Evaluating Parallel R-Tree Implementations on a Network of Workstations*

Ning An  Liujian Qian  Anand Sivasubramaniam  Tom Keefe

Department of Computer Science & Engineering
The Pennsylvania State University
University Park, PA 16802.
Phone: (814) 865-1406
Fax: (814) 865-3176
anand@cse.psu.edu

Technical Report CSE-98-006
May 1998

Abstract

It is becoming increasingly important that a Geographical Information System delivers high performance to efficiently store, retrieve and process the voluminous data that it needs to handle. It is necessary to employ processing and storage parallelism for scalable long-term solutions. With the demise of many custom-built parallel machines, it is imperative that we use off-the-shelf technology to provide this parallelism. A closely-coupled network of workstations is a viable alternative.

In this paper, we explore techniques for distributing a spatial data structure (R-tree) across a network of workstations. We provide a framework to explore design alternatives in distributing the R-tree across workstations. We also develop an extensive system to implement and evaluate these alternatives. Specifically, we show this by implementing two distribution schemes, and evaluating their performance for insert and spatial search operations on two different data sets.

Keywords: Geographical Information Systems, Parallel Processing, Spatial Data Structures, Network of Workstations.

*This research is supported in part by a NSF Career Award MIP-9701475, EPA grant R825195-01-0, and equipment grants from NSF and IBM.