Range Reporting

Peter Gabrielsen 20114179
Christoffer Hansen 20114637

February 5, 2016

Abstract

blabla

Contents

1 Introduction

2 Model of computation

We will argue the results of this thesis in terms of the external memory model of Aggarwal and Vitter [AV88]. The external memory model (or I/O model) measures the efficiency of an algorithm by counting the total number of reads and writes performed. In detail the model consists of two levels of memory; a bounded internal memory of size M and an unbounded external memory. For a total of N records we define an *IO* operation to the process of transferring B consecutive records between the two levels of memory as depicted in figure 1. We restrict all computations on records to be done in internal memory. Throughout the thesis we will let K denote the total number of records in the output.

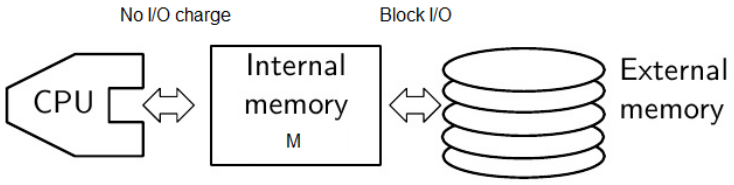

Figure 1: The IO Model. Only reads/writes between internal and external memory is charged.

The fundamental bounds in the external memory model is that scanning can be done in $\mathcal{O}(\text{Scan}) = \mathcal{O}(N/B)$, sorting in $\mathcal{O}(\text{Sort}) = \mathcal{O}(N/B \log_{M/B} N/B)$ and searching in $\mathcal{O}(\log_B N)$. We denote $\mathcal{O}(N/B)$ as being linear in terms of IOs. Note that the B factor is very important as $N/B < \mathcal{O}(N/B \log_{M/B} N/B) \ll N$.

For convenience we will assume $M > B^2$. This assumption is known as the *tall-cache assumption* in the cache-oblivious model and basically states

that the number of blocks M/B is larger than the size of each block B [Pro99].

3 Preliminaries

3.1 Amortization

3.2 Global Rebuilding

The term *global rebuilding* refers to the standard technique of making a (typically small) static data structure dynamic. We simply store all updates in a *update block* and once a certain threshold has been collected we rebuild the data structure [**?**]. For data structures that does not allow the space for deleted records to be reoccupied we *mark* (or *weak delete*) the elements. Whenever αN elements have been marked, for some constant $\alpha > 0$, the entire data structure is rebuilt from scratch with only the non-marked elements. The cost of rebuilding is at most a constant factor higher than the cost of inserting αN elements and so the amortized cost of global rebuilding can be charged to the insertions of the deleted elements [MSS03].

3.3 Filtering

Charge part of query cost to output.

4 Sublogarithmic Updates

5 External Memory Priority Search Tree

Bibliography

[AV88] Alok Aggarwal and S. Vitter, Jeffrey. The input/output complexity of sorting and related problems. *Commun. ACM*, 31(9):1116–1127, September 1988.

[MSS03] Ulrich Meyer, Peter Sanders, and Jop Sibeyn, editors. *Algorithms for Memory Hierarchies: Advanced Lectures*. Springer-Verlag, Berlin, Heidelberg, 2003.

[Pro99] Harald Prokop. Cache-oblivious algorithms, 1999.