

Scalable Computing Software Laboratory Technical Report

Department of Computer Science

Illinois Institute of Technology

Automatic Construction of Pre-execution Prefetching Thread for Parallel Applications

Yong Chen, Surendra Byna, Xian-He Sun, Rajeev Thakur, William Gropp

{chenyon1, sbyna, sun}@iit.edu, thakur@mcs.anl.gov, wgropp@uiuc.edu

November 2007

Technical Report № IIT/CS-2007-22

<http://www.cs.iit.edu>

10 West 31st Street, Chicago, IL 60616

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IIT-SCS and will probably be copyrighted if accepted for publication. It has been issued as a Technical Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IIT-SCS prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g. payment of royalties).

Automatic Construction of Pre-execution Prefetching Thread for Parallel Applications

Yong Chen¹ Surendra Byna^{1,2} Xian-He Sun¹ Rajeev Thakur² William Gropp³

¹ *Department of Computer Science, Illinois Institute of Technology, Chicago, IL*
{chenyon1, bynasur, sun} @iit.edu

² *Mathematics & Computer Science Division, Argonne National Laboratory, Argonne, IL*
thakur@mcs.anl.gov

³ *Department of Computer Science, University of Illinois Urbana-Champaign, Urbana, IL*
wgropp@uiuc.edu

Abstract

Parallel applications are usually able to achieve high computational performance but suffer from large latency in I/O accesses. I/O prefetching is an effective solution to masking the latency; however, most of existing I/O prefetching techniques are conservative and the effectiveness is limited by low accuracy and coverage. As the processor performance has been increasing rapidly, multi-core/many-core architecture processor has been the trend for future high-performance computers, and the computing power is virtually free, we recently proposed a novel pre-execution approach for hiding I/O latency. In this study, we analyze the pre-execution prefetching thread construction problem in detail and present the methodology, and automatic construction with programming slicing technique, as well as design considerations in various aspects.

Keywords: I/O performance, pre-execution prefetching, program slicing, parallel application performance