THE UNIVERSITY OF ALABAMA

DEPARTMENT OF COMPUTER SCIENCE

Colloquium Talk

High Performance Computing at Extreme Scale: Data Reduction, Reilience, and Scalability

3:00 - 4:00PM

FRIDAY, JANUARY 26, 2018

SEC 3437

ABSTRACT

Abstract: High-performance computing (HPC) is now almost everywhere, including scientific research and business, such as meteorology, quantum physics and chemistry, geophysics, molecular dynamics, artificial intelligence, and transaction processing, because of HPC platform's powerful computational capability. The next generation HPC platform will be able to achieve ten to the power of eighteen floating-point operations per second (ExaFLOPS). However, some issues are becoming increasingly severe towards exascale HPC systems. One issue is that today's HPC applications are producing extremely large amounts of data, such that data storage in disks, transmission on networks, and processing in post-analysis are becoming more challenging for scientific research. For this issue, we developed a series of data reduction techniques that can significantly reduce the data size with no loss of scientific integrity. Another issue is that supercomputers are being built with an increasing number of complex components, each of which has grown on-chip transistor density. Together with a renewed emphasis on limiting power and energy consumption, this is anticipated to result in these systems being increasingly susceptible to faults and errors. For this issue, we designed a series of fault-tolerant techniques, such as online ABFT and lossy checkpointing, to protect HPC applications from hard and soft errors with low performance overhead.

DINGWEN TAO

Dingwen Tao received his Bachelor's degree in Mathematics from the University of Science and Technology of China (USTC) in June 2013. Currently, he is a Ph.D. candidate in Computer Science at the University of California, Riverside, working under the direction of Dr. Zizhong (Jeffrey) Chen. He is expecting to defend his dissertation in summer 2018. His current research interests include High-Performance Computing, Parallel and Distributed Systems, Big Data Analytics, Cloud and Cluster Computing, Extreme-scale Machine Learning and Deep Learning, Lossy Compression Algorithms, and Reliability, Fault Tolerance, and Resilience for HPC.