Fall 2019 Seminar Series

Title
Accurate and Efficient Off-CPU Performance Analysis
Wednesday, November 13
10:00 – 11:00 a.m.

Abstract
For the purpose of performance optimization, we often need to identify events that are limiting the throughput of the application or creating long latencies. Such events can be categorized into two types: events that execute certain instructions on the CPU (i.e. on-CPU events) and events that wait for other events (i.e. off-CPU events), and both of them can lead to performance problems. While on-CPU analysis is quite well studied, we find existing off-CPU analysis methods are either inaccurate or incomplete. Our works try to develop theoretical models to accurately capture problematic off-CPU events and develop methods to efficiently record corresponding events in both the application and the OS kernel. In this talk, I will present two of our recent works: wPerf tries to identify off-CPU events limiting the throughput of an application and TailMRI tries to identify off CPU events leading to tail latencies. Our evaluation shows that, by optimizing the problems reported by these tools, we can achieve up to 4.8x improvement in throughput and up to 60x reduction in tail latencies in the applications we have studied.

Biography
Yang Wang received the bachelor's and master's degrees in computer science and technology from Tsinghua University, in 2005 and 2008, respectively, and the doctorate degree in computer science from the University of Texas at Austin, in 2014 (advisors Dr. Lorenzo Alvisi and Dr. Mike Dahlin). He is now an assistant professor in the Department of Computer Science and Engineering, the Ohio State University. His current research interests include distributed systems, fault tolerance, scalability, and performance analysis.