DISSERTATION DEFENSE

Hasan Maruf

Practical Memory Disaggregation

Tuesday, March 21, 2023
12:00 – 2:00pm
4941 Beyster
Hybrid – Zoom
Meeting ID: 973 4642 3832

ABSTRACT: In today's datacenters, compute and memory resources are tightly coupled. This causes fleet-wide resource underutilization and increases the Total Cost of Ownership (TCO) for large-scale datacenters. Modern datacenters are embracing a paradigm shift towards disaggregation, where each resource type is decoupled and connected through a network fabric. Memory, being the prime resource for high-performance services, is becoming an attractive target for disaggregation. Disaggregating memory from compute enables flexibility to scale them independently and better resource utilization. As memory consumes 30-40% of the total rack power and operation cost, proper utilization of stranded resources through disaggregation can save billions of dollars in TCO.

With the advent of ultra-fast networks and coherent interfaces like CXL, disaggregation has become popular over the last few years. There are, however, many open challenges for its practical adoption including the latency gap between local and remote memory access, resiliency, deployability in existing infrastructure, adopting heterogeneity in cluster resources, and providing isolation while maintaining the quality of service. To make memory disaggregation widely adoptable, besides hardware support, software stacks need to be performant while considering all these challenges so that such systems do not degrade application performance beyond a noticeable margin. This dissertation proposes a comprehensive solution to address the host-level, network-level, and end-to-end aspects of practical memory disaggregation.

To bridge the still-sizeable latency gap between local vs. remote memory access, we design Leap – a prefetch-enabled low-latency kernel-space datapath that isolates each application's access to the remote memory. Relying on memory across multiple machines in a disaggregated cluster makes applications susceptible to a wide variety of uncertainties. Hydra addresses this issue by enabling a low-latency, low-overhead, and highly available erasure-coded resilient remote memory datapath at single-digit μs tail latency. Memtrade addresses the deployability challenge – it enables memory disaggregation on public clouds even in the absence of the latest networking hardware and protocols (e.g., RDMA, CXL). TPP allows application-transparent page placement for heterogeneous memory systems and enables efficient memory disaggregation even within a single server. Altogether, this dissertation provides insights on how to ensure a performant, reliable, and easily deployable system for next-generation disaggregated cloud infrastructure.

CHAIR: Prof. Mosharaf Chowdhury