HYOGI SIM

Phone: +1-865-574-6167
simh@ornl.gov

1 Bethel Valley Road, MS6008
Oak Ridge, TN 37831

CURRENT POSITION

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

EDUCATION

PhD in Computer Science, Virginia Tech, VA — *expected in 2020*
Advisor: Dr. Ali R. Butt

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

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

BS 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 — *Aug 2011 – Dec 2013, Sep 2013 – Feb 2015*
- Senior Developer, Pitapat Mobile, S. Korea — *Jan 2011 – June 2011*
- Senior Developer, Golf School co., S. Korea — *2002 – 2003*
- Military Conscription in Combat Police Force, S. Korea — *May 1999 – Jul 2001*

SKILLS & EXPERTISE

Skillful in Linux system/kernel programming, distributed systems, file systems, database systems

- Programming Languages: C, C++, Python, Java, PHP, C#
- File & Storage Systems: Linux VFS, FUSE, Device Mapper, SCSI OSD-2
- Parallel & Distributed File Systems: GlusterFS, Lustre, Ceph, GPFS, NFS, HDFS, HPSS
- Non-Volatile Memory: Linux MTD, Flash Translation Layer
- Databases: MySQL, PostgreSQL, SQLite, MS SQL Server, 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, glide, visual studio, latex, gnuplot
- Web Development: HTML, CSS, PHP, Javascript, JQuery, ASP.NET
- Package Management: RPM, Portage (Gentoo), pkgsrc (NetBSD)

PROJECT EXPERIENCE IN FILE & STORAGE SYSTEMS

A metadata indexing framework for HPC archives — *2018 – present, Oak Ridge National Lab.*

- Developing a software that extracts metadata from the HPSS archive and builds a centralized data catalog
- Developing a framework that extracts metadata from document and self-describing scientific files
- Developing command-line utilities for users and REST API for middlewares
- *(HPSS, IBM DB2, PostgreSQL, Apache Tika, XDR, cURL, SQLite)*

UnifyFS: An ephemeral burst buffer file system for HPC *2017 – present, Oak Ridge National Lab.*

- Designing and developing an ephemeral distributed file system for node-local burst buffers to facilitate checkpointing of HPC scientific applications
- Developing a file system metadata management framework using a KV store
- *(FUSE, MPI, LevelDB, Mercury, Margo, Argobots, HDF5)*

Large-scale file system snapshot and workload analysis *2016 – 2019, Oak Ridge National Lab.*

- Analyzed the daily snapshot data of the Spider II file system, the world's largest Lustre deployment, in OLCF using a big data analysis framework
- Analyzed the eight-year worth data transfer workloads of 80 PB HPSS archival storage system in OLCF
- *(Lustre, HPSS, Spark, SparSQL, Parquet, Python, MySQL, SQLite)*

TagIt: File system-integrated search and discovery service for HPC
2015 – 2016, Oak Ridge National Lab.

- Designed and developed a file system integrated metadata indexing framework that supports user-defined taggings in GlusterFS and CephFS
- Developed an active execution framework associated with the metadata indexing framework
- *(GlusterFS, CephFS, Linux, SQLite)*

AnalyzeThis: An analysis workflow-aware storage system
2013 – 2015, Oak Ridge National Lab./Virginia Tech

- Designed an active execution framework by extending SCSI T10 OSD-2 protocol
- Extended the exofs OSD file system to support the extended OSD-2 protocol
- Developed a FUSE file system that manages an array of active OSD devices
- *(Linux Kernel, SCSI T10 OSD-2, Linux tgt, FUSE, SQLite)*

Hierarchical data management in media servers *2007 – 2009, Hanyang University*

- Developed an online content popularity analyzer
- Developed a light-weight file system for storage class memory devices (PRAM) in media cache servers
- *(Linux Kernel, Windows Media Server)*

Flash memory-based transactional file system *2006 – 2008, Hanyang University*

- Developed a relational database that directly manages a NAND flash memory for storing EPG (Electronic Program Guide) data in a TV set-top box
- Developed a NAND flash memory-aware buffer manager in PostgreSQL
- Developed a software framework that identifies an FTL mapping algorithm of a USB flash drives
- *(Linux MTD, Wisconsin Storage System, PostgreSQL, ARM-based embedded board)*

OTHER PROJECT EXPERIENCES

A programming framework for PIM architecture *2015- 2016, Oak Ridge National Lab.*

- Developed a high-level programming interface (C/C++) and a runtime environment for Processing-In-Memory devices.
- Developed a Linux device driver that emulates PIM devices in a NUMA architecture
- *(NUMA, pthread, Intel TBB, Linux)*

Development of web-based social marketing game *2011, Pitapat Mobile*

- Developed a social quiz/marketing web application using the 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 self-driving vehicles *2008, Metabuild co.*

- Developed a 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, Hanyang University*

- 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, SQL Server, Delphi)*

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

- Developed and maintained a commercial company website that served 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. Valle´e, 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. Valle´e, 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. Valle´e, 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, *Extracting Metadata from the ORNL HPSS Archive to Improve its Usability*, Knowledge is Power: Unleashing the Potential of Your Archives through Metadata, Bird of a Feather (BoF) Session in 2019 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '19), Denver, CO, Nov 2019
- 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(https://deixismagazine.org/2015/11/analysis-restaurant/): In November 2015, DEIXIS online magazine featured AnalyzeThis as a monthly highlight, based on the SC '15 paper.

PROFESSIONAL SERVICES

Program Committees

- International Symposium on Benchmarking, Measuring and Optimizing (Bench) *2019, 2018*
- Workshop on Big Scientific Data Benchmarks, Architecture, and Systems (SDBA) *2019*

Reviews

- Bill Gates Foundation, Alzheimer's Data Initiatives RFP review *2019*
- International Conferences on Computer Science and Application Engineering (CSAE) *2019*
- CORAL (Collaboration of Oak Ridge, Argonne, and Livermore) 2 RFP review *2018*
- Journal of Parallel and Distributed Computing (JPDC) *2018*
- IEEE Transactions on Parallel and Distributed Systems (TPDS) *2018*

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 *Spring 2012, Fall 2011 (Teaching Assistant)*
- File Structure, Hanyang University *Spring 2009, Fall 2007 (Leading lab sessions)*
- Data Structures, Hanyang University *Spring 2007 (Leading lab sessions)*
- Object-Oriented Programming, Hanyang University *Spring 2005 (Leading lab sessions)*

AWARDS & SCHOLARSHIPS

- Graduate Research Assistantship, Virginia Tech *2011 – 2014*
- BK21 Scholarship, Hanyang University *2009*
- Excellent Academic Record Scholarship, Hanyang University *2005*