

Jong Youl Choi

Oak Ridge National Laboratory
PO BOX 2008 MS6290
Oak Ridge, TN 37831-6290
Phone: 812-606-8435
choij@ornl.gov
<http://www.ornl.gov/~jyc/>

Research Interests

Areas of research interest span data mining and machine learning algorithms, high-performance data-intensive computing, and parallel and distributed systems for Cloud and HPC. More specifically, researching high-performance data mining algorithms and developing efficient run-time environments in Cloud and HPC system.

Education

2004 – 2012	Indiana University, Bloomington, Indiana, USA. Ph.D candidate. in Computer Science (Major: Software Systems, Minor: Algorithm)
Thesis	Unsupervised learning of finite mixture models with deterministic annealing for large-scale data analysis
Advisor	Professor Geoffrey Fox
May 2004	New York University, New York, New York, USA M.S. in Computer Science
Feb 1998	Hanyang University, Seoul, South Korea B.S. in Industrial Chemistry (a.k.a., Chemical Engineering)

Professional Experience

Oct. 2012 – Current	HPC Data Research Scientist Scientific Data Group, Computer Science and Mathematics Division Oak Ridge National Laboratory, Oak Ridge, Tennessee Group Leader : Dr. Scott Klasky
Feb. 2012 – Sept. 2012	Post-doc, Univ. of Tennessee, Knoxville, Tennessee Working in Scientific Data Group, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee Advisor : Dr. Scott Klasky
2006 – 2012	Research Assistant, Indiana University, Bloomington, Indiana Department : Pervasive Technology Institute Advisor : Professor Geoffrey Fox
Jun. 07 – Sep. 07	Research Internship, Microsoft Research, Redmond, Washington Department : Technical Computing @ Microsoft group Advisor : Dr. Savas Parastatidis
Oct. 05 – Jan. 06	Research Internship, École Polytechniques Fédérales de Lausanne (EPFL), Lausanne, Switzerland Department : Computer Communications and Applications Laboratory Advisor : Professor Jean-Pierre Hubaux

Sep. 04 – Sep. 05	Research Assistant, Indiana University, Bloomington, Indiana Department : Computer Science
Jun. 03 – Aug. 03	Advisor : Professor Markus Jakobsson Research Internship, New York University, New York, New York Department : MIS Department of Stern School of Business
Feb. 98 – Sep. 01	Advisor : Professor Shinkyu Yang Project Manager, Samsung SDS, Seoul, South Korea Developed an Electronic Data Interchange (EDI) system and managed projects for e-Government Finance Management systems for the South Korean government

Summary of Research Accomplishments

ADIOS-P (2012 – present)	Data provenance and predictive pre-fetcher in ADIOS Developing data provenance and predictive data pre-fetcher modules for ADIOS[eScience '12].
DA-PLSI (2011 – present)	Mixture model problems with the Deterministic Annealing A mixture model problem is to find an optimal mixture distribution of conditional probabilities and it is very common in many data mining areas, such as text mining, image processing, speech recognition, to name a few. However, traditional solutions largely depend on the EM method which can only find local solutions. With the novel Deterministic Annealing method, I am currently researching on solving Probabilistic Latent Semantic Indexing (PLSI), one of mixture model problems, in order to find global solutions as well as finding model parameters in an adaptive way.
PlotViz (2010 – present)	3D data point visualization for large and high-dimensional data Developing a visualization tool for large and high-dimensional data, as an application for data-intensive life science data analysis. Consuming an output from high-performance dimension reduction algorithms, like parallel DA-GTM, the tool visualizes a large data set in 3D. The program is written in C++ by using VTK and Qt and available from http://salsahpc.indiana.edu/pviz3/ [CCPE '11][ECMLS '10].
DA-GTM (2009 – present)	Parallel Generative Topographic Mapping and Deterministic Annealing Developed the high-performance parallel Generative Topographic Mapping (GTM), an algorithm for dimension reduction and visual data mining, as a data-intensive application for large and high-dimensional life science data analysis. To overcome a local optimum problem appearing in the conventional GTM, applied an innovative global optimization method, Deterministic Annealing (DA), and devised a new DA-GTM algorithm [CCPE '11, CCPE '11, Bioinformatics '10][HPDC '10, ECMLS '10, ICCS '10, CCGRID '10].

Folksonomy Mining (2008)	Collaborative tagging system for folksonomy mining Developed Collective Collaborative Tagging (CCT) system for folksonomy data mining to build a recommendation and searching system. Also, performed analysis of folksonomy data by exploiting graph structures of tagging [CCPE '09] [GCE '08, CTS '08].
V-Lab (2007)	Virtual labs in Clouds : V-Lab-Protein and V-Lab-Microarray Developed virtual collaborative labs in Clouds for bioinformatics applications, Virtual Collaborative Lab for Protein Sequence Analysis (V-Lab-Protein) and Microarray Data Analysis (V-Lab-Microarray). Designed V-Labs to provide virtual and volatile cloud computing resources by using Amazon's EC2 and S3 services. Equipped with workflow engines along with a user-friendly graphic workflow composer, V-Labs provide easy-to-use cloud computing environments to bioinformatics researchers [eScience '08, BIBM '07]. This project is also known as one of the first scientific applications using cloud computing infrastructures and it has been introduced in the October 2007 issue of Nature's News.
Big Red Portal (2007)	A portal development for IU's Big Red super computer, including MEME job submission and job dashboard portlet [TeraGrid '07]
PRECIP and Spysshield (2007)	Privacy protection from spyware Implemented two anti-spyware architectures, PRECIP and Spysshield, to protect user privacy. Developed Spysshield as a Browser Helper Object (BHO), also known as a toolbar plug-in, of the Internet Explorer [RAID '07, NDSS '08]
Packet Vaccine (2006)	Malware detection and signature generation from packets Implemented a system for malicious packet detection by using a packet-based black-box approach. The system can identify malicious codes inside a network packet for fast detection [TISSEC '08] [CCS '06]
Tamper-evident schemes (2005)	Tamper-evident signatures and mix networks Developed a tamper-evident signature scheme and a mix network. A new temper-evident signature scheme can allow a user to detect whether a signer is corrupted or not. Designed a temper-evident mix network to protect the system from malicious attempts [DASC '06, FC '06]
Vehicular Network Protocol (2005)	Secure protocol designs for vehicular networks Designed a secure vehicular network protocol to achieve both auditability and privacy [Q2SWinet '05]
VPN (2004)	Performance study for Virtual Private Network (VPN) Research on IPSec overheads for VPN servers [ICNP '05]

Journal Publications

CCPE '11	J. Y. Choi , S.-H. Bae, J. Qiu, B. Chen, and D. Wild, "Browsing large scale cheminformatics data with dimension reduction," <i>Concurrency and Computation: Practice and Experience</i> , 2011. [PDF]
CCPE '11	T. Gunarathne, S. Wu, J. Y. Choi , S.-H. Bae, and J. Qiu, "Cloud computing paradigms for pleasingly parallel biomedical applications," <i>Concurrency and Computation: Practice and Experience</i> , 2011. [PDF]

Bioinformatics '10	J. Qiu, J. Ekanayake, T. Gunarathne, J. Choi , S.-H. Bae, H. Li, B. Zhang, T.-L. Wu, Y. Ruan, S. Ekanayake, A. Hughes, and G. Fox, "Hybrid cloud and cluster computing paradigms for life science applications," <i>BMC Bioinformatics</i> , vol. 11, no. Suppl 12, p. S3, 2010. [PDF]
CCPE '09	M. Pierce, G. Fox, J. Choi , Z. Guo, X. Gao, and Y. Ma, "Using Web 2.0 for scientific applications and scientific communities," <i>Concurrency and Computation: Practice and Experience</i> , vol. 21, no. 5, pp. 583–603, 2009. [PDF]
TISSEC '08	X. Wang, Z. Li, J. Choi , J. Xu, M. Reiter, and C. Kil, "Fast and black-box exploit detection and signature generation for commodity software," <i>ACM Transactions on Information and System Security (TISSEC)</i> , vol. 12, no. 2, p. 11, 2008.

Selected Conference/Workshop Publications

eScience '12	J. Y. Choi , H. Abbasi, D. Pugmire, N. Podhorszki, S. Klasky, C. Capdevila, M. Parashar, M. Wolf, J. Qiu, and G. Fox, "Mining hidden mixture context with adios-p to improve predictive pre-fetcher accuracy," in <i>eScience</i> , 2012.
ECMLS '10	J. Y. Choi , S.-H. Bae, J. Qiu, G. Fox, B. Chen, and D. Wild, "Browsing large scale cheminformatics data with dimension reduction," in <i>Workshop on Emerging Computational Methods for Life Sciences (ECMLS), in conjunction with the 19th ACM International Symposium on High Performance Distributed Computing (HPDC) 2010</i> , HPDC '10, (Chicago, Illinois), pp. 503–506, ACM, June 2010. [PDF]
HPDC '10	S.-H. Bae, J. Y. Choi , J. Qiu, and G. C. Fox, "Dimension reduction and visualization of large high-dimensional data via interpolation," in <i>Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing</i> , HPDC '10, (Chicago, Illinois), pp. 203–214, ACM, June 2010. [PDF]
ICCS '10	J. Y. Choi , J. Qiu, M. Pierce, and G. Fox, "Generative Topographic Mapping by Deterministic Annealing," in <i>Proceedings of the 10th International Conference on Computational Science and Engineering (ICCS 2010)</i> , 2010. [PDF]
CCGRID '10	J. Y. Choi , S.-H. Bae, X. Qiu, and G. Fox, "High performance dimension reduction and visualization for large high-dimensional data analysis," <i>Cluster Computing and the Grid, IEEE International Symposium on</i> , vol. 0, pp. 331–340, 2010. [PDF]
GCE '08	J. Choi , J. Rosen, S. Maini, M. Pierce, and G. Fox, "Collective Collaborative Tagging System," in <i>Grid Computing Environments Workshop, 2008. GCE'08</i> , pp. 1–7, 2008. [PDF]
CTS '08	M. Pierce, G. Fox, J. Rosen, S. Maini, and J. Choi , "Social networking for scientists using tagging and shared bookmarks: a Web 2.0 application," in <i>Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on</i> , pp. 257–266, 2008. [PDF]
eScience '08	Y. Yang, J. Choi , K. Choi, M. Pierce, D. Gannon, and S. Kim, "BioVLAB-Microarray: Microarray Data Analysis in Virtual Environment," in <i>IEEE Fourth International Conference on eScience</i> , 2008. [PDF]

BIBM '07	J. Choi , Y. Yang, S. Kim, and D. Gannon, "V-lab-protein: Virtual collaborative lab for protein sequence analysis," in <i>IEEE Workshop on High-Throughput Data Analysis for Proteomics and Genomics, Workshop at BIBM 2007</i> , 2007. [PDF]
TeraGrid '07	M. Nacar, J. Choi , M. Pierce, and G. Fox, "Building a Grid Portal for Teragrid's Big Red," <i>Proceedings of TeraGrid</i> , vol. 2007, 2007.
RAID '07	Z. Li, X. Wang, and J. Y. Choi , "Spyshield: preserving privacy from spy add-ons," in <i>Proceedings of the 10th international conference on Recent advances in intrusion detection</i> , RAID'07, (Berlin, Heidelberg), pp. 296–316, Springer-Verlag, 2007. [PDF]
NDSS '08	X. Wang, Z. Li, N. Li, and J. Choi , "Precip: Towards practical and retrofittable confidential information protection," in <i>Network and Distributed System Security Symposium (NDSS)</i> , 2008.
CCS '06	X. Wang, Z. Li, J. Xu, M. Reiter, C. Kil, and J. Choi , "Packet vaccine: Black-box exploit detection and signature generation," in <i>Proceedings of the 13th ACM conference on Computer and communications security</i> , p. 46, ACM, 2006.
DASC '06	J. Y. Choi , P. Golle, and M. Jakobsson, "Tamper-evident digital signature protecting certification authorities against malware," in <i>Proceedings of the 2nd IEEE International Symposium on Dependable, Autonomic and Secure Computing</i> , DASC '06, (Washington, DC, USA), pp. 37–44, IEEE Computer Society, 2006. [PDF]
FC '06	J. Choi , P. Golle, and M. Jakobsson, "Auditable privacy: On tamper-evident mix networks," in <i>Proceeding of Financial Cryptography and Data Security</i> (G. D. Crescenzo and A. Rubin, eds.), vol. 4107 of LNCS, (Anguilla, British West Indies), pp. 126–141, Springer, February 2006.
Q2SWinet '05	J. Y. Choi , M. Jakobsson, and S. Wetzel, "Balancing auditability and privacy in vehicular networks," in <i>Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks</i> , Q2SWinet '05, (New York, NY, USA), pp. 79–87, ACM, 2005. [PDF]
ICNP '05	C. Shue, Y. Shin, M. Gupta, and J. Choi , "Analysis of IPSec overheads for VPN servers," in <i>IEEE ICNP's NPsec Workshop</i> , 2005.

Technical Skills

Languages	C/C++, C#, Fortran, Java, Python, PL/SQL, Unix shell scripts
Parallel Programming	MPI, OpenMP, Microsoft Dryad
High-performance Platform	Cray, TeraGrid, Microsoft HPC
Performance Tuning	Vampire, PAPI, and VTune
Statistical Tool	R, MATLAB, Octave
Visualization	VTK, Qt, OpenGL, Microsoft XNA
Operating System	Unix/Linux, Mac OS X, Microsoft Windows, Microsoft HPC Server
Database	Oracle DBMS, Oracle DB tuning and backup & recovery, PostgreSQL, MySQL
Certification	Oracle Certified Professional (OCP) with 3+ year field experience

List of References

- | | |
|--------------|--|
| Geoffrey Fox | Associate Dean for Graduate Studies and Research, School of Informatics and Computing, Indiana University - Bloomington
Distinguished Professor of Informatics and Computing and Physics, Indiana University - Bloomington
Director of Digital Science Center of the Pervasive Technology Institute and Director of the Community Grids Laboratory, Indiana University - Bloomington
Informatics West, Room 207
901 E. 10th St.
Bloomington, IN 47408
Phone : (812) 856-7977
Email : gcf@indiana.edu |
| Judy Qiu | Assistant Professor
School of Informatics and Computing
Indiana University at Bloomington
Lindley Hall, Room 201D
150 S. Woodlawn Ave.
Bloomington, IN 47405
Email : xqiu@indiana.edu |
| David Wild | Assistant Professor
School of Informatics and Computing
Indiana University at Bloomington
Lindley Hall, Room 330B
150 S. Woodlawn Ave.
Bloomington, IN 47405
Email : djwild@indiana.edu |