UNIVERSITI PUTRA MALAYSIA

CACHE REPLACEMENT ALGORITHM USING HIERARCHICAL ALLOCATION SCHEDULING

MOHAMMAD FAIZAL BIN MOHD SHARIF

FK 2014 44

CACHE REPLACEMENT ALGORITHM USING HIERARCHICAL ALLOCATION SCHEDULING

By
MOHAMMAD FAIZAL BIN MOHD SHARIF

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

May 2014

DEDICATION

This work is lovingly dedicated to
OUR ALMIGHTY GOD ALLAH
My wife Mazuin binti Mohammad
My son Mohammad Aqmal
My daughters Marsha and Mahira

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science.

CACHE REPLACEMENT ALGORITHM USING HIERARCHICAL ALLOCATION SCHEDULING

By

MOHAMMAD FAIZAL BIN MOHD SHARIF

May 2014

Chair: Harlisya binti Harun, PhD
Faculty: Engineering

Cache management has become one of the most popular areas of research in improving the performance of computation. Cache memory is the nearest block of storage between the main memory and the processor. Caches are very fast but have a small memory storage size; therefore only selected instructions or data are to be kept in cache memory. A cache management or also known as a cache replacement policy is a mechanism to manage the process of selecting, keeping, replacing, or evicting the instruction inside cache memory. It is important to determine instructions that need to be kept (or not to be evicted) in order to avoid any disruption for future processes.

Therefore, this research is an effort of developing a conceptual model of the cache replacement policy. Predicting a relative important instruction in cache replacement policy can improve the cache performance. The predicted instruction will have the highest priority to be kept in cache memory for future use. Therefore, latency between the processor and memory can be reduced.

The existing cache replacement (i.e. least recently used (LRU)) algorithm depends on usage of the data being referenced. Data that are least referenced will be removed during cache replacement if the cache is already full. There is no determination of relative important instruction in this policy. By removing this, based on the least recent, it might cause a potential delay in the future processing if the removed instruction depends on it.

To alleviate this limitation, the Hierarchical Allocation Scheduling (HAS) model is proposed in this research. HAS is to schedule the instructions based on the priority of instructions from the aspect of space and time. The core principle is to keep the relatively important instruction in cache memory from being evicted. The development of the HAS model is based on the idea of hierarchical temporal memory (HTM) developed by Numenta Inc. HTM is an approach in which the scheduling is derived on priority and similarity of instruction from the aspect of space and time. The implementation of the HAS model is developed using the OCTAVE software. A simulation of the model is performed to analyse its behaviour in a hypothetical cache management scenario.

Specifically, the goal of the experimentation is to study the situation in which the HAS model can accurately predict a single instruction with the highest relative importance. Analyses indicate that it is the most single instruction occurrence when the data range (D) is defined as being between 0 to 9 and the window size (k) is fixed at 10. This only contributes to the simulated behaviour of the model regardless to its actual implementation in the cache replacement policy.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
Sebagai memenuhi keperluan untuk ijazah Sarjana Sains

ALGORITHM PENGGANTIAN CACHE MENGGUNAKAN PENJADUALAN PERUNTUKAN HIERARKI

Oleh

MOHAMMAD FAIZAL BIN MOHD SHARIF

May 2014

Pengerusi: Harlisya binti Harun, PhD
Fakulti: Kejuruteraan

Pengurusan cache telah menjadi salah satu penyelidikan yang paling popular dalam bidang peningkatan prestasi pengkomputeran. Ingatan cache adalah blok terdekat di antara menyimpan ingatan utama dan pemproses. Kelajuan cache sangat cepat tetapi ianya mempunyai saiz menyimpan memori yang kecil; oleh itu hanya arahan atau data terpilih akan disimpan dalam ingatan cache. Pengurusan cache atau juga dikenali sebagai polisi penggantian cache adalah satu mekanisma untuk menguruskan proses memilih, menjaga, menggantikan atau menyingkirkan arahan-arahan di dalam ingatan cache. Ia adalah penting untuk menentukan arahan yang perlu untuk disimpan (atau tidak akan disingkir) untuk mengelakkan sebarang gangguan bagi proses-proses yang akan datang.

Oleh itu, kajian ini adalah satu usaha untuk membangunkan konsep model dalam polisi penggantian cache. Meramalkan arahan penting secara relatif dalam polisi penggantian cache boleh meningkatkan prestasi cache. Arahan-arahan yang diramalkan akan mempunyai keutamaan yang tinggi untuk disimpan dalam ingatan cache untuk kegunaan masa depan. Oleh itu, kependaman antara pemproses dan ingatan dapat dikurangkan.

Algorithm penggantian cache yang sedia ada (iaitu “least recently used” (LRU)) bergantung kepada penggunaan data yang dirujuk. Data yang paling kurang dirujuk akan disingkirkan semasa gentian cache jika isi kandungan cache telah penuh. Tiada penentuan arahan penting secara relatif dalam polisi ini. Dengan ini, berdasarkan kaedah rujukan yang kurang, ia berkemungkinan penyebab kelewatan pemprosesan di masa hadapan jika arahan-arahan yang disingkirkan bergantung kepada arahan tersebut.

Untuk mengatasi kekangan ini, model penjadualan peruntukan hierarki (HAS) adalah dicadangkan dalam kajian ini. HAS adalah untuk menjadualkan arahan-arahan berdasarkan keutamaan arahan dari segi ruang dan masa. Prinsip utamanya adalah untuk menyimpan arahan-arahan penting secara relatif dalam ingatan cache daripada yang disingkirkan. Pembangunan model HAS adalah berdasarkan kepada idea hirarki memori sementara (HTM) yang telah dibangunkan oleh Numenta Inc. HTM adalah satu pendekatan di mana penjadualan berdasarkan keutamaan dan kesamaan arahan-arahan dari segi ruang dan masa. Pelaksanaan model HAS dibangunkan dengan menggunakan perisian OCTAVE. Satu simulsi model dilakukan untuk menganalisa perilaku di dalam satu senario pengurusan hipotetikal cache.

Secara khususnya, matlamat ujikaji ini adalah untuk mempelajari keadaan di mana model HAS mampu meramalkan satu arahan dengan tepat mengikut kepentingan relatif tertinggi. Analisis menunjukkan berlakunya arahan tunggal yang tinggi apabila julat data (D) ditakrifkan di antara 0 hingga 9 dan saiz tetingkap (k) ditetapkan pada 10. Ini hanyalah menyumbang kepada simulasi perilaku pada model tanpa mengambil kira pelaksanaannya yang sebenar di dalam penggantian polisi cache

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Harlisya binti Harun for the continuous support of my master in research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master research.

Besides my advisor, I would like to thank the rest of my thesis committee, Prof. Ir. Dr. Norman bin Mariun for their encouragement, insightful comments, and hard questions.

My sincere thanks also go to Dean of Universiti Kuala Lumpur Malaysian Institute of Aviation Technology (UniKL MIAT) and Organization Development Department of UniKL for offering me the full time study for my master. Also, thanks to my fellow friend at UniKL MIAT, Puan Safura binti Abdul Rahman for proofread my thesis.

Last but not the least, I would like to thank my family, my parents for supporting me spiritually throughout my life.

APPROVAL

I certify that a Thesis Examination Committee has met on 16th May 2014 to conduct the final examination of Mohammad Faizal bin Mohd Sharif on his thesis entitled **"Cache Replacement Algorithm using Hierarchical Allocation Scheduling (HAS)"** in accordance with the Universities and University Collages Act 1971 and Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science in Aerospace Engineering.

Members of the Thesis Examination Committee were as follows:

H. R. Harijono Djojodihardjo, Sc.D., IPU
Professor
Faculty of Engineering (Aerospace)
Universiti Putra Malaysia
(Chairman)

Makhfudzah binti Mokhtar, PhD
Faculty of Engineering (Computer and Communication System)
Universiti Putra Malaysia
(Internal Examiner)

Rizal Zahari, PhD
Associate Professor
Faculty of Engineering (Aerospace)
Universiti Putra Malaysia
(Internal Examiner)

Nur'Aini Abdul Rashid, PhD
Associate Professor
School of Computer Science
Universiti Sains Malaysia
Malaysia
(External Examiner)

_____

NORITAH OMAR, PhD
Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 21 July 2014

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the Master of Science. The members of the Supervisory Committee were as follows:

Harlisya binti Harun, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Norman bin Mariun, PhD
Professor Ir.
Faculty of Engineering
Universiti Putra Malaysia
(Member)

_____

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No.: Mohammad Faizal Bin Mohd Sharif, GS29524

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature:	_____	Signature:	_____
Name of Chairman of Supervisory Committee:	**Harlisya binti Harun, PhD** Senior Lecturer	Name of Member of Supervisory Committee:	**Norman bin Mariun, PhD** Professor Ir.

TABLE OF CONTENTS