

# **HYOGI SIM**

Oak Ridge National Laboratory  
*simh@ornl.gov, (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 2019*

- 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**

**Oak Ridge National Laboratory:** *HPC Systems Engineer* — *Feb 2017-present*

**Oak Ridge National Laboratory:** *Postmasters Research Assistant* — *Mar 2015-Jan 2017*

**Oak Ridge National Laboratory:** *Summer Intern* — *Jan-Aug 2013*

**Virginia Tech:** *Research and Teaching Assistant* — *Aug 2011-Dec 2013, Sep 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*

## **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 EXPERIENCES ON FILE & STORAGE SYSTEMS**

**Scientific User Behavior Analysis from the HPC File System Snapshot** — *2016-2017, Oak Ridge National Lab.*

- Building an analysis framework using big data tools to analyze the daily file system snapshots of Spider II, a 32 PB centralized Lustre parallel file system.
- Identifying scientific user behavior and patterns through the file system snapshot analysis.  
(Spark, SparkSQL, Parquet, Python, Lustre)

**File System-Integrated Search and Discovery Services for High Performance Computing** — *2015-2016, Oak Ridge National Lab. and Virginia Tech*

- Designed and developed a file system-integrated metadata indexing framework that supports the user-defined tagging in GlusterFS.
- Developed a command-line utility using the GlusterFS API to allow interactive, tagging-based file search queries in SQL syntax.
- Developed a computation offloading framework via the 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 Lab. and Virginia Tech

- Designed an active execution framework based on SCSI T10 OSD-2 specification and implemented the extended OSD-2 protocol on the Linux TGT.
- Extended the host-side OSD initiator driver and exofs in Linux Kernel to support the active execution framework.
- Designed and developed a FUSE file system that manages the array of active OSD devices.
- Integrated a workflow manager within the file system to support scientific workflow processing across the array of active OSD devices.
- Developed a provenance management framework based on the OSD-2 object abstraction using a light-weight database within the FUSE file system  
(Linux Kernel, SCSI T10 OSD-2 Protocol, Linux TGT, FUSE, SQLite)

**Hierarchical Data Management in Media Servers with Hybrid Storage Architecture** — 2007-2009, Database Lab., Hanyang University

- Developed a content popularity analyzer that periodically analyzes request log files and ranks frequently requested contents in a commercial media server running Windows Media Server.
- Developed a light-weight file system based on ext2 for the caching servers equipped with storage class memory devices, e.g., PRAM.  
(Linux Kernel, Windows Media Server)

**Development of a NAND Flash Memory-Based File System Supporting Transaction and Record Structure** — 2006-2008, Database Lab., Hanyang University

- Developed a DBMS that integrates FTL and directly manages a raw NAND flash memory chip via Linux MTD layer for managing EPG (Electronic Program Guide) data in a set top box.
- Developed a flash-aware buffer management policy in PostgreSQL to generates a flash-friendly LBA sequence for NAND flash memory.
- Developed 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 EXPERIENCES

**A Programming Framework for Processing-In-Memory Architecture** — 2015-2016, Oak Ridge National Lab.

- Developed a high-level programming interface (C/C++) and a runtime environment that facilitates data analysis tasks with an array of PIM devices.
- Developed 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** — Jan-June 2011, Pitapat Mobile, S. Korea

- Developed a social quiz/marketing web application using the Microsoft ASP.NET.
- Integrated 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** — Dec 2008, Metabuild co., S. Korea

- Developed 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

- 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, Microsoft Excel)

**Construction of a Protein Function Database** — 2006-2008, Database Lab., 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)

## RESEARCH PUBLICATIONS

- **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 — to appear
- Ali Anwar, Yue Cheng, Hai Huang, Dongyoon Lee, Jingoo Han, **Hyogi Sim**, Fred Douglass, 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**, Raghu 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

## 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:** 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:** Advanced Operating System, Computer Algorithms, Database System Implementation, Real-Time Systems, Cryptography, Database Tuning

## 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

## AWARDS & SCHOLARSHIPS

**Virginia Tech:** Graduate Research Assistantship — *Fall 2011, 2012, 2013, Spring 2012, 2014*

**Hanyang University:** BK21 Scholarship — *2009*, Academic Record Scholarship — *2005*