Effectiveness Evaluation of Replacement Policies for On-Chip Caches in Multiprocessors

Jobin Jose, National Institute of Technology, Tiruchirappalli, India*

Shameedha Begum, National Institute of Technology, Tiruchirappalli, India

https://orcid.org/0000-0003-4543-5692

Ramasubrmanian N., National Institute of Technology, Tiruchirappalli, India

ABSTRACT

A cache plays a vital role in improving performance in the multicore environment, especially the last level cache (LLC). The improvements in performance are based on the block size, associativity, and replacement policies. Most of the papers concentrate on traditional least recently used (LRU)-based replacement policies for their replacement decisions. Unfortunately, the replacement decisions do not enhance performance of the cache as expected. An enhanced modified pseudo LRU policy is proposed, which is an approximation of LRU. The proposed methodology uses counters to enhance the confidence of replacement decisions based on the history of the replaceable blocks in cache. It is very clear from the simulation results that the replacement scheme proposed exhibits better performance improvement in terms of miss ratio of about 3% and energy efficiency of about 2% on average.

KEYWORDS

Cache, Last Level Cache (LLC), Least Recently Used (LRU), Pseudo Least Recently Used (PLRU), Replacement Policies

INTRODUCTION

The processing power and energy dissipation are the major concerns nowadays in the multi-core architecture and thus leads the technology front into a new dimension called Low power VLSI (Hanson, H. et. al (June 2003)). As in the case, multiple processing cores have to perform their operation simultaneously or in parallel that requires a need for a large amount of memory, as the capability of memory doesn't increase in phase with the multicore chips that lead to access limitations to processing cores. The current researches are focusing towards the effort to design memory hierarchies that would be efficient (Laha et. al (2002). These designs eliminate the gap in the performance of various cache levels (Asaduzzaman, A. et. al (2009)) thus reducing the cache level accesses, in addition to it the time on an average required to access the memory is also quantifiably reduced, thus improving the performance of execution. Deep investigations are performed by numerous researches on Level 2 (L2) caches for various reasons (Patidar, K., 2015). The introduction of the new level enables some

*Corresponding Author

number of hits in the cache level L2 which actually follows the same data misses in the level L2 of the cache, which is hidden (Flautner and K., et. al (May 2002g)). The cache helps to utilize parallelism at the instruction level (Carr, S. (1996)) along with the processor, to achieve execution of instructions in out of order mode, there by determining the cache lines no blocking event. So, it is seriously hard to hide the miss penalty of L2 cache. In the case of considering the optimizations of L1, that result in shorter hit times, is complex on the grounds of execution time in comparison to the hits in L2 cache (Anjana, J. G., and Prasanth, M. (2014, May)). Replacement policy is an important element, that is related to all cache levels in a multilevel hierarchy of memory. The efficiency of a replacement depends on mainly, the hit rate or miss rate and sometimes the access latency, bandwidth utilization and response time of a cache system.

The lower most level in the hierarchy of the fastest memory cache is termed as Last level cache (LLC) in modern processors and allows efficient optimizations on the cache with smarter replacement policies, because these are the major bottlenecks in improving the performance which are being shared amongst the cores. Since LLCs are considered as the highly associative cache (Gutierrez, A. et. al (2014)), whose size is large compared to other level caches, the replacement policy becomes the key technique to optimize the performance in terms of hits and misses, also ensures effective utilization with much more energy efficiency.

All the traditional cache replacement policies come in force nowadays has no room to allocate a necessary block to the cache lines. A miss in the cache line decides whether there should be an eviction or not. The eviction candidates are determined according to the provisions applied by various replacement policies. In exception to the random policies, all other policies will detect the cache line to be removed taking into account of its access frequencies. Considering the Least Recently Used (LRU) (Dan, A. and Towsley, D. (1990)) replacement, a considerable number of status bits are needed to indicate whether the cache block is accessed or not and updating is done accordingly. Owing to the increase in number of bits caused by set associativity between the main memory and cache, results in increase in complexity of cost and computation. The random policy is used as the best way to reduce the complexity associated with the LRU, mainly in finding an item in the cache. All those algorithms increase the hardware complexity also. Most of the modern processors choose to use pseudo LRU to bring down the cost of the hardware and increase the system performance by the process of approximating the mechanism of LRU. Recent studies show that incorporating the techniques of LRU with certain limit in associativity, as their performance optimization method as far as replacement policy is concerned, few of them started working on the upgradation of LRU by remodeling replacement decisions. An approach to the tree-based Pseudo LRU policy is proposed in (Jiménez, D. A. (*2013*)), to maintain the cache effectively by incorporating a counter, which increases the confidence in the eviction process so that none of the blocks are being replaced unnecessarily. This paper evaluates the effectiveness of PLRU policies over LRU by incorporating a simple counter to minimize the changes in the tree structure. This ensures that the thrashing applications are getting the necessary space in the cache. Simulation results show that its performance is increased on an overall basis when compared to PLRU.

The subsequent two sections of the paper describe related works and the methodology used. The successive two sections clearly explain the method proposed by the authors and the experimental setup with its appropriate results. The conclusion is depicted in the final section.

RELATED WORKS

Traditionally, the replacement decisions on a cache block, that are made depend on whether the cache access in the future resulted a miss in any cache line. The history of replacement is started with the famous belady's optimal policy (Belady, L. A. (1966)) which discards the page whose next access is farthest in the future. Saying in another way, the decision for replacement is made, depending on the value of the reuse distance. The reuse distance is calculated by taking the difference in the values of

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/effectiveness-evaluation-of-replacement-policies-for-on-chip-caches-in-multiprocessors/289202

Related Content

Tapping the Wisdom of Crowd Phenomenon in Healthcare Social Networks
Alaa Al-Kadiand Samir Chatterjee (2012). *International Journal of Business Data Communications and Networking (pp. 42-56).*
www.irma-international.org/article/tapping-wisdom-crowd-phenomenon-healthcare/72884

Game Theory-Based Coverage Optimization for Small Cell Networks
Yiqing Zhou, Liang Huang, Lin Tianand Jinglin Shi (2016). *Game Theory Framework Applied to Wireless Communication Networks (pp. 184-211).*
www.irma-international.org/chapter/game-theory-based-coverage-optimization-for-small-cell-networks/136639

Reviewing HTTP and RTSP Work in Two Actual Commercial Media Delivery Platforms for Multimedia Services and Mobile Devices
Martin Zimmermannand Gilbert Seilheimer (2012). *Multimedia Services and Streaming for Mobile Devices: Challenges and Innovations (pp. 91-110).*
www.irma-international.org/chapter/reviewing-http-rtsp-work-two/58700

Dependency of Transport Functions on IEEE802.11 and IEEE802.15.4 MAC/PHY Layer Protocols for WSN: A Step towards Cross-layer Design
Atif Sharif, Vidyasagar M. Potdarand A. J. D. Rathnayaka (2010). *International Journal of Business Data Communications and Networking (pp. 1-30).*
www.irma-international.org/article/dependency-transport-functions-ieee802-ieee802/45137

Implementation and Evaluation of Skip-Links: A Dynamically Reconfiguring Topology for Energy-Efficient NoCs
Simon J. Hollisand Chris Jackson (2013). *Adoption and Optimization of Embedded and Real-Time Communication Systems (pp. 181-203).*
www.irma-international.org/chapter/implementation-evaluation-skip-links/74248