CAS CS 561 A1: Data Systems Architectures
Data-intensive Systems and Computing Lab
Department of Computer Science
College of Arts and Sciences, Boston University
http://bu-disc.github.io/CS561/

CS561 A1 Spring 2021 - Research Project

Title: Concurrency-Aware Graph/Tree Traversal Algorithms

Background: Modern storage devices like solid-state disks (SSDs) have immense parallelism opportunities because of their architecture. This means that multiple concurrent I/Os are needed to saturate the whole bandwidth of the device. Otherwise, the device remains underutilized. To exploit the concurrency of these devices, recent research has focused on developing new I/O schedulers for SSDs taking advantage of the existing parallelism. However, there have been very little works that focus on exploiting the concurrency for better algorithm design. The goal of this project is to utilize the internal parallelism of modern storage device for better algorithm design: specifically, designing better graph/tree traversal algorithms.

For example, using the device concurrency, algorithms for tree and graph traversal will be able to access multiple nodes concurrently. By doing so, they can have a wider/deeper search space. Consider a depth-first search (DFS) algorithm that can offer DFS guarantees, while at the same time covering a wider search-space by loading concurrently sibling nodes. On the other hand, one can construct an algorithm with breadth-first-search (BFS) guarantees that follows a few promising paths as deep as the might get using the concurrency.

Objective: The objective of the project is to develop concurrency-aware graph/tree traversal algorithms that are optimized for modern storage devices

(a) Get familiarized with modern storage device's properties like read/write asymmetry, concurrency.

(b) Design concurrency-aware graph/tree traversal algorithms.

[1] Feng Chen, Binbing Hou, and Rubao Lee. 2016. Internal Parallelism of Flash Memory-Based Solid-State Drives. ACM Trans. Storage 12, 3 (2016), 13:1—-13:39. https://doi.org/10.1145/2818376
[2] Feng Chen, Rubao Lee, and Xiaodong Zhang. 2011. Essential roles of exploiting internal parallelism of flash memory based solid state drives in high-speed data processing. In Proceedings of the IEEE International Symposium on High Performance Computer Architecture (HPCA). 266–277.