CECS
CENTER FOR EMBEDDED & CYBER-PHYSICAL SYSTEMS
UNIVERSITY OF CALIFORNIA · IRVINE

CECS/Dept. of Computer Science Seminar

"Storage-Centric Computing for Genomics and Metagenomics"

Nika Mansouri Ghiasi

Ph.D. candidate of SAFARI Research Group,
ETH Zürich

Thursday, October 17th
3:00-4:00 p.m.
Location: DBH 4011

Abstract:

Genomics and metagenomics applications have enabled significant advancements in many critical areas. The exponential growth of genomic data poses unprecedented challenges in genomics and metagenomic applications. These applications suffer from significant data movement overheads from the storage system. To fundamentally address these overheads, we make a case for storage-centric computing.

First, we propose GenStore, the first in-storage processing system designed for genome sequence analysis that greatly reduces both data movement and computational overheads of genome sequence analysis by exploiting low-cost and accurate in-storage filters. We address the challenges of in-storage processing, supporting reads with 1) different read lengths and error rates, and 2) different degrees of genetic variation. Through rigorous analysis of read mapping processes, we design low-cost hardware accelerators and data/computation flows inside a NAND flash-based SSD. Our evaluation using a wide range of real genomic datasets shows that GenStore significantly improves the read mapping performance of state-of-the-art software (hardware) baselines by 2.07-6.05× (1.52-3.32×) for read sets with high similarity to the reference genome and 1.45-33.63× (2.70-19.2×) for read sets with low similarity to the reference genome.

Second, we propose MegIS, the first in-storage processing system designed to significantly reduce the data movement overhead of the end-to-end metagenomic analysis pipeline. MegIS is enabled by our lightweight design that effectively leverages and orchestrates processing inside and outside the storage system. Through our detailed analysis of the end-to-end metagenomic analysis pipeline and careful hardware/software co-design, we address in-storage processing challenges for metagenomics via specialized and efficient 1) task partitioning, 2) data/computation flow coordination, 3) storage technology-aware algorithmic optimizations, 4) data mapping, and 5) lightweight in-storage accelerators. MegIS's design is flexible, capable of supporting different types of metagenomic input datasets, and can be integrated into various metagenomic analysis pipelines. Our evaluation shows that MegIS outperforms the state-of-the-art performance- and accuracy-optimized software metagenomic tools by 2.7×–37.2× and 6.9×–100.2×, respectively, while matching the accuracy of the accuracy-optimized tool. MegIS achieves 1.5×–5.1× speedup compared to the state-of-the-art metagenomic hardware-accelerated (using processing-in-memory) tool, while achieving significantly higher accuracy.

Biography:

Nika Mansouri Ghiasi is a Ph.D. candidate in the SAFARI Research Group at ETH Zürich, working with Professor Onur Mutlu. Her current research interests are in computer architecture, emerging technologies, bioinformatics, and their interactions.

Hosted By: Prof. Alex Veidenbaum