Hee Won Lee

36 Liberty Ridge Rd, Basking Ridge, NJ 07920, USA
knowpd@gmail.com • +1 (919) 800-8993

SUMMARY

Infrastructure and cloud systems expert with 20+ years of experience designing, scaling, and optimizing distributed systems across public, private, and hybrid cloud environments. Former Principal Engineer at Samsung Electronics, where I led a 30+ engineer team operating cloud platforms supporting over 10,000 virtual machines. Specialized in high-performance infrastructure using Kubernetes, Ceph, and other open-source technologies. At AT&T Labs, developed resilient, cost-efficient storage and compute systems for carrier-grade environments. Ph.D. in Computer Science with deep expertise in system-level virtualization (QEMU-KVM), networking, and performance engineering. Proven track record of building foundational infrastructure, mentoring engineering teams, and delivering enterprise-scale innovation.

HIGHLIGHED PUBLICATIONS

MIST: Mitigating Host-Side Interference for Storage Traffic in Virtualized Data Centers.
Hee Won Lee and Moo-ryong Ra
In *IEEE 9th International Conference on Cloud Computing (CLOUD 2016)*.

Fighting with Unknowns: Estimating the Performance of Scalable Distributed Storage Systems with Minimal Measurement Data.
Moo-ryong Ra and Hee Won Lee.
In *IEEE 35th Symposium on Mass Storage Systems and Technologies (MSST 2019)*.

Integrated Simulation and Emulation using Adaptive Time Dilation.
Hee Won Lee, David Thuente, and Mihail L. Sichitiu.
In *ACM 2nd SIGSIM Conference on Principles of Advanced Discrete Simulation (PADS 2014)*.

EXPERIENCE

Samsung Electronics, Hwaseong, South Korea

- Principal Engineer   Dec 2020 – Feb 2024
  - Infrastructure Engineering
    - Designed and implemented high-performance virtual machines by applying CPU pinning and SR-IOV PCI passthrough techniques.
    - Developed real-time monitoring and telemetry systems using Telegraf/InfluxDB/Grafana
    - Designed and implemented auto-provisioning of Azure resources using Bicep (similar to Terraform)
    - Integrated Azure AI Search into an on-premises chat service built on LangChain, leveraging a large language model with retrieval-augmented generation (RAG).
  - Leadership
    - As Head of the Cloud Platform Group, I led a team of 30+ cloud engineers, managing the operations of the in-house cloud platform (for 10,000+ VMs), and enterprise-wide Active Directory services (for 60,000+ employees).

AT&T Labs Research, Bedminster, New Jersey, USA

- Principal Inventive Scientist   Mar 2015 – Jul 2020
  - Compute
    - Designed and implemented Redis High Availability with Kubernetes in CI/CD environment
    - Integrated Redis HA helm charts into Radio Access Network Intelligent Controller release 4 (field trial version)
    - Evaluated the performance of Redis with Intel Optane Persistent Memory
    - Analyzed cost benefits by applying Intel Optane Persistent Memory to AT&T's Content Delivery Network (CDN)
    - Evaluated and compared the performance of *Redis* vs *Aerospike* vs *Cassandra* using the YCSB benchmark tool
  - Storage
    - Designed and implemented I/O traffic protection mechanisms for cloud storage
    - Designed and implemented containerized Ceph with Kubernetes helm charts
    - Performed extensive resiliency tests for containerized Ceph storage with Kubernetes
    - Developed best recovery practices from various failure scenarios for containerized Ceph storage with Kubernetes
    - Evaluated cache performance across host and VM page cache, dm-cache, and Ceph RBD cache
    - Set up a testbed for NVMe over Fabrics with CPU offloading on RoCE-based lossless networks, and then evaluated the performance and verified zero CPU usage on the target host machine
  - Networking
    - Designed and implemented a system that controls network bandwidth/delay between multi-site OpenStack clusters

North Carolina State University, Raleigh, North Carolina, USA

- Research Assistant   Jan 2010 – Dec 2014

- Designed and implemented a virtualization platform for evaluating distributed applications running on diverse OSs (Linux, FreeBSD, Windows, JunOS)
- Redesigned and customized the QEMU-KVM hypervisor to synchronize distributed VMs with virtual time

Korea Telecom, Daejeon, South Korea

- Software Engineer   Jan 2002 – Jul 2009
  - Capacity Planning for KT Backbone Network and IPTV Services
    - Built a virtual network environment of KT backbone networks using OPNET simulation tool
    - Performed link capacity planning for IPTV services
  - VoIP Network Management System
    - Conducted requirement management and then designed architecture of VoIP-NMS
    - Deployed VoIP-NMS to production
  - IP Core Network Management System
    - Designed and implemented a system component that collects network equipment information from MIB (Management Information Base) using SNMP (Small Network Management Protocol)
    - Developed SQL scripts for Oracle database
    - Developed a system component that exchanges information with Service Assurance System using XML-RPC
    - Set up an MPLS testbed using Cisco Routers 7204/7500, Juniper Router M5 and Ixia Traffic Generator

CERTIFICATIONS

- **Microsoft Azure Solutions Architect Expert**
  - Microsoft Certified: Azure Administrator Associate (AZ-104)
  - Microsoft Certified: Azure Solutions Architect Expert (AZ-305)

- **Microsoft Cybersecurity Architect Expert**
  - Microsoft Certified: Azure Security Engineer Associate (AZ-500)
  - Microsoft Certified: Cybersecurity Architect Expert (SC-100)

EDUCATION

North Carolina State University, Raleigh, North Carolina, USA

- Ph.D. in Computer Science   Aug 2009 – May 2015
  - Thesis: Network Emulation with Adaptive Time Dilation
  - Adviser: Prof. Mihail L. Sichitiu, Co-adviser: Prof. David Thuente
  - Concentration: Virtualization, Networked Systems, Distributed Systems

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

- M.E. in Software Engineering   Aug 2004 – Aug 2005

Korea University, Seoul, South Korea

- B.E. in Electrical Engineering   Mar 1995 – Feb 2002

SOFTWARE DEVELOPMENT EXPERIENCES

Virtualization

- Hypervisor (QEMU-KVM) [C++]
  - Designed and implemented a spinlock on shared memory to synchronize distributed VMs (processes created by QEMU-KVM hypervisor) and simulation nodes (processes created by NS-3) with virtual time
  - Implemented a virtual time synchronization daemon with TCP/UDP sockets using Boost Asio Library

- High-Performance Emulation System [C++, Python]
  - Emulated higher performance with a time dilation technique using unmodified OSs (Linux, FreeBSD, Windows, and Junos) and real application workloads (VLC media player)
  - Modified the QEMU-KVM hypervisor for virtual time

- Hybrid Simulation and Emulation System [C++, Python]
  - Built an integrated simulation (NS-3) and emulation (QEMU-KVM) framework in a distributed environment

Storage

- Cloud Storage
  - Designed a system architecture for making a commodity SSD array a usable iSCSI block storage backend for OpenStack clusters
  - Designed and implemented the backend of OpenStack Cinder for dynamic configuration and QoS support
  - Designed and implemented containerized Ceph with Kubernetes helm charts with an one-osd-per-pod approach

- Persistent Memory
  - Set up a testbed with Intel Optane Persistent Memory (in collaboration with Intel team)
  - Analyzed the performance characteristics of Intel Optane Persistent Memory
  - Measured the latency of Redis with Intel Optane Persistent Memory using the *memtier_benchmark* tool

- I/O Traffic Protection [Python]

- Designed and implemented an I/O bandwidth reservation algorithm that guarantees the minimum I/O bandwidth of a virtual host connected to a shared storage volume through iSCSI
- Designed a NUMA-aware CPU pinning algorithm that protects I/O traffic from CPU interference
- Evaluated the system performance for diverse types of I/O pattern using FIO and Filebench
  - RDMA over Converged Ethernet (RoCE)
    - Set up a testbed for lossless RoCE using PFC-capable Mellanox Spectrum Switch and Dell servers equipped with ConnectX-5 NICs
    - Evaluated the performance of NVMe over Fabric for zero CPU usage on target host machine

Networking

- Network Traffic Controller [Ansible]
  - Designed and implemented a system that controls network bandwidth/delay between OpenStack clusters using Tc/NetEm

- Network Link Emulation [C++, Python]
  - Interconnected distributed VMs through virtual links using Tc/NetEm with adaptive time dilation

- Software-Defined Networking [Python]
  - Used OpenFlow POX controller and Open vSwitch to migrate VirtualBox-based virtual networks

- Routing Algorithms [Java]
  - Implemented modified Dijkstra's algorithm with negative weights
  - Implemented vertex/edge disjoint paths and elementary circuits search algorithms
  - Implemented CPLEX code with Integer Linear Programming for Routing and Wavelength Assignment

- Queueing Systems [Java]
  - Implemented discrete event simulation models for queuing systems (M/M/1, M/M/m, M/G/m)

- Wireless Networking [C++]
  - Designed and implemented an OMNeT++ model for IVG (Inter-Vehicle Geocast) routing protocol in vehicle-to-vehicle ad-hoc network simulation
  - Connected KVM-based virtual nodes through a wireless simulation model of OMNeT++

- Peer-to-Peer File Sharing System [Java]
  - Mapped a file signature (created by encoding text document with m-bit string) onto a hash space
  - Used Apache Mina, Apache FtpServer, and JavaDB

CPU & Memory

- CPU/Memory Control [Python]
  - Designed and implemented a CPU/Memory control algorithm that enforces CPU/Memory ceiling and allocates a relative share of CPU time using Linux Control Groups (Cgroups)

- Multi-Threading [C++]
  - Created a thread for searching text file paths and multiple worker threads for counting words using Boost Filesystem and Thread Libraries
  - Implemented multiple producer, multiple consumer thread-safe queue and map

- Prioritized Preemptive Schedulers [C]
  - Implemented a prioritized preemptive scheduler using POSIX thread library
  - Designed and implemented a priority-based scheduling algorithm using XINU kernel (a small UNIX OS)

- Demand Paging [C]
  - Implemented a demand paging system that allows for more address space than physically available one (XINU)

PUBLICATIONS   **CONFERENCES**

[9] **Dejavu: Reinforcement Learning-based Cloud Scheduling with Demo and Competition**.
Seonwoo Kim, Yoonsung Nam, Minwoo Park, <u>Hee Won Lee</u>, Seyeon Kim, Sangtae Ha.
In *IEEE 21st International Conference on Mobile Ad-Hoc and Smart Systems (MASS 2024)*.
Seoul, South Korea, Sep 2024.

[8] **TIPS: Making Volatile Index Structures Persistent with DRAM-NVMM Tiering**.
Madhava Krishnan, Wook-Hee Kim, Xinwei Fu, Sumit Kumar Monga, <u>Hee Won Lee</u>, Minsung Jang, Ajit Mathew, and Changwoo Min.
In *USENIX Annual Technical Conference (ATC 2021)*.
Virtual Event, Jul 2021.

[7] **eMRC: Efficient Miss Ratio Approximation for Multi-Tier Caching**.
Zhang Liu, <u>Hee Won Lee</u>, Yu Xiang, Dirk Grunwald, and Sangtae Ha.
In *USENIX 19th Conference on File and Storage Technologies (FAST 2021)*.
Virtual Event, Feb 2021.

[6] **EF-Dedup: Enabling Collaborative Data Deduplication at the Network Edge**.
Shijing Li, Tian Lan, Bharath Balasubramanian, Moo-ryong Ra, <u>Hee Won Lee</u>, and Panta Rajesh.
In *IEEE 39th International Conference on Distributed Computing Systems (ICDCS 2019)*.
Dallas, TX, USA, Jul 2019.

[5] **Fighting with Unknowns: Estimating the Performance of Scalable Distributed Storage Systems with Minimal Measurement Data**.
Moo-ryong Ra and <u>Hee Won Lee</u>.
In *IEEE 35th Symposium on Mass Storage Systems and Technologies (MSST 2019)*.
Santa Clara, CA, USA, May 2019.

[4] **MIST: Mitigating Host-Side Interference for Storage Traffic in Virtualized Data Centers**.
<u>Hee Won Lee</u> and Moo-ryong Ra
In *IEEE 9th International Conference on Cloud Computing (CLOUD 2016)*.
San Francisco, CA, USA, Jun 2016.

[3] **Integrated Simulation and Emulation using Adaptive Time Dilation**.
<u>Hee Won Lee</u>, David Thuente, and Mihail L. Sichitiu.
In *ACM 2nd SIGSIM Conference on Principles of Advanced Discrete Simulation (PADS 2014)*.
Santa Clara, CA, USA, May 2014.

[2] **Resource Management System for Next Generation Services**.
Seung-Hee Han, Bom-Su Kim, Chan-Kyou Hwang, <u>Hee Won Lee</u>, Byung-deok Chung.
In *IEEE International Conference on Advanced Technologies for Communications 2019*.
Hai Phong, Vietnam, Oct 2009.

[1] **Reusability Enhancement by Using Flexible Topology Architecture for Network Management System**.
<u>Hee Won Lee</u>, Chan Kyou Hwang, Jae-Hyoung Yoo, Ho-Jin Choi, Sungwon Kang, and Dan H. Lee.
In *IEEE/ACIS 7th International Conference on Computer and Information Science (ICIS 2008)*.
Portland, Oregon, USA, May 2008.

JOURNALS

[3] **FPV Video Adaptation for UAV Collision Avoidance**.
Simran Singh, <u>Hee Won Lee</u>, Tuyen X. Tran, Yu Zhou, Mihail L. Sichitiu, Ismail Güvenç, and Arupjyoti Bhuyan.
IEEE Open Journal of the Communications Society, vol. 2, pp. 2095–2110.
Aug 2021.

[2] **Network Link Emulation With Adaptive Time Dilation**.
<u>Hee Won Lee</u>, Mihail L. Sichitiu, and David Thuente.
Journal of Parallel and Distributed Computing, vol. 104, pp. 88–98.
Jun 2017.

[1] **High-performance Emulation of Heterogeneous Systems using Adaptive Time Dilation**.
<u>Hee Won Lee</u>, Mihail L. Sichitiu, and David Thuente.
International Journal of High Performance Computing Applications, vol. 29, issue 2.
May 2015.

WORKSHOPS, POSTERS AND TECHNICAL REPORTS

[3] **Accelerating Applications in the Fast-Moving Devices with Proactive Provisioning (Poster)**.
HyunJong Lee, <u>Hee Won Lee</u>, Moo-Ryong Ra, Yu Xiang, and Jason Flinn.
In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys 2019)*.
Seoul, South Korea, Jun 2019.

[2] **IOArbiter: Dynamic Provisioning of Backend Block Storage in the Cloud**.
Moo-Ryong Ra and <u>Hee Won Lee</u>.
arXiv preprint arXiv:1904.09984.
Apr 2019.

[1] **End-User IPTV Traffic Measurement of Residential Broadband Access Networks**.
Young J. Won, Mi-Jung Choi, Byung-Chul Park, James W. Hong, <u>Hee Won Lee</u>, Chan Kyu Hwang, and Jae-Hyoung Yoo.
In *IEEE Network Operations and Management Symposium (NOMS) Workshops*.
Apr 2008.

PATENTS

[1] **Dynamic Provisioning of Storage in the Cloud**.
Moo-ryong Ra and <u>Hee Won Lee</u>.
US 10,530,703 B2.
Date of Patent: January 7, 2020.

RESEARCH MENTORING

- **Simran Singh**, North Carolina State University
  - Project: UAV real-time video applications with cellular infrastructure support
  - 2019 fall internship at AT&T

- **Madhava Krishnan**, Virginia Tech
  - Project: A general purpose in-memory key-value store architecture for emerging persistent memory
  - 2019 summer internship at AT&T

- **HyunJong (Joseph) Lee**, University of Michigan–Ann Arbor
  - Project: Predictive migration of edge computations
  - 2018 summer internship at AT&T and 2021 winter internship at Samsung

- **Zhang Liu**, University of Colorado–Boulder
  - Project: Multi-tier cache orchestration with miss ratio curve
  - 2017 summer internship at AT&T

HONORS & AWARDS

- **Outstanding Teaching Assistant**, North Carolina State University   Apr 2011
  - Awarded the day of 29 April 2011
    - For excellent service as a teaching assistant during the 2010-11 calender year, the Department of Computer Science recognizes Hee Won Lee as an Outstanding Teaching Assistant
  - Teaching Assistent Courses
    - Algorithm (2011)
    - Operating System (2010, 2012, 2013)
    - Computer Networks (2011, 2012)
    - Software Engineering (2010)

- **Army Commendation Medal**, U.S. Army   Feb 2001
  - Awarded an ARCOM (Army Commendation Medal) for dedications and outstanding leadership in U.S. Army

REFERENCES

- **Prof. Mihail L. Sichitiu**
  (PhD Program Advisor)
  Deaprtment of Electrical and Computer Engineering, North Carolina State University
  890 Oval Drive, 3114 Engineering Building II, Raleigh, NC 27606, USA
  mlsichit@ncsu.edu • +1 (919) 515-7348

- **Prof. Sangtae Ha**
  (Research Collaborator)
  Department of Computer Science, University of Colorado Boulder
  ECCR 1B14, 1045 Regent Drive 430 UCB, Boulder CO, 80309, USA
  sangtae.ha@colorado.edu • +1 (303) 492-7031

- **Dr. Yu Xiang**
  (AT&T Colleague)
  Principal Inventive Scientist at AT&T
  1405 US-206, Bedminster, NJ 07921, USA
  yx247e@att.com • +1 (240) 583-0968

[CV compiled on 2025-05-08]