Hyogi Sim

Oak Ridge National Laboratory

simh@ornl.gov, +1-865-574-6167

Current Position

HPC Systems Engineer   *National Center for Computational Sciences, Oak Ridge National Laboratory*

Education

Ph.D. in Computer Science, Virginia Tech, VA   *— expected in 2020*
- Advisor: Dr. Ali R. Butt
- Thesis: Exploiting Storage-Side Computing Power for Fast and Efficient Scientific Data Discovery *(working title)*

M.S. in Computer Science, Virginia Tech, VA   *— Dec. 2014*
- Advisor: Dr. Ali R. Butt
- Thesis: AnalyzeThis: An Analysis Workflow-Aware Storage System

M.S. in Electronics and Computer Engineering, Hanyang University, S. Korea   *— Feb. 2008*
- Thesis: A Study of Performance Impact of Merging Storage Layers on Flash-Based DBMS

B.S. in Urban Planning, Hanyang University, S. Korea   *— Feb. 2005*

Work History

HPC Systems Engineer, Oak Ridge National Laboratory	*Feb. 2017 – present*
Postmasters Research Assistant, Oak Ridge National Laboratory	*Mar. 2015 – Jan. 2017*
Summer Intern, Oak Ridge National Laboratory	*Jan. 2013 – Aug. 2013*
Graduate Research and Teaching Assistant, Virginia Tech	*Aug. 2011 – Dec. 2013, Sep. 2013 – Feb. 2015*
Senior Developer, Pitapat Mobile, S. Korea	*Jan. 2011 – June 2011*
Part-Time Programmer, Dept. of English Education, Hanyang University, S. Korea	*Aug. 2009, Feb. 2010*
Part-Time Programmer, Metabuild co., S. Korea	*Dec. 2008*
Senior Developer, Golfschool co., S. Korea	*2002 – 2003*
Military Conscription in Combat Police Force, S. Korea	*May 1999 - Jul. 2001*

Skills & Expertise

Skillful in Linux application/system/kernel programming, parallel/distributed file systems, database systems.
- Programming Languages: C, C++, Python, JAVA, PHP, C#, BASH
- File & Storage: FUSE, Linux VFS, Device Mapper, SCSI OSD-2, Linux TGT, blktrace, exofs, ext3
- Parallel & Distributed File Systems: GlusterFS, Lustre, Ceph, NFS, HDFS
- Non-Volatile Memory & SSD: Linux MTD, Flash Translation Layer
- Databases: MySQL, PostgreSQL, SQLite, MS SQLServer, IBM DB2, WiSS, HyperDex
- Parallel Programming: MPI, Pthread, OpenMP, Spark
- Parallel I/O Library: netCDF, HDF5, SCR
- I/O Benchmarks: fio, mdtest, IOR, TPC-C, TPC-H
- Tools: gcc, gdb, cscope, ctags, autotools, git, svn, eclipse, glade, visual studio, latex, gnuplot
- Web Development: HTML, CSS, PHP, Javascript, JQuery, ASP.NET
- Package Management: RPM, Portage (Gentoo Linux), Pkgsrc (NetBSD)

Project Experience in File and Storage Systems

A metadata indexing framework for HPC archival storage systems   *2018 – present, Oak Ridge National Lab.*
- Developing a software framework that periodically collects metadata from HPSS and builds data catalog.
- Developing a framework that extracts metadata from well-known document files without user intervention.
- Developing command-line utilities for system administrators and users.
- *(HPSS, IBM DB2, PostgreSQL, Apache Tika, XDR, cURL)*

Development of a userspace checkpoint file system for HPC burst buffers   *2017 – present, Oak Ridge National Lab.*
- Designing and developing an ephemeral, distributed file system for node-local burst buffers to facilitate checkpointing of HPC applications.

○ *(FUSE, MPI, MDHIM, LevelDB, Mercury, Margo, Argobots)*

Large-scale file system snapshot and workload analysis   *2016 – 2019, Oak Ridge National Lab.*
  ○ Analyzing the snapshot data of Spider II (32 PB Lustre) file system.
  ○ Analyzing the eight year worth workloads of 80 PB HPSS archival storage system.
  ○ *(Lustre, HPSS, Spark, SparSQL, Parquet, Python, MySQL, SQLite)*

File system-integrated search and discovery services for HPC   *2015-2017, Oak Ridge National Lab.*
  ○ Designing and developing a file system-integrated metadata indexing framework that supports user-defined tagging in GlusterFS and CephFS.
  ○ *(GlusterFS, CephFS, Linux, SQLite)*

Analysis-aware storage system for high performance computing   *2013 – 2015, Oak Ridge National Lab./Virginia Tech*
  ○ Designing an active execution framework based on SCSI T10 OSD-2 specification and implemented the extended OSD-2 protocol on the Linux TGT.
  ○ Extending the Linux exofs file system to support the extended OSD protocol.
  ○ Developing a FUSE file system that manages the array of active OSD devices.
  ○ *(Linux Kernel, SCSI T10 OSD-2 Protocol, Linux TGT, FUSE, SQLite)*

Hierarchical data management in media servers with hybrid storage architecture   *2007 – 2009, Hanyang University*
  ○ Developing a content popularity analyzer that dynamically analyzes incoming requests to determine popular media contents in a commercial media server.
  ○ Developing a light-weight file system for storage class memory devices (PRAM) in media cache servers.
  ○ *(Linux Kernel, Windows Media Server)*

Development of a NAND Flash Memory-Based File System Supporting Transaction and Record Structure   *2006 – 2008, Hanyang University*
  ○ Developing a DBMS that directly manages a NAND flash memory via Linux MTD layer for storing EPG (Electronic Program Guide) data in a set-top box.
  ○ Developing a new PostgreSQL buffer manager that generates an optimized LBA sequence for NAND flash memory.
  ○ Developing a framework that generates a set of pre-defined I/O requests, measures latency, and identifies a FTL mapping algorithm of a NAND flash memory-based storage device.
  ○ *(Linux Kernel, Linux MTD, Wisconsin Storage System, PostgreSQL, ARM-based embedded board)*

Other Project Experience

A programming framework for processing-in-memory architecture   *2015 – 2016, Oak Ridge National Lab.*
  ○ Developing a high-level programming interface (C/C++) and a runtime environment that facilitates data analysis tasks with an array of PIM devices.
  ○ Developing a kernel device driver that emulates PIM devices in a NUMA architecture.
  ○ *(NUMA, Pthread, Intel TBB, Linux Kernel)*

Development of a web-based social marketing game   *2011, Pitapat Mobile, S. Korea*
  ○ Developing a social quiz/marketing web application using the Microsoft ASP.NET.
  ○ Integrating the web application to the Facebook app platform using the Facebook API.
  ○ *(C#, Javascript, JQuery, HTML, PHP, SQL Server, MySQL, Amazon EC2)*

Development of an image transfer library for a self-driving vehicle   *2008, Metabuild co., S. Korea*
  ○ Developing a userspace library that transfers requested pixel data from the CMOS camera to a lane-recognition module in a self-driving vehicle.
  ○ *(Linux, C, ARM-based embedded board)*

Development of an online survey website   *Feb 2010, Aug 2009, Dept. of English Education, Hanyang University, S. Korea*
  ○ Developing an online-survey web application including administrative tools.
  ○ Developing a web application that visualizes interactions among educators.
  ○ *(Linux, PHP, HTML, Javascript, MySQL, GD, Apache, Microsoft Excel)*

Construction of a protein function database   *2006 – 2008, Hanyang University*
  ○ Developed a GUI environment that allows medical researchers to verify protein interaction graphs that are automatically extracted from academic articles.
  ○ *(Windows, SQLServer, Delphi)*

Development of an online lesson website   *2002 – 2003, Golfschool co., S. Korea*

- Developed and maintained a commercial company website that serves online golf lessons and an online shopping mall of sporting goods.
- *(Linux, Apache, PHP, MySQL, HTML, Javascript)*

Publications

- Hyogi Sim, Sudharshan S. Vazhkudai, *Profiling the Usage of an Extreme-Scale Archival Storage System*, Proceedings of the 27th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOSTS '19), Rennes, France, October 2019
- Hyogi Sim, Arnab K. Paul, Eli Tilevich, Ali R. Butt, *CSLIM: Automated Extraction of IoT Functionalities from Legacy C Codebases*, Proceedings of the 8th International Workshop on Computing and Networking for IoT and Beyond (ComNet-IoT '19) in conjunction with ICDCN '19, Bangalore, India, pages 6, January 2019
- Ali Anwar, Yue Cheng, Hai Huang, Dongyoon Lee, Jingoo Han, Hyogi Sim, Fred Douglis, and Ali R. Butt, *BESPOKV: Application Tailored Scale-Out Key-Value Stores*, Proceedings of the 2018 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '18), Dallas, TX
- Hyogi Sim, Geoffroy R. Vallée, Youngjae Kim, Sudharshan S. Vazhkudai, Devesh Tiwari, Ali R. Butt, *An Analysis Workflow-Aware Storage System for Multi-Core Active Flash Arrays*, IEEE Transactions on Parallel and Distributed Systems (TPDS), Vol. 29, 2018
- Sangkuen Lee, Hyogi Sim, Sudharshan S. Vazhkudai, *A Programmable Shared-Memory System for an Array of Processing-In-Memory Devices*, Cluster Computing: The Journal of Networks, Software Tools and Applications, 2018
- Feiyi Wang, Hyogi Sim, Cameron Harr, Sarp Oral, *Diving into Petascale Production File Systems through Large Scale Profiling and Analysis*, In Proceedings of the 2nd Joint International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISC '17) in conjunction with SC, Denver, CO
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Geoffroy R. Vallée, Seung-Hwan Lim, Ali R. Butt, *TagIt: An Integrated Indexing and Search Service for File Systems*, Proceedings of the 2017 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '17), Denver, CO
- Seung-Hwan Lim, Hyogi Sim, Raghul Gunasekaran, Sudharshan S. Vazhkudai, *Scientific User Behavior and Data-Sharing Trends in a Petascale File System*, Proceedings of the 2017 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '17), Denver, CO
- Sangkuen Lee, Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, *AnalyzeThat: A Programmable Shared-Memory System for an Array of Processing-In-Memory Devices*, IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID '17), Madrid, Spain, May 2017
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Geoffroy R. Vallée, Seung-Hwan Lim, Ali R. Butt, *TagIt: An Integrated Search and Discovery Service for Extreme-Scale File Systems*, Poster in the 2016 USENIX Annual Technical Conference (ATC '16), Denver, CO, June 2016
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Devesh Tiwari, Ali Anwar, Ali R. Butt, Lavanya Ramakrishnan, *AnalyzeThis: An Analysis Workflow-Aware Storage System*, Proceedings of the 2015 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '15), Austin, TX
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Devesh Tiwari, Ali Anwar, Ali R. Butt, Lavanya Ramakrishnan, *AnalyzeThis: An Analysis Workflow-Aware Storage System*, Poster in the 2015 USENIX Annual Technical Conference (ATC '15), Santa Clara, CA, July 2015
- Hyogi Sim, Hoyoung Jung, Sungmin Park, Sooyong Kang, Jaehyuk Cha, *Identifying the FTL Mapping Scheme for USB Flash Devices*, The 4th International Conference on Convergence Technology and Information Convergence, CTIC 2009, Oct. 12
- Sooyong Kang, Sungmin Park, Hoyoung Jung, Hyogi Sim, Jaehyuk Cha, *Performance Tradeoffs in Using NVRAM Write Buffer for Flash Memory-based Storage Devices*, IEEE Transactions on Computers, Vol. 58, Issue 6 (Jun. 2009) Pages 744-758
- Hoyoung Jung, Hyogi Sim, Sungmin Park, Sooyong Kang, Jaehyuk Cha, *LRU-WSR: Integration of LRU and Writes Sequence Reordering for Flash Memory*, IEEE Transactions on Consumer Electronics, Volume 54, Issue 3 (Aug. 2008)
- Sungmin Park, Hoyoung Jung, Hyogi Sim, Sooyong Kang, Jaehyuk Cha, *Using Non-Volatile RAM as a Write Buffer for NAND Flash Memory-based Storage Devices*, 2008 IEEE International Symposium on Modeling, Analysis & Simulation of Computer & Telecommunication Systems, MASCOTS 2008, Sept 8-10, Baltimore, MD
- Sungmin Park, Hoyoung Jung, Hyogi Sim, Sooyong Kang, Jaehyuk Cha, *Write Buffer-aware Address Mapping for NAND Flash Memory Devices*, 2008 IEEE International Symposium on Modeling, Analysis & Simulation of Computer & Telecommunication Systems, MASCOTS 2008, Sept 8-10, Baltimore, MD
- Hoyoung Jung, Kyunghoon Yoon, Hyogi Sim, Sungmin Park, Sooyong Kang, Jaehyuk Cha, *LIRS-WSR: Integration of*

LIRS and Write Sequence Reordering for Flash Memory, The 2007 International Conference on Computational Science and Its Applications, ICCSA LNCS 2007, Aug. 29

Presentations

- Hyogi Sim, *Making a Peta-Scale Archival Storage System Searchable*, High Performance Storage Systems User Forum 2019 (HUF 2019), Indiana University, Bloomington, IN, October 2019
- Hyogi Sim, *Profiling the Usage of an Extreme-Scale Archival Storage System*, High Performance Storage Systems User Forum 2019 (HUF 2019), Indiana University, Bloomington, IN, October 2019

Recognitions

- Analysis Restaurant: In November 2015, *DEIXIS online magazine* featured AnalyzeThis storage system as a monthly highlight, based on the SC '15 paper (*https://deixismagazine.org/2015/11/analysis-restaurant/*).

Graduate-Level Courses

- Virginia Tech, VA: Statistics in Research, Multiprocessor Programming, Research Method in Computer Science, Advanced Parallel Computation, Advanced Topics in System and Network Security, Operating Systems, Software Refactoring
- Hanyang University, S. Korea: Advanced Operating System, Computer Algorithms, Database System Implementation, Real-Time Systems, Cryptography, Database Tuning

Teaching Experience

Computer Organization II, Virginia Tech, VA	*— Spring 2012, Fall 2011 (Teaching assistant)*
File Structure, Hanyang University, S. Korea	*— Spring 2009, Fall 2007 (Leading lab sessions)*
Data Structures, Hanyang University, S. Korea	*— Spring 2007 (Leading lab sessions)*
Object Oriented Programming, Hanyang University, S. Korea	*— Spring 2005 (Leading lab sessions)*

Awards & Scholarships

Graduate Research Assistantship, Virginia Tech, VA	*— Fall 2011, 2012, 2013, Spring 2012, 2014*
BK21 Scholarship, Hanyang Univeristy, S. Korea	*— 2009*
Academic Record Scholarship, Hanyang Univeristy, S. Korea	*— 2005*