Thesis Defense

Computer Science Master's Program

"Performance Interference Detection for Cloud-Native Applications using Unsupervised Machine Learning Models"

By **Eli Bakshi**

Abstract:

Contemporary cloud-native applications frequently adopt the microservice architecture, where applications are deployed within multiple containers that run on cloud virtual machines (VMs). These applications are typically hosted on public cloud platforms, where VMs from multiple cloud subscribers compete for the same physical resources on a cloud server. When a cloud subscriber application running on a VM competes for shared physical resources from other applications running on the same VM or from other VMs co-located on the same cloud server, performance interference may occur when the performance of an application degrades due to shared resource contention. Detecting such interference is crucial for maintaining the Quality-of-Service of cloud-native Web applications. However, cloud subscribers lack access to underlying host-level hardware metrics traditionally used for interference detection without needing to instrument high overhead-inducing per-request response time values. Machine learning (ML) techniques have proven effective in detecting performance interference using metrics available at the subscriber level, though these techniques have predominantly focused on supervised models with pre-existing labeled data sets that can distinguish between normal and interference conditions. In contrast, this work proposes an unsupervised clustering-based ML approach to identify performance interference in cloud-native applications. The proposed approach implements a lightweight method for collecting container metrics in normal and interference scenarios and applies a dimensionality reduction technique to mitigate redundancy and noise in the collected dataset. We then apply a density-based clustering approach to this unlabeled data set to classify interference in two applications running on the AWS EC2 cloud: a microbenchmark Web application called Acme Air and a large-scale production-realistic Web benchmark called DeathStarBench. Results indicate that our density-based clustering approach effectively distinguishes between normal and interference conditions and achieves an average Density-Based Clustering Validation (DBCV) index of 0.805 and a cluster homogeneity of 0.856 across both applications.

Date: Wednesday, June 12th, 2024
Time: 2:00 PM – 4:00 PM
Location: 14-238b
Committee: Dr. Mukherjee, Dr. Mukhopadhyay, and Dr. Humer