# HAIYANG SHI

2015 Neil Avenue, Columbus, OH 43210 shi.876@buckeyemail.osu.edu 614-607-4016

## Summary of Qualifications

- 7+ years of research and work experience in designing and developing high-performance and scalable systems
- In-depth knowledge of high-performance interconnects (e.g., Remote Direct Memory Access (RDMA)) and protocols, Erasure Coding (EC), In-Network Computing, and distributed systems
- Skilled in programming languages: C/C++, Java, Python, SQL, and Shell scripting
- Years of experience in using message queues, databases (e.g., MySQL, HBase, TitanDB), and Big Data processing frameworks (e.g., Hadoop, Spark)
- Strong problem-solving and communication skills
- Self-motivated and efficient in cross-group and cross-culture collaboration

## Education

**The Ohio State University** Columbus, OH
*Ph.D. in Computer Science and Engineering. Advisor: Prof. Xiaoyi Lu* *Aug 2016 – Dec 2020*

**Tianjin University** Tianjin, China
*B.E. in Computer Science and Technology* *Aug 2008 – Jul 2012*

## Experience

**Graduate Research Assistant**
***The Ohio State University**, Columbus, OH* *Aug 2016 - Dec 2020*

- **High-Performance Erasure Coding Schemes for Next-Generation Storage Systems**: Research project; Multiple papers get published in top conferences (C++/C/Java)
  - Leading the research and implementation of involved EC libraries and schemes
  - Designed and developed a multi-rail EC library leveraging heterogeneous hardware, a tripartite graph based EC scheme taking advantage of EC NIC (Network Interface Card) offloading, and a set of in-network EC primitives to accelerate EC computation in distributed storage systems
  - Invented a set of micro-benchmarks to evaluate the performance of various EC schemes
- **Fast and Scalable Remote Procedure Call (RPC) Frameworks**: Accelerated Apache Thrift and brpc for modern data centers (*C++*)
  - Leading the design and development of RDMA-oriented RPC frameworks
  - Characterizing and applying networking optimizations for fast message delivery
  - Designed and implemented efficient flow control and communication protocols
  - Participated in benchmarking and testing the frameworks in several data centers
  - RPC frameworks have been exploited in an enterprise database system and demonstrate up to 20% speedup on 1TB TPC-H benchmark
- **Big Data Processing Frameworks over High-Performance Networks**: High-performance derivative of Apache Hadoop accelerated by RDMA technologies (*C, Java*)
  - Redesigned and implemented RDMA-accelerated I/O pipelines for both replication and erasure coding storage schemes in HDFS 3.x
  - Involved in benchmarking and testing related software stacks
  - Benchmark results show that the enhanced Hadoop improves overall throughput by up to 3.1×
  - Open-sourced as a part of OSU HiBD project (http://hibd.cse.ohio-state.edu)

**Research Intern**

***Alibaba**, Seattle, WA. Mentor: Liwei Peng* *May 2019 – Aug 2019*

- Benchmarked distributed training of Transformer and BERT models with varied configurations to study the impact of RDMA interconnects and all-reduce algorithms
- Developed tools to analyze the bottleneck of distributed machine learning

**Software Engineer**

***MiningLamp**, Beijing, China* *May 2015 – Jul 2016*

- **Scopa**: A property graph based analysis system presenting data as meaningfully defined objects and relationships: people, places, events, and the connections between them. (*Java, HBase, Spark, TitanDB*)
  - One of the major engineers in designing and developing the system
  - Developed a graph traversal engine for the underlying graph database - TitanDB (or JanusGraph)

***Sina Weibo (China's Twitter)**, Beijing, China* *Jul 2012 – Apr 2014*

- Participated in designing and developing several subsystems for managing profiles and posts of hundreds of millions of users (*PHP, Message Queue, Memcached, MySQL*)

## Selected Publications

**[SC'20] INEC: Fast and Coherent In-Network Erasure Coding**, *H. Shi, and X. Lu*, In Proceedings of the 33rd International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2020. (To Appear).

**[SC'19] TriEC: Tripartite Graph Based Erasure Coding NIC Offload**, *H. Shi, and X. Lu*, In Proceedings of the 32nd International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2019. **Best Student Paper Finalist**.

**[HPDC'19] UMR-EC: A Unified and Multi-Rail Erasure Coding Library for High-Performance Distributed Storage Systems**, *H. Shi, X. Lu, D. Shankar, and D. K. Panda*, In Proceedings of the 28th ACM International Symposium on High-Performance Parallel and Distributed Computing (HPDC), 2019.

**[Bench'18] EC-Bench: Benchmarking Onload and Offload Erasure Coders on Modern Hardware Architectures**, *H. Shi, X. Lu, and D. K. Panda*, In Proceedings of International Symposium on Benchmarking, Measuring, and Optimizing (Bench), 2018. **Best Paper**.

**[HotI'17] Characterizing Deep Learning over Big Data (DLoBD) Stacks on RDMA-capable Networks**, *X. Lu, H. Shi, H. Javed, R. Biswas, and D. K. Panda*, In Proceedings of the 25th Annual Symposium on High-Performance Interconnects (HotI), 2017.

## Awards

- SC'19 Best Student Paper Finalist
- Bench'18 Best Paper Award

## Professional Involvement

- Publicity chair of HPBDC'20
- Reviewer of TPDS'19
- External/Sub reviewer of TPDS Special Section on AI/ML/DL 2020, HiPC'20, Bench'20, UCC'20