HYOGI SIM

Oak Ridge National Laboratory, Virginia Tech
hyogi@vt.edu, (540) 391-0202

Objective: To obtain a research and development position that will both utilize and enrich my current skills.

CURRENT POSITION

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

RESEARCH INTERESTS

- High Performance Computing, Large-Scale Data Management, Active Storage, Parallel Processing
- File and Storage Systems, Distributed Systems, Operating Systems, Database Management Systems
- Non-Volatile Memory, Flash Memory

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*

CURRENT RESEARCH 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)

Analysis of User Behaviors through the Daily Parallel File System Snapshot — *2016-present, Oak Ridge National Laboratory*
- Building an analysis framework based on Spider II (Lustre PFS) daily snapshot data.
- Analyzing user behaviors and interactions through the observed file system activities.
  (Spark, SparkSQL, Python, Lustre)

Developing a Fast NVMe-Based Temporal Storage Framework for Data-Intensive Applications — *2016-present, Virginia Tech*
- Designing a zero-copy, out-of-core application object serialization framework.
- Implementing a user-space layer that directly manages NVMe devices and bypasses the kernel file system.
  (Linux, NVMe)

PAST RESEARCH EXPERIENCE

Shared-Memory Programming Framework for Processing-In-Memory Architecture — *2015-2016, Oak Ridge National Laboratory*
- Developed a high-level data structure and a programming framework that facilitate the data analysis using PIM devices.
- Developed a runtime system that optimizes the data placement among PIM devices.
- Developed a device driver which emulates PIM devices in a NUMA architecture.
  (NUMA, Pthread, Linux Kernel)

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

- Developed a storage-side data reduction and automatic metadata extraction framework based on the tagging-based file search.
- Implemented the prototype framework in GlusterFS and command-line utilities.
  (GlusterFS, SQLite, Linux)

Exploiting Active Flash in High Performance Computing Storage — *2013-2015, Virginia Tech and Oak Ridge National Laboratory*

- Developed an active storage target framework based on SCSI T10 OSD-2 specification.
- Enhanced OSD initiator driver and exofs in Linux to support the active processing.
- Designed and developed a FUSE file system for the active OSD devices.
- Developed a workflow scheduler within FUSE to orchestrate scientific workflow tasks across an array of active OSD devices.
- Developed provenance management via SQLite within the file system.
  (Linux Kernel, SCSI T10 OSD-2 Protocol, FUSE, SQLite)

Managing Multimedia Data for Content Servers with Hybrid Storage — *2007-2009, Database Lab., Hanyang University*

- Developed a content server log analyzer to analyze content popularity.
- Developed a prototype Linux file system for storage class memory device.
  (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 storage devices.
- Developed a DBMS that directly manages the raw NAND Flash memory device.
  (Linux, Wisconsin Storage System, ARM-based embedded board)

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)

WORK EXPERIENCE

Oak Ridge National Laboratory — *Postmasters, Mar 2015-present*

Oak Ridge National Laboratory — *Research Intern, Jan-Aug 2013*

- Designed and developed a node local SSD-based framework for processing near storage in a HPC environment. (Linux Kernel, SCSI T10 OSD-2 Protocol, FUSE, SQLite)

Pitapat Mobile, S. Korea — *Jan-June 2011*

- Developed a social marketing web application and a social quiz game running on Facebook. (C#, Javascript, JQuery, HTML, PHP, SQL Server, MySQL, Amazon EC2)

Dept. of English Education, Hanyang University — *Feb 2010, Aug 2009*

- Developed an online-survey web application. (Linux, PHP, MySQL, Apache)
- Developed a web-based application that visualizes interactions among educators. (Linux, PHP, GD, Apache)

Metabuild co., S. Korea — *Dec 2008*

- Developed a user-space module that delivers requested pixel data from the CMOS camera to the lane-recognition application for self-driving vehicles. (Linux, C, ARM-based embedded board)

Military Conscription in a Combat Police Force, S. Korea — *May 1999 - July 2001*

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

TEACHING EXPERIENCE

Computer Organization II — *Spring 2012, Fall 2011 (Teaching assistant)*
- Graduate Teaching Assistant, Dept. of Computer Science, Virginia Tech

File Structure — *Spring 2009, Fall 2007 (Leading lab sessions)*
- Graduate Teaching Assistant, Division of Computer Science and Engineering, Hanyang University

Data Structures — *Spring 2007 (Leading lab sessions)*
- Graduate Teaching Assistant, Division of Computer Science and Engineering, Hanyang University

Object Oriented Programming — *Spring 2005 (Leading lab sessions)*
- Graduate Teaching Assistant, Dept. of Computer Science Education, Hanyang University

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*

TECHNICAL SKILLS

Skillful in Linux application/system/kernel programming.
- **Languages:** C *(proficient)*, C++, Python, BASH, PHP, JAVA, C#, Javascript
- **File & Storage:** Linux VFS, Device Mapper, FUSE, SCSI, iSCSI, OSD, exoFS, GlusterFS
- **SSD:** NAND Flash Memory, Storage Class Memory, Flash Translation Layer
- **Databases:** PostgreSQL, MySQL, MS SQLServer, IBM DB2, SQLite, Redis, HyperDex
- **Tools:** gcc, gdb, vim, Glade, Eclipse, Visual Studio, LaTex, gnuplot