Stony Brook University
The Graduate School

Doctoral Defense Announcement

Abstract

Cache Oblivious Data Structures for Massive Data Sets

By

Haodong Hu

The cache-oblivious model is one of the most successful models of a memory hierarchy. The cache-oblivious model allows programmers to reason about a simple two-level memory model without knowing any memory parameters, but to prove results about a multilevel memory model. Thus, algorithms and structures based on the cache-oblivious model have the advantage of platform independence and simultaneously optimal on all levels of a memory hierarchy. The disk-access model (DAM), another successful memory model, assumes a two-level memory model with the full knowledge of memory parameters. Like the DAM model, the performance of the cache-oblivious model is measured by memory (or block) transfers between two adjacent memory levels with block size.

In this dissertation, we build highly efficient, optimized cache-oblivious structures in support of cache-oblivious B-trees and other dictionaries. We contribute the theory of the searching cost in the cache-oblivious model by presenting the generalized van Emde Boas layout. We develop two dynamic cache-oblivious structures: the adaptive packed-memory array and the partially deamortized packed-memory array to overcome the traditional packed-memory array's deficiencies and make it practical. We also propose the atomic key B-tree in support of atomic keys with variable length, whose structure is as simple as the traditional B-tree while having theoretical and practical guarantees.

Date: November 08, 2007 **Program**: Computer Science
Time: 11:20AM **Dissertation Advisor**: Michael A. Bender
Place: Computer Science Building, Room 1310.