

Relative Competitiveness of Cache Replacement Policies*

Jan Reineke
Saarland University, Germany
reineke@cs.uni-sb.de

Daniel Grund
Saarland University, Germany
grund@cs.uni-sb.de

Categories and Subject Descriptors

B.3.3 [Memory Structures]: Performance Analysis and Design Aids—*Worst-case analysis*; I.6.5 [Simulation and Modeling]: Model Development—*Modeling Methodologies*

General Terms

Performance, Design, Algorithms, Theory

Keywords

Cache performance, replacement policy, worst-case execution time, WCET analysis, predictability

1. INTRODUCTION

Caches are commonly employed to hide the latency gap between memory and the CPU by exploiting locality in memory accesses. On today's architectures a cache miss may take several hundred CPU cycles. In order to fulfill stringent timing requirements, caches are now also used in hard real-time systems. In such systems, guarantees have to be made concerning the best- and worst-case execution times (BCET and WCET) of tasks. To obtain tight bounds on the execution times of a task, timing analyses *must* take into account the cache architecture. However, developing cache analyses – analyses that statically determine whether a memory access associated with an instruction will always be a hit or a miss – is a difficult problem. Precise and efficient analyses have been developed for set-associative caches that employ the least-recently-used (LRU) replacement policy. Other commonly used policies, like first-in-first-out (FIFO) or Pseudo-LRU (PLRU) have proven to be more difficult to analyze.

Relative competitive analyses yield upper (lower) bounds on the number of misses (hits) of a policy P relative to the number of misses (hits) of another policy Q . For example, a competitive analysis may find out that policy P will incur at most 30% more misses than policy Q and at most 20% less hits in the execution of any task. Note that P and Q may have different associativities.

*This work has profited from discussions within the ARTIST2 Network of Excellence. It is supported by the German Research Foundation (DFG) as part of SFB/TR AVACS and the German-Israeli Foundation (GIF). The second author is supported by a scholarship in the DFG GK 623.

We propose the following approach to determine bounds on the number of cache hits and misses by a task T under FIFO(k), PLRU(l)¹, or any another replacement policy:

1. Determine competitiveness of the desired policy P relative to a policy Q for which a cache analysis exists, like LRU.
2. Perform cache analysis of task T for policy Q to obtain a cache-performance prediction, i.e. upper (lower) bounds on the number of misses (hits) by Q .
3. Calculate upper (lower) bounds on the number of misses (hits) for P using the cache analysis results for Q and the competitiveness results of P relative to Q .

A limitation of this approach is that it only produces upper (lower) bounds on the number of misses (hits) for the whole program execution. It does not reveal at which program points the misses (hits) will happen, something many timing analyses need. Relative competitiveness results can also be used to obtain sound *may* and *must* cache analyses [1], i.e. analyses that can classify individual accesses as hits or misses.

We have developed a tool to automatically compute relative competitiveness results for a large class of replacement policies, including LRU, FIFO, and PLRU.

2. RELATIVE COMPETITIVENESS

Competitiveness as introduced by Sleator and Tarjan captures the worst-case performance of an online replacement policy relative to the optimal offline policy. Relative competitiveness captures the worst-case performance of an online policy relative to another online policy.

Formally, it can be defined as follows, where $m_P(q, s)$ is the number of misses incurred by policy P , starting in cache-set state q , processing memory access sequence s :

DEFINITION 1 (RELATIVE MISS-COMPETITIVENESS).
A policy P is k -miss-competitive relative to policy Q with additive constant c , if

$$m_P(p, s) \leq k \cdot m_Q(q, s) + c$$

for all access sequences $s \in S$ and compatible cache-set states $p \in C^P, q \in C^Q$.

In other words, policy P will incur at most k times the number of misses of policy Q plus a constant c on any access sequence. Hit-competitiveness is defined analogously:

¹ k and l denote the respective associativities of FIFO(k) and PLRU(l).

Associativity:	Miss-Competitiveness								Hit-Competitiveness							
	2	3	4	5	6	7	8		2	3	4	5	6	7	8	
LRU vs FIFO	2,1	3,2	4,3	5,4	6,5	7,6	8,7		0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
FIFO vs LRU	2,1	3,2	4,3	5,4	6,5	7,6	8,7		$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, 1$	$\frac{1}{2}, \frac{3}{2}$	$\frac{1}{2}, 2$	$\frac{1}{2}, \frac{5}{2}$	$\frac{1}{2}, 3$	$\frac{1}{2}, 3$	$\frac{1}{2}, \frac{7}{2}$
LRU vs PLRU	1,0	–	2,1	–	–	–	5,4		1,0	–	$\frac{1}{2}, 1$	–	–	–	–	$\frac{1}{8}, \frac{15}{8}$
PLRU vs LRU	1,0	–	∞	–	–	–	∞		1,0	–	$\frac{1}{2}, 1$	–	–	–	–	$\frac{1}{4}, \frac{3}{2}$
FIFO vs PLRU	2,1	–	4,4	–	–	–	8,8		$\frac{1}{2}, \frac{1}{2}$	–	$\frac{1}{4}, \frac{5}{4}$	–	–	–	–	$\frac{1}{11}, \frac{19}{11}$
PLRU vs FIFO	2,1	–	∞	–	–	–	∞		0,0	–	0,0	–	–	–	–	0,0

Figure 1: Miss- and Hit-Competitiveness ratios k and additive constants c relating FIFO, PLRU, and LRU at the same associativity. PLRU is only defined for powers of two. As an example of how this should be read, LRU(4) is 2-miss-competitive relative to PLRU(4) with additive constant 1, whereas PLRU(4) is not miss-competitive relative to LRU(4) at all. ∞ indicates that there is no k such that the policy on the left is k -miss-competitive relative to the policy on the right.

DEFINITION 2 (RELATIVE HIT-COMPETITIVENESS).

A policy P is k -hit-competitive relative to policy Q with subtractive constant c , if

$$h_P(p, s) \geq k \cdot h_Q(q, s) - c$$

for all access sequences $s \in S$ and compatible states $p \in C^P, q \in C^Q$.

Notice, that the two definitions are only redundant if $k = 1$, i.e. if policy A is 1-miss-competitive relative to policy B then A is also 1-hit-competitive relative to B . One can obtain sound *may* and *must* analyses [1] from competitiveness results in the special case of 1-competitiveness:

THEOREM 1 (*MAY AND MUST ANALYSES*). *If policy P is 1-miss-competitive relative to policy Q with additive constant 0, then*

- (i) *A must analysis for Q is a sound must analysis for P .*
- (ii) *A may analysis for P is a sound may analysis for Q .*

PROOF. See [3]. \square

Consider a policy A that is k -miss-competitive relative to policy B . A is also l -miss-competitive relative to B for $l > k$. However, the former statement is clearly a better characterization of the policy’s relative competitiveness. The best characterization of the relative competitiveness of a policy is called its *competitive ratio*. This is what our tool computes.

We have reduced the problem of computing these competitive ratios to the minimum cycle ratio problem over finite graphs which represent the possible relative behavior of the two policies under consideration. For an explanation of the theory behind our fully-automatic tool confer [3].

3. ANALYSIS RESULTS

Figure 1 depicts relative miss- and hit-competitiveness results among FIFO, PLRU, and LRU if compared at the same level of associativity, obtained automatically with our tool. For an explanation of the three policies confer [4].

Miss-Competitiveness.

By Sleator/Tarjan [5], FIFO(k) and LRU(k) are k -miss-competitive relative to the optimal offline policy MIN (also called OPT). This implies at least k -miss-competitiveness (of these two) relative to any online algorithm. In contrast, PLRU(4) and PLRU(8) are not k -miss-competitive relative to LRU(4) and LRU(8), respectively, for any k . This is

particularly interesting as it has been suggested in [2] to model a PLRU(k)-cache by an LRU(k)-cache to simplify WCET prediction, as it “does not add a significant error.”

It is also interesting to compare policies of different associativities. One of the obtained results is that PLRU(k) is 1-miss-competitive (therefore also 1-hit-competitive) relative to LRU($1 + \log_2 k$), which follows from a theorem in [4].

Hit-Competitiveness.

The hit-competitiveness results show less symmetry than in the case of miss-competitiveness. LRU(k) is *not* hit-competitive relative to FIFO(k) for any k . In contrast, FIFO(k) is $\frac{1}{2}$ -hit-competitive relative to LRU(k) for all of the investigated k . We have proven the generalization to arbitrary k by hand (see [3]). Considering the prominence of the two policies it is quite surprising that this relation has apparently not been discovered before.

Again, it was worthwhile to compare policies of different associativities. Our analysis results suggest that an LRU($2k - 1$) is 1-hit-competitive (and therefore also 1-miss-competitive) relative to FIFO(k), which could also be proven for arbitrary k . Due to Theorem 1, a *may* cache analysis [1] for LRU($2k - 1$) may be used as a *may* analysis for FIFO(k) as well. *may* analyses are used to infer misses. No analysis has been published to date, that is able to infer misses for a FIFO cache.

Increasing the associativity of FIFO relative to LRU never results in 1-competitiveness, i.e. FIFO(l) is not 1-competitive relative to LRU(k) for any $k > 1$ and l .

4. REFERENCES

- [1] C. Ferdinand and R. Wilhelm. Efficient and precise cache behavior prediction for real-time systems. *Real-Time Systems*, 17(2-3):131–181, 1999.
- [2] A. Hergenhan and W. Rosenstiel. Static timing analysis of embedded software on advanced processor architectures. In *DATE '00*, pages 552–559, New York, NY, USA, 2000. ACM Press.
- [3] J. Reineke and D. Grund. Relative competitive analysis of cache replacement policies. In *LCTES*, 2008 (to appear).
- [4] J. Reineke, D. Grund, C. Berg, and R. Wilhelm. Timing predictability of cache replacement policies. *Real-Time Systems*, 37(2):99–122, November 2007.
- [5] D. D. Sleator and R. E. Tarjan. Amortized efficiency of list update and paging rules. *Commun. ACM*, 28(2):202–208, 1985.