Reducing Shared Cache Misses via Dynamic Coscheduling on Multicores

by

Wael Amr Hossam El Din Lotfy

A Thesis Submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE
in
COMPUTER ENGINEERING

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2015

Reducing Shared Cache Misses via Dynamic Coscheduling on Multicores

by

Wael Amr Hossam El Din Lotfy

A Thesis Submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE
in
COMPUTER ENGINEERING

Under the Supervision of

Dr. Ihab ElSayed Talkhan
Professor
Computer Engineering Department
Cairo University

Dr. Hany Mohamed ElSayed
Assistant Professor
Communication Engineering Department
Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2015

Reducing Shared Cache Misses via Dynamic Coscheduling on Multicores

by

Wael Amr Hossam El Din Lotfy

A Thesis Submitted to the

Faculty of Engineering at Cairo University

In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE
in
COMPUTER ENGINEERING

Approved by the examining committee:

Prof. Dr. Ihab ElSayed Talkhan, *Thesis Main Advisor*

Associate Prof. Dr. Ibrahim Mohamed Kamar, *Internal Examiner*

Prof. Dr. Ahmed Abdelrahman Abou-Auf, *External Examiner*

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2015

Engineer: Wael Amr Hossam El Din Lotfy
Date of Birth: 07 / 07 / 1984
Nationality: Egyptian
E-mail: wael.amrhossam@gmail.com
Phone: 01285692539
Address: Haydiek El Ahram, Masakin El Sharkaa El Watania, El Haram.
Registration Date: 01 / 10 / 2008
Awarding Date: / /
Degree: Masters of Science
Department: Computer Engineering

Supervisors: Prof. Dr. Ihab ElSayed Talkhan
Assistant Prof. Dr. Hany Mohamed ElSayed

Examiners: Prof. Dr. Ihab E. Talkhan (Cairo University)
Associate Prof. Dr. Ibrahim M. Qamar (Cairo University)
Prof. Dr. Ahmed A. Abou-Auf (American University)

Title of Thesis:
Reducing Shared Cache Misses Via Dynamic Coscheduling On Multicores

Keywords:
Cache Miss Rate, Dynamic Scheduling, Multicore, Symbiosis

Summary:

Multicore technology enables having two processors or more working together on the same single chip. It enables the system to perform more tasks with higher overall system performance. However, this performance can't be exploited well due to the high miss rate in the second level shared cache among the cores which represents one of the multicore's challenges. This problem can be reduced by using the scheduling techniques. This thesis addresses the dynamic co-scheduling of tasks in multicore real-time systems. The focus is on the basic idea of the megatask technique for grouping the tasks that may affect the shared cache miss rate, and the Pfair scheduling that is then used for reducing the concurrency within the grouped tasks while ensuring the real time constraints. So all the grouped tasks could not be run at the same time. Consequently the shared cache miss rate is reduced. The dynamic co-scheduling is proposed through the combination of the symbiotic technique with the megatask technique for co-scheduling the tasks based on the collected information.

Experiments show that the proposed dynamic co-scheduling can decrease the shared cache miss rate compared to the static one by up to 52%.

Advisor's Signature:

- ----------

- ----------

Acknowledgments

First of all, I would like to thank ALLAH for his guidance and help. While I wrote this thesis, it is ultimately the result of inspiration from, and interaction with, a large number of people. I thank my advisors, Dr. Hany Mohamed Elsayed and Dr. Ihab Elsayed Talkhan, for their support, guidance, and patience throughout my graduate career.

I also thank Islam Mohamed Hatem Atta for his supporting and providing me valuable feedback on my research.

Finally, I would like to thank my family and friends for being supportive in all that I do and for their prayers; particularly my father Amr HossamElDin,my mother, Eman Salem, my sister ,Ghadah Amr and my friend, Wael Nabil.

Dedication

I dedicate this thesis to every student who cares about the research and the knowledge for developing better nations and achieving a better humanitarian level.

Table of Contents

List of Tables

List of Figures

Abstract

Multicore technology enables having two processors or more working together on the same single chip. It enables the system to perform more tasks with higher overall system performance. However, this performance can't be exploited well due to the high miss rate in the second level shared cache among the cores which represents one of the multicore's challenges. This problem can be reduced by using the scheduling techniques. Although the scheduling techniques have been recently proposed for multicores platforms, most of them have aimed at the system throughput or cache performance without ensuring the real time constraints or measuring the shared cache miss rate. A recent work targeted scheduling for reducing cache miss rate while ensuring the real time constraints among Multicore platforms. However, this scheduling was static, i.e. it does not adapt itself to operating conditions.

This thesis addresses the dynamic co-scheduling of tasks in multicore real-time systems. The focus is on the basic idea of the megatask technique for grouping the tasks that may affect the shared cache miss rate, and the Pfair scheduling that is then used for reducing the concurrency within the grouped tasks while ensuring the real time constraints. So all the grouped tasks could not be run at the same time. Consequently the shared cache miss rate is reduced. The dynamic co-scheduling is proposed through the combination of the symbiotic technique with the megatask technique for co-scheduling the tasks based on the collected information using two schemes. The first scheme is measuring the temporal working set size of each running task at run time, while the second scheme is collecting the shared cache miss rate of each running task at run time.

The used simulation methodology is based on implementing a cache simulator that is based on SESC simulator. The PIN tool is used to generate memory access request files using different applications from the SPECJVM2008 benchmarks, then these files are used as an input for the cache simulator.

Experiments show that the proposed dynamic co-scheduling can decrease the shared cache miss rate compared to the static one by up to 52%. This indicates that the dynamic co-scheduling is important to achieve high performance with shared cache memory for running high workloads like multimedia applications that require real-time response and continuous-media data types.

1 Introduction

In 1965, Gordon E. Moore, the intel co-founder, explained the growth of semiconductors technology in a formula. This formula states that the number of transistors on a chip will double approximately every two years. This observation has since been dubbed "Moore's Law" and is now enormously influential. ”Moore's Law” has been considered in the processor industry. As transistors are increased on a chip, the processor clock frequencies are also increased at the same time. ”Moore's law” has reached its limit in 2010. One reason behind this is that the faster processors have power consumption and thermal dissipation. This needs special cooling systems and leads to higher cost. Hence processor industry has moved towards the multicore technology since the delivered performance of single cores can not meet the needed requirements for running different applications like web servers, multimedia programs and databases. Multicore technology is introduced to increase the required performance and power efficiency. However, there are new challenges for this technology.

Next section contains some figures that show the architectures for the uniprocessor, multiprocessors and multicores, then it tackles the challenges of the multicores. These challenges open the door in research to propose new software and hardware solutions.

1.1 Chip Multiprocessors

Chip Multiprocessors (can also be named a Multicore processor) refers to a single chip that integrates two or more processors in an area that would have originally been filled with a single large uniprocessor. This solves the power consumption problem when adding more transistors to the uniprocessor and switching it at higher and higher frequencies.

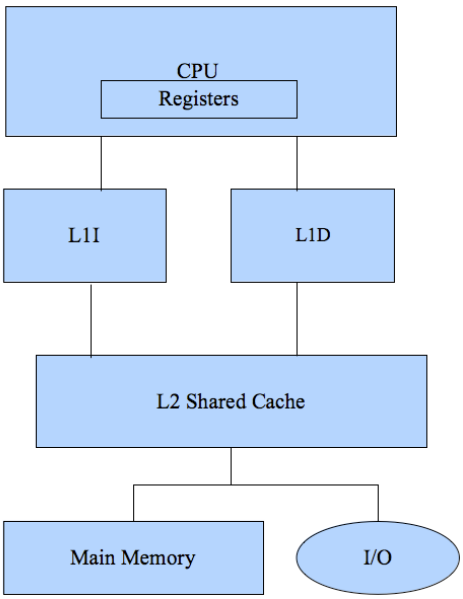

Figure 1.1 Conventional Microprocessor

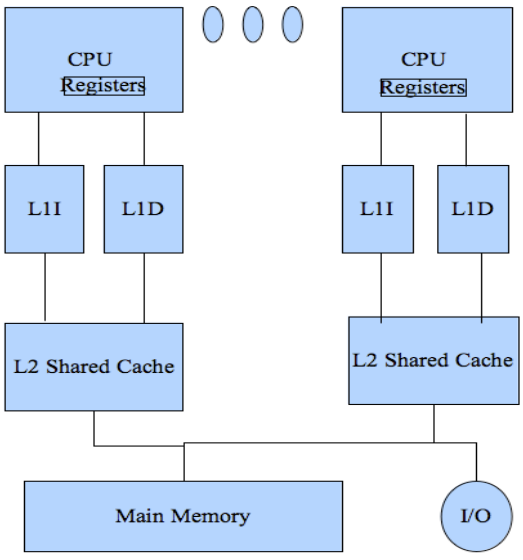

Figure 1.2 Simple Chip Multiprocessor

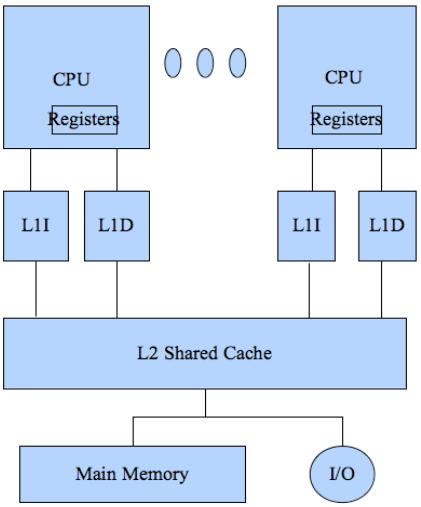

Figure 1.3 Shared Cache Multiprocessor

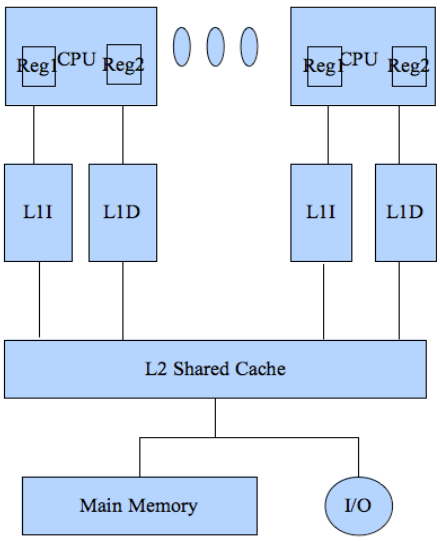

Figure 1.4 Multithreaded Shared Cache Multiprocessor