HYOGI SIM

Oak Ridge National Laboratory, Virginia Tech
simh@ornl.gov, (540) 391-0202

CURRENT POSITION

Postmasters Research Assistant: Technology Integration Group, Oak Ridge National Laboratory
Ph.D. Student: Department of Computer Science, Virginia Tech

EDUCATION

Ph.D. in Computer Science, Virginia Tech, VA — *present*
 - Advisor: Dr. Ali R. Butt

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*
 - Advisor: Dr. Jaehyuk Cha and Dr. Sooyong Kang
 - 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*

SKILLS & EXPERTISE

Skillful in Linux application/system/kernel programming, parallel/distributed file systems, database systems.
 - **Programming Languages**: C *(proficient)*, C++, Python, JAVA, PHP, C#, BASH
 - **File & Storage**: Linux VFS, Device Mapper, FUSE, SCSI OSD-2, Linux TGT, blktrace, exofs, ext3
 - **Parallel & Distributed File Systems**: GlusterFS, Ceph, Lustre, 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
 - **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

CURRENT PROJECTS

Multi-Tiered Storage System in High Performance Computing — *2016-present, Oak Ridge National Laboratory*
 - Designing a software-defined storage system that orchestrates heterogeneous storage tiers based on user-specified workflow execution policies.
 - Designing and developing a NoSQL store-based global metadata manager that provides a unified namespace abstraction over heterogeneous storage tiers.
 - Developing a client-side shared library that allows applications to interact with the metadata manager.
   (FUSE, Lustre, Ceph, MySQL, HyperDex)

Scientific User Behavior Analysis from the Daily Parallel File System Snapshot — *2016-present, Oak Ridge National Laboratory*
 - Building an analysis framework using big data tools to analyze Spider II (Lustre PFS) daily snapshot data.
 - Analyzing user behaviors and interactions through the observed file system activities.
   (Spark, SparkSQL, Parquet, Python, Lustre)

Fast NVMe-Based Temporal Storage Framework for Data-Intensive Applications — *2016-present, Oak Ridge National Laboratory, Virginia Tech*
 - Designing a zero-copy, out-of-core framework to persist application in-memory objects.
 - Implementing a user-space layer that directly manages NVMe devices and bypasses the kernel file system.
 - Identifying a potential extension of the NVMe protocol for providing an object-based interface.
   (Linux, NVMe, NVMeDirect)

DEVELOPMENT EXPERIENCES ON FILE & STORAGE SYSTEMS

File System-Integrated Search and Discovery Services for High Performance Computing — *2015-2016, Oak Ridge National Laboratory and Virginia Tech*
 - Designed and developed a file system-integrated metadata indexing framework that supports a user-defined tagging in GlusterFS.

- Developed a command-line utility with the GlusterFS API to allow interactive, tagging-based file search queries in SQL syntax.
- Developed a computation offloading framework via a tagging-based file search, similar to 'find -exec'.
- Developed an automatic metadata extraction framework based on the tagging-based file search.
  (GlusterFS, SQLite, Linux)

Analysis-Aware Storage System for High Performance Computing — *2013-2015, Oak Ridge National Laboratory and Virginia Tech*
- Developed an active code execution framework on the Linux TGT implementation based on the extended SCSI T10 OSD-2 specification.
- Extended the OSD initiator driver and exofs in Linux Kernel to support the active processing.
- Designed and developed a FUSE file system that manages the array of active OSD devices.
- Integrated a workflow manager within the file system to schedule scientific workflow tasks across the array of active OSD devices.
- Developed a provenance management framework based on the OSD-2 object abstraction and the database integration in the FUSE file system.
  (Linux Kernel, SCSI T10 OSD-2 Protocol, Linux TGT, FUSE, SQLite)

Managing Multimedia Data for Content Servers with Hybrid Storage Architecture — *2007-2009, Database Lab., Hanyang University*
- Developed a content server log analyzer to analyze content popularity of a commercial media server.
- Developed a stripped-down file system based on ext2 for storage class memory devices, i.e., PRAM.
  (Linux Kernel, Windows Media Server)

Design of a NAND Flash Memory-Based File System Supporting Transaction and Record Structure — *2006-2008, Database Lab., Hanyang University*
- Developed a framework to identify FTL mapping schemes of NAND flash memory-based storage devices.
- Developed a DBMS that directly manages a raw NAND flash memory chip for managing EPG (Electronic Program Guide) data in a set top box.
- Developed a buffer management policy in PostgreSQL for NAND flash memory-based storage devices.
  (Linux Kernel, Linux MTD, Wisconsin Storage System, ARM-based embedded board, PostgreSQL)

OTHER DEVELOPMENT EXPERIENCES

Shared-Memory Programming Framework for Processing-In-Memory Architecture — *2015-2016, Oak Ridge National Laboratory*
- Developed a high-level programming framework that facilitates the data analysis using PIM devices.
- Developed a runtime system that optimizes the data placement across an array of PIM devices.
- Developed a device driver that emulates PIM devices in a NUMA architecture.
  (NUMA, Pthread, Intel TBB, Linux Kernel)

Pitapat Mobile, S. Korea: *Senior Developer, Startup Member — Jan-June 2011*
- Developed a social marketing web application and a social quiz game for desktop and mobile devices.
- Integrated the application to the Facebook app platform using the Facebook API.
  (C#, Javascript, JQuery, HTML, PHP, SQL Server, MySQL, Amazon EC2)

Metabuild co., S. Korea: *Part-Time Programmer — Dec 2008*
- Developed a user-space module that delivers requested pixel data from the CMOS camera to the lane-recognition module in a self-driving vehicle.
  (Linux, C, ARM-based embedded board)

Dept. of English Education, Hanyang University: *Part-Time Programmer — Feb 2010, Aug 2009*
- Developed an online-survey web application including administrative tools.
- Developed a web application that visualizes interactions among educators.
  (Linux, PHP, HTML, Javascript, MySQL, GD, Apache)

Construction of a Protein Function Database — *2006-2008, Database Lab., Hanyang University*
- Developed a GUI environment that verifies protein interaction graphs extracted from academic documents.
  (Windows, SQLServer, Delphi)

Golfschool co., S. Korea: *Programmer, Startup Member — 2002-2003*
- Developed and maintained a commercial website that serves online lectures and sells sporting goods.
  (Linux, Apache, PHP, MySQL, HTML, Javascript)

WORK HISTORY

Oak Ridge National Laboratory: *Postmasters Research Assistant — Mar 2015-present*
Oak Ridge National Laboratory: *Summer Intern — Jan-Aug 2013*
Virginia Tech: *Research and Teaching Assistant — Aug 2011-Dec 2013, Step 2013-Feb 2015*
Pitapat Mobile, S. Korea: *Senior Developer, Startup Member — Jan-June 2011*
Dept. of English Education, Hanyang University, S. Korea: *Part-Time Programmer — Feb 2010, Aug 2009*
Metabuild co., S. Korea: *Part-Time Programmer — Dec 2008*
Golfschool co., S. Korea: *Programmer, Startup Member — 2002-2003*
Military Conscription in a Combat Police Force, S. Korea *— May 1999 - July 2001*

RESEARCH PUBLICATIONS

- **Hyogi Sim**, Youngjae Kim, Sudharshan S. Vazhkudai, Geoffroy R. Vallée, Seung-Hwan Lim, Ali R. Butt, *A Fast and Efficient Scientific Data Discovery Service — currently under review*
- **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

GRADUATE-LEVEL COURSES

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

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/*).

AWARDS & SCHOLARSHIPS

Virginia Tech: Graduate Research Assistantship *— Fall 2011, 2012, 2013, Spring 2012, 2014*
Hanyang University: BK21 Scholarship *— 2009*, Academic Record Scholarship *— 2005*