Sandeep **Kumar**

Research Scientist @ Intel, Bengaluru, India

--

(+91) 8277361995 | sandeep007734@gmail.com | sandeep007734.github.io | sandeep007734 | sandeep007734

Professional Summary

Systems researcher specializing in **AI/ML performance optimization** and **memory tiering**. Expertise in memory management for DNN training (PyTorch), including Transformer and CNN architectures, LLM inference, and RAG systems. Proven track record of optimizing memory-intensive workloads through tiered memory architectures, kernel-level enhancements, and hands-on experience with hardware accelerators (Intel IAA, DSA).

Contributor to the **Linux kernel** (memory management subsystem) with upstream patches. Currently leading a research team at Intel focused on AI infrastructure optimization, CXL memory technologies, and next-generation memory tiering solutions.

Selected Publications

1. *TierScape: Harnessing Multiple Compressed Tiers to Tame Server Memory TCO*
   Sandeep Kumar, Aravinda Prasad, and Sreenivas Subramoney. In **EuroSys** , 2026.
2. *TierTrain: Proactive Memory Tiering for CPU-Based DNN Training*
   Sathvik Swaminathan, **Sandeep Kumar**, Aravinda Prasad, and Sreenivas Subramoney. In **ISMM**, Seoul, South Korea, 2025.
3. *Efficient Memory Tiering in a Virtual Machine*
   Chandra Prakash, **Sandeep Kumar**, Aravinda Prasad, and Sreenivas Subramoney. In arXiv, 2025.
4. *A Tug-of-War between Static and Dynamic Memory in Intel SGX*
   Sandeep Kumar, Abhisek Panda, Advait Nerlikar and Smruti R. Sarangi. In **VLSID, Bengaluru, India** , 2024.
5. *Telescope: Telemetry for Gargantuan Memory Footprint Applications*
   Alan Nair, **Sandeep Kumar**, Aravinda Prasad, Ying Huang, Andy Rudoff, Sreenivas Subramoney. In **USENIX, ATC, USA**, 2024.
6. *Enabling Migration of Large SGX Enclaves in a Data Center*
   Sandeep Kumar, Abhisek Panda, Smruti R. Sarangi. In **arXiv** , 2024.
7. *Perspector: Benchmarking Benchmark Suites*
   Sandeep Kumar, Abhisek Panda, Smruti R. Sarangi. In **DATE**, Antwerp, Belgium, 2023.
8. *SecureLease: Maintaining Execution Control in The Wild using Intel SGX*
   Sandeep Kumar, Abhisek Panda, and Smruti R. Sarangi. In **Middleware**, Quebec City, Canada, 2022.
9. *SGXGauge: A Comprehensive Benchmark Suite for Intel SGX*
   Sandeep Kumar, Abhisek Panda, and Smruti R. Sarangi. In **ISPASS**, Singapore, 2022.
10. *SecureFS: A Secure File System for Intel SGX*
    Sandeep Kumar and Smruti R. Sarangi. In **RAID**, Spain, 2021.
11. *Page Table Management for Heterogeneous Memory Systems*
    Sandeep Kumar, Aravinda Prasad, Smruti R. Sarangi, and Sreenivas Subramoney. In **ISMM**, Canada, 2021.
12. *F-LaaS: A Control-Flow-Attack Immune License-as-a-Service Model*
    Sandeep Kumar, Diksha Moolchandani, Takatsugu Ono, and Smruti Sarangi. In **IEEE SCC**, Milan, Italy, 2019.
13. *Scalable Performance Tuning of Hadoop MapReduce: A Noisy Gradient Approach*
    Sandeep Kumar, Sindhu Padakandla, Chandrashekar L, Priyank Parihar, Gopinath K, and Shalabh Bhatnagar. In **IEEE Cloud**, Hawaii, USA, 2017.

Complete list here: https://sandeep007734.github.io/publications/

Patents

1. Context-Aware Memory Tiering for Machine Learning Training
   Sathvik Swaminathan, Sandeep Kumar, Aravinda Prasad, and Sreenivas Subramoney
2. Methods and Apparatus to Profile Page Tables for Memory Management.
   Aravinda Prasad, Sandeep Kumar, Sreenivas Subramoney, and Andy Rudoff

*Multiple patents pending in the filing pipeline.

Education

Indian Institute of Technology Delhi — *New Delhi, India*
PH.D. IN COMPUTER SCIENCE — *Jul. 2017 - Jun. 2024*
- Advised by Prof. S. R. Sarangi

Indian Institute of Science — *Bengaluru, India*
M.E. IN COMPUTER SCIENCE — *Jul. 2011 - Aug. 2013*
- Advised by Prof. K. Gopinath

Guru Gobind Singh Indraprastha Univ. — *New Delhi, India*
B.TECH. IN COMPUTER SCIENCE — *Jul. 2007 - Aug. 2011*

Work Experience

Intel — *Bengaluru, Karnataka*
RESEARCH SCIENTIST — *March 2022 – Present*
- Profiled and optimized AI and cloud workloads (LLM inference, DNN training, VectorDBs) on Intel Xeon platforms using perf, VTune, and eBPF-based tracing.
- Built a fine-grained telemetry framework (Telescope) for production workloads with large memory footprints; submitted patches to the Linux kernel; published at USENIX ATC'24.
- Designed and implemented multi-tier memory management policies in the Linux kernel (EuroSys'26).
- Collaborated with a team contributing patches to the PyTorch ecosystem to profile CPU-side bottlenecks and improve training performance.

Intel Labs — *Bengaluru, Karnataka*
GRADUATE RESEARCH INTERN — *July 2020 – January 2021*
- Investigated page table management overheads for heterogeneous (DRAM + PMEM) memory systems; contributed to a published ISMM paper and a subsequent patent.

Indian Institute of Science — *Bengaluru, Karnataka*
RESEARCH ASSOCIATE — *September 2014 – July 2017*
- Researched scalable performance tuning of Hadoop MapReduce using noisy gradient methods; demonstrated workload-driven configuration optimization at cloud scale.

Dell R&D — *Bengaluru, India*
SOFTWARE DEVELOPMENT ENGINEER — *July 2013 – June 2014*
- Developed systems software in C/C++ for storage infrastructure; performance profiling and regression analysis of enterprise storage workloads.

Interests & Activities

Reading — Goodreads Profile
goodreads.com/sandeep007734

Running, Cycling, Hiking — Strava Profile
strava.com/athletes/3569697

References

Prof. Smruti R. Sarangi — *IIT Delhi, India*
PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE
srsarangi@cse.iitd.ac.in

Prof. K. Gopinath — *IISc Bengaluru, India*
PROFESSOR, COMPUTER SCIENCE AND AUTOMATION
gopi@iisc.ac.in