COMPUTER SCIENCE
DEFENSE

Measurement and Analysis Methods of Performance Problems in Distributed Systems

Lei Zhang
Emory University

Abstract: Today's distributed systems invest significant computational and storage resources to accommodate their large scale of data, but more resources does not automatically improve performance. To deliver high performance, new types of large-scale solutions, such as the cloud computing and microservices paradigms, follow the design of deploying loosely coupled components that perform but, in the process, making it harder to maintain a global view of system performance. The ensuing growing complexity of system architectures, diagnosing and understanding performance problems has become both critically important and highly challenging. ¡br¿¡br¿ The aim of my thesis is to fill in some missing but significant parts towards monitoring and analyzing performance problems in distributed system, by asking the question: What is the performance bottleneck of distributed systems performance, and how should we improve it? First, my thesis proposes a novel retroactive tracing abstraction where full telemetry information about a distributed request can be retrieved "back in time" soon after a problem is detected without unduly burdening any node in the system, with an always-on distributed tracing system. Second, my thesis frames the challenges of data placement in modern memory hierarchies in a generalized paging model outside of traditional assumptions, and provides an offline data placement algorithm towards optimal placement decisions. Last, my thesis derives a rule-of-thumb expression for cache warmup times, specifically how long caches in storage systems and CDNs need to be warmed up before their performance is deemed to be stable.

Monday, November 8, 2021, 12:00 pm
https://emory.zoom.us/j/94559953414

COMPUTER SCIENCE
EMORY UNIVERSITY