi, "*Multi-Source Grid Scheduling for Divisible Loads*", 2006 Conference on Information Sciences and Systems, Princeton University, Princeton NJ, March 2006.

7. D. Yu, etc, "*BNL Wide Area data transfer for rhic and atlas: Experience and Plan*", the 2006 Conference for Computing in High Energy and Nuclear Physics, Mumbai, India, Feb, 2006.

8. D. Yu, etc, "*TeraPaths: A QoS-Enabled Collaborative Data Sharing Infrastructure For Peta-Scale Computing Research*", the 2006 Conference for Computing in High Energy and Nuclear Physics (CHEP06), Mumbai, India, Feb, 2006.

9. M. Moges, D. Yu, and T.G. Robertazzi, "*Divisible Load Scheduling with Multiple Sources: Closed Form Solutions*", 2005 Conference on Information Sciences and Systems, The Johns Hopkins University, Baltimore, Maryland, March 2005.

10. M. Moges, D. Yu, and T.G. Robertazzi, "*Grid Scheduling Divisible Loads from Multiple Sources via Linear Programming*", Proceedings of the IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2004), Cambridge, Massachusetts, Nov. 2004.

11. S. Viswanathan, B. Veeravalli, D. Yu, T.G. Robertazzi, *"Design and Analysis of a Dynamic Scheduling Strategy with Resource Estimation for Large-Scale Grid Systems"*, In the Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing (held in Conjunction with SuperComputing 2004), Pittsburgh, Pennsylvania, USA, pp. 163-171, November 2004.

12. D. Yu and T. G. Robertazzi, "*Divisible Load Scheduling for Grid Computing*", the 15th IASTED International Conference on Parallel and Distributed Computing And Systems, November, 2003, Marian Del Rey, CA, USA.

13. H. M. Wong, D. Yu, B. Veeravalli and T. G. Robertazzi, "*Data Intensive Grid Scheduling: Multiple Sources with Capacity Constraints*", the 15th IASTED International Conference on Parallel and Distributed Computing And Systems, November, 2003, CA, USA.

14. R. Baker, D. Yu and T. Wlodek, "*A Model for Grid User Management*", 2003 Conference for Computing in High Energy and Nuclear Physics, La Jolla, California, March, 2003.

15. R. Baker, D. Yu and J. Smith, "*GridMonitor, Integration of Large Scale Facility Fabric Monitoring with Meta Data Service in Grid Environment*", 2003 Conference for Computing in High Energy and Nuclear Physics, La Jolla, California, March, 2003.

16. D. Yu and A. Zhang, "*ClusterTree: Integration of Cluster Representation and Nearest Neighbor Search for Image Databases*", IEEE International Conference On Multimedia and Expo (ICME), pp.1713-1716, New York City, New York, USA, August, 2000.

17. Y. Sun, D. Yu, R. Acharya, R. S. Gaborski "*Knowledge-based method for fully automatic contour detection in radiographs*", SPIE's International Symposium on Medical Imaging, Vol.1, No.24, pp.256-264, San Diego, California, USA, February, 2000.

18. D. Yu, S. Chatterjee, and A. Zhang, "*Efficiently Detecting Arbitrary Shaped Clusters in Image Databases*", Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'99), pp.187-192, Chicago, Illinois, November, 1999.

19. D. Yu and A. Zhang, "*ACQ: An Automatic Clustering and Querying Approach for Large Image Databases*", Proceedings of ACM Multimedia 99, part 2, pp.95-98, Orlando, 1999.

Dantong Yu

SAMPLE PRESENTATIONS

1. "*Terapaths*", Dantong Yu, DOE Office of Science High-Performance Network Research PI Meeting, September 28~September 30, 2005.

2. "*USATLAS Tier 1 Facility and Service Challenge*", Dantong Yu, HEPiX 2005 at Germany.

3. *"BNL VO Management and Grid Map File Generation",* Dantong Yu at US ATLAS Grid Workshop at Lawrence Berkeley National Laboratory, Berkeley, CA, July 29, 2002.

4. *"Atlas Monitoring Infrastructure in Grid Environment",* presented by Dantong Yu at the Particle Physics Data Grid (PPDG) Collaboration Meeting, Canada, February 20, 2002.

5. *"Network-related Issues in a Grid Environment"*, presented by Dantong Yu at the first HENP Networking Working Group Meeting, Ann Arbor, Michigan, Oct. 26, 2001.

6. *"End-to-End Network Performance Monitoring and Tuning",* presented by Dantong Yu at the Particle Physics Data Grid (PPDG) Collaboration Meeting, Madison, Wisconsin, 2001.