Relative Competitive Analysis of Cache Replacement Policies

Jan Reineke, Daniel Grund

Department of Computer Science
Saarland University

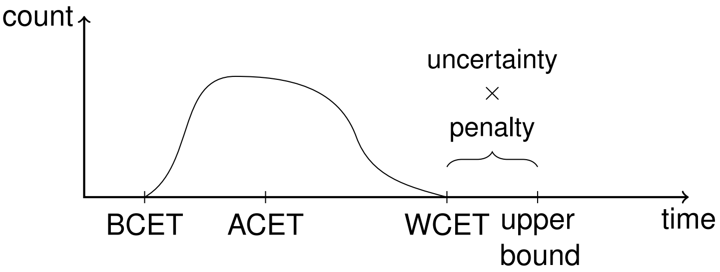
LCTES, June 2008

Outline

Motivation

- Caches used in hard real-time systems
- Need to derive upper and lower bounds on WCET and BCET

⟶ Need cache analysis



- In literature: almost exclusively LRU
- In practice: LRU, FIFO, PLRU, Pseudo Round-Robin, . . .

Approach

1. Determine competitiveness of the policy P relative to policy Q.

$$\boxed{\mathbf{m_P \leq k \cdot m_Q + c}}$$

Approach

1. Determine competitiveness of the policy P relative to policy Q.

$$\mathbf{m_P \leq k \cdot m_Q + c}$$

2. Compute performance of task T for policy Q by cache analysis.

$$\mathbf{m_Q(T)}$$

Approach

1. Determine competitiveness of the policy *P* relative to policy *Q*.

   $$\mathbf{m_P \leq k \cdot m_Q + c}$$

2. Compute performance of task *T* for policy *Q* by cache analysis.

   $$\mathbf{m_Q(T)}$$

3. Calculate upper bounds on the number of misses for *P* using the cache analysis results for *Q* and the competitiveness results of *P* relative to *Q*.

   $$\boxed{\mathbf{m_P \leq k \cdot m_Q + c}}\ \boxed{\mathbf{m_Q(T)}} = \boxed{\mathbf{m_P(T)}}$$

Cache Analysis

Two types of cache analysis:

1. Global guarantees: bounds on cache hits/misses [GMM98, CPHL01]
2. Local guarantees: classification of individual accesses [FMW97, FW99, WHW$^+$97, RM05]

$\longrightarrow$ Can provide both!

Outline

Relative Competitiveness

- Competitiveness (Sleator and Tarjan, 1985): worst-case performance of an online policy *relative to the optimal offline policy* (MIN, OPT, BEL)
  - used to evaluate online policies, many extensions
- Relative competitiveness (Reineke and Grund, 2008): worst-case performance of an online policy *relative to another online policy*

Examples – Relative Miss-Competitiveness

P is 3-miss-competitive relative to Q with additive constant 4.
If Q incurs 7 misses, then P can incur at most $3 \cdot 7 + 4 = 25$ misses.

Examples – Relative Miss-Competitiveness

P is 3-miss-competitive relative to Q with additive constant 4.
If Q incurs 7 misses, then P can incur at most $3 \cdot 7 + 4 = 25$ misses.

P is $\frac{1}{2}$-miss-competitive relative to Q.

$$\Longrightarrow m_P \leq \frac{1}{2} \cdot m_Q \text{ on all access sequences.} \quad \lightning$$

Examples – Relative Miss-Competitiveness

P is 3-miss-competitive relative to Q with additive constant 4.
If Q incurs 7 misses, then P can incur at most $3 \cdot 7 + 4 = 25$ misses.

P is $\frac{1}{2}$-miss-competitive relative to Q.

$$\Longrightarrow m_P \leq \frac{1}{2} \cdot m_Q \text{ on all access sequences.} \quad \lightning$$

Best: P is 1-miss-competitive relative to Q.

Examples – Relative Miss-Competitiveness

P is 3-miss-competitive relative to *Q* with additive constant 4.
If *Q* incurs 7 misses, then *P* can incur at most $3 \cdot 7 + 4 = 25$ misses.

P is $\frac{1}{2}$-miss-competitive relative to *Q*.

$\Longrightarrow m_P \leq \frac{1}{2} \cdot m_Q$ on all access sequences. $\lightning$

Best: *P* is 1-miss-competitive relative to *Q*.

Worst: *P* is not-miss-competitive (or ∞-miss-competitive) relative to *Q*.

Definition – Relative Miss-Competitiveness

Notation

$$m_P(p, s) \quad = \quad \textit{number of misses that policy } P \textit{ incurs on access sequence } s \in M^* \textit{ starting in state } p \in C^P$$

Definition – Relative Miss-Competitiveness

Notation

$$m_P(p, s) \quad = \quad \textit{number of misses that policy } P \textit{ incurs on access sequence } s \in M^* \textit{ starting in state } p \in C^P$$

Definition (Relative miss competitiveness)

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 $s \in M^*$ and compatible cache-set states $p \in C^P, q \in C^Q$.

Definition – Relative Miss-Competitiveness

Notation

$$m_P(p, s) \quad = \quad \textit{number of misses that policy } P \textit{ incurs on access sequence } s \in M^* \textit{ starting in state } p \in C^P$$

Definition (Relative miss competitiveness)

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 $s \in M^*$ and compatible cache-set states $p \in C^P, q \in C^Q$.

Definition (Competitive miss ratio of P relative to Q)

The smallest k, such that P is k-miss-competitive relative to Q.

Examples – Relative Hit-Competitiveness

P is $\frac{2}{3}$-hit-competitive relative to Q with subtractive constant 3.
If Q has 27 hits, then P has at least $\frac{2}{3} \cdot 27 - 3 = 15$ hits.

Examples – Relative Hit-Competitiveness

P is $\frac{2}{3}$-hit-competitive relative to Q with subtractive constant 3.
If Q has 27 hits, then P has at least $\frac{2}{3} \cdot 27 - 3 = 15$ hits.

P is 2-hit-competitive relative to Q.

$$\Longrightarrow h_P \geq 2 \cdot h_Q \text{ on all access sequences.} \lightning$$

Examples – Relative Hit-Competitiveness

P is $\frac{2}{3}$-hit-competitive relative to Q with subtractive constant 3.
If Q has 27 hits, then P has at least $\frac{2}{3} \cdot 27 - 3 = 15$ hits.

P is 2-hit-competitive relative to Q.

$\Longrightarrow h_P \geq 2 \cdot h_Q$ on all access sequences. $\lightning$

Best: P is 1-hit-competitive relative to Q.
Equivalent to 1-miss-competitiveness.

Examples – Relative Hit-Competitiveness

P is $\frac{2}{3}$-hit-competitive relative to Q with subtractive constant 3.
If Q has 27 hits, then P has at least $\frac{2}{3} \cdot 27 - 3 = 15$ hits.

P is 2-hit-competitive relative to Q.

$\Longrightarrow h_P \geq 2 \cdot h_Q$ on all access sequences. $\lightning$

Best: P is 1-hit-competitive relative to Q.
Equivalent to 1-miss-competitiveness.

Worst: P is 0-hit-competitive relative to Q.
Analogue to ∞-miss-competitiveness.

Definition – Relative Hit-Competitiveness

Notation

$$h_P(p, s) \quad = \quad \textit{number of hits that policy } P \textit{ incurs on access sequence } s \in M^* \textit{ starting in state } p \in C^P$$

Definition (Relative hit-competitiveness)

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 $s \in M^*$ and compatible cache-set states $p \in C^P, q \in C^Q$.

Definition (Competitive hit ratio of P relative to Q)

The greatest k, such that P is k-hit-competitive relative to Q.

1-Competitiveness

Let P be 1-(miss-)competitive relative to Q with constant 0:

$$m_P(p,s) \leq 1 \cdot m_Q(q,s) + 0$$
$$\Leftrightarrow m_P(p,s) \leq m_Q(q,s)$$

1-Competitiveness

Let P be 1-(miss-)competitive relative to Q with constant 0:

$$m_P(p, s) \leq 1 \cdot m_Q(q, s) + 0$$
$$\Leftrightarrow m_P(p, s) \leq m_Q(q, s)$$

1. If Q “hits” so does P, and
2. if P “misses” so does Q.

1-Competitiveness

Let P be 1-(miss-)competitive relative to Q with constant 0:

$$m_P(p,s) \leq 1 \cdot m_Q(q,s) + 0$$
$$\Leftrightarrow m_P(p,s) \leq m_Q(q,s)$$

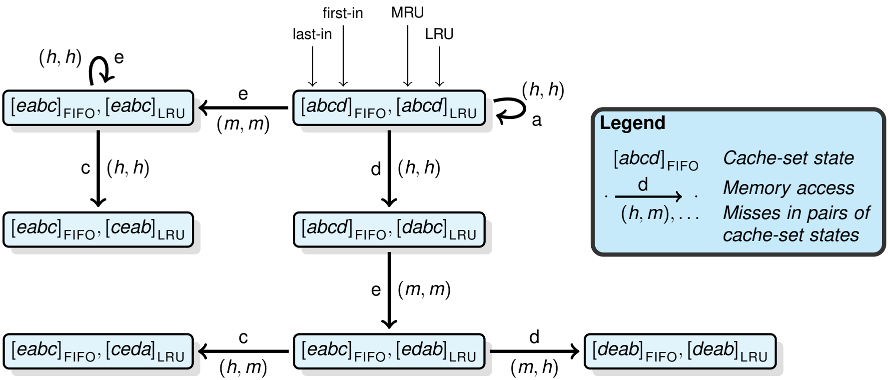
1. If Q “hits” so does P, and
2. if P “misses” so does Q.

As a consequence,

1. a *must*-analysis for Q is also a sound *must*-analysis for P, and
2. a *may*-analysis for P is also a sound *may*-analysis for Q.

Relative Competitiveness – Automatic Computation

P and *Q* induce transition system (running example):



Competitive miss ratio = maximum ratio of misses in policy *P* relative to the number of misses in policy *Q* in transition system

Transition System is ∞ Large

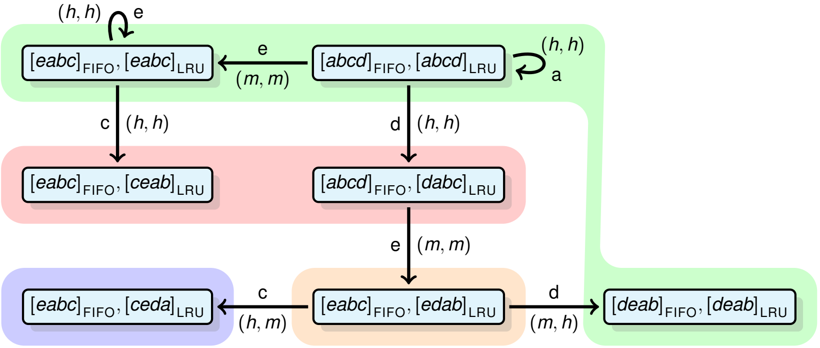
Problem: The induced transition system is ∞ large.

Goal: Construct *finite transition system* with same properties.

Observation: Only the *relative positions* of elements matter:

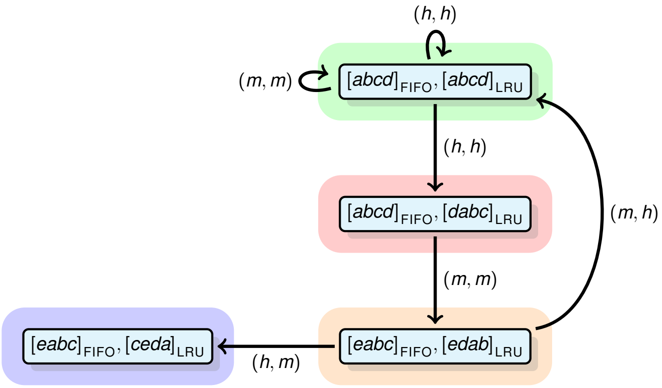
$$\begin{tikzcd} {[abc]_{\mathrm{LRU}}, [bde]_{\mathrm{FIFO}}} \arrow[d, "c"', "(h,m)"] \arrow[r, phantom, "\approx"] & {[fgl]_{\mathrm{LRU}}, [ghm]_{\mathrm{FIFO}}} \arrow[d, "l"', "(h,m)"] \\ {[cab]_{\mathrm{LRU}}, [cbd]_{\mathrm{FIFO}}} \arrow[r, phantom, "\approx"] & {[lfg]_{\mathrm{LRU}}, [lgh]_{\mathrm{FIFO}}} \end{tikzcd}$$

≈-Equivalent States in Running Example



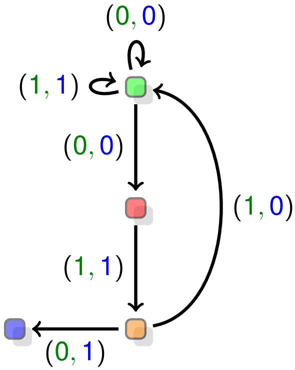
Finite Quotient Transition System

Merging $\approx$-equivalent states yields a finite quotient transition system:



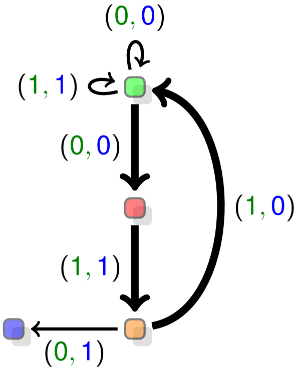
Competitive Ratio = Maximum Cycle Ratio

Competitive miss ratio = maximum ratio of misses in policy P relative to the number of misses in policy Q in transition system



Competitive Ratio = Maximum Cycle Ratio

Competitive miss ratio = maximum ratio of misses in policy P relative to the number of misses in policy Q in transition system



Maximum cycle ratio = $\frac{0+1+1}{0+1+0} = 2$

Outline

Miss-Competitiveness Results

Miss-competitiveness *ratios, constants* relating FIFO, PLRU, and LRU at the same associativity:

Associativity:	2	3	4	5	6	7	8
LRU vs FIFO	$2, 1$	$3, 2$	$4, 3$	$5, 4$	$6, 5$	$7, 6$	$8, 7$
FIFO vs LRU	$2, 1$	$3, 2$	$4, 3$	$5, 4$	$6, 5$	$7, 6$	$8, 7$
LRU vs PLRU	$1, 0$	–	$2, 1$	–	–	–	$5, 4$
PLRU vs LRU	$1, 0$	–	∞	–	–	–	∞
FIFO vs PLRU	$2, 1$	–	$4, 4$	–	–	–	$8, 8$
PLRU vs FIFO	$2, 1$	–	∞	–	–	–	∞

Example:

LRU(4) is 2-miss-competitive relative to PLRU(4) with constant 1.

PLRU(4) is not miss-competitive relative to LRU(4) at all.

Miss-Competitiveness Results

Miss-competitiveness *ratios, constants* relating FIFO, PLRU, and LRU at the same associativity:

Associativity:	2	3	4	5	6	7	8
LRU vs FIFO	2, 1	3, 2	4, 3	5, 4	6, 5	7, 6	8, 7
FIFO vs LRU	2, 1	3, 2	4, 3	5, 4	6, 5	7, 6	8, 7
LRU vs PLRU	1, 0	–	2, 1	–	–	–	5, 4
PLRU vs LRU	1, 0	–	∞	–	–	–	∞
FIFO vs PLRU	2, 1	–	4, 4	–	–	–	8, 8
PLRU vs FIFO	2, 1	–	∞	–	–	–	∞

Example:

LRU(4) is 2-miss-competitive relative to PLRU(4) with constant 1.

PLRU(4) is not miss-competitive relative to LRU(4) at all.

Miss-Competitiveness Results

Miss-competitiveness *ratios, constants* relating FIFO, PLRU, and LRU at the same associativity:

Associativity:	2	3	4	5	6	7	8
LRU vs FIFO	2, 1	3, 2	4, 3	5, 4	6, 5	7, 6	8, 7
FIFO vs LRU	2, 1	3, 2	4, 3	5, 4	6, 5	7, 6	8, 7
LRU vs PLRU	1, 0	–	2, 1	–	–	–	5, 4
PLRU vs LRU	1, 0	–	∞	–	–	–	∞
FIFO vs PLRU	2, 1	–	4, 4	–	–	–	8, 8
PLRU vs FIFO	2, 1	–	∞	–	–	–	∞

Example:

LRU(4) is 2-miss-competitive relative to PLRU(4) with constant 1.
PLRU(4) is not miss-competitive relative to LRU(4) at all.

Hit-Competitiveness Results

Hit-competitiveness *ratios, constants* relating FIFO, PLRU, and LRU at the same levels of associativity:

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

Generalizations

Identified patterns and proved generalizations by hand.

Generalizations

Identified patterns and proved generalizations by hand.

Previously unknown facts:

PLRU(k) is 1 comp. rel. to LRU($1 + log_2 k$) with constant 0,

$\longrightarrow$ LRU-*must*-analysis can be used for PLRU

Generalizations

Identified patterns and proved generalizations by hand.

Previously unknown facts:

PLRU(k) is 1 comp. rel. to LRU($1 + log_2 k$) with constant 0,

$\longrightarrow$ LRU-*must*-analysis can be used for PLRU

FIFO(k) is $\frac{1}{2}$ hit-comp. rel. to LRU(k),

Generalizations

Identified patterns and proved generalizations by hand.

Previously unknown facts:

PLRU(k) is 1 comp. rel. to LRU($1 + log_2 k$) with constant 0,
$\longrightarrow$ LRU-*must*-analysis can be used for PLRU

FIFO(k) is $\frac{1}{2}$ hit-comp. rel. to LRU(k), whereas
LRU(k) is 0 hit-comp. rel. to FIFO(k),

Generalizations

Identified patterns and proved generalizations by hand.

Previously unknown facts:

PLRU(k) is 1 comp. rel. to LRU($1 + log_2 k$) with constant 0,
$\longrightarrow$ LRU-*must*-analysis can be used for PLRU

FIFO(k) is $\frac{1}{2}$ hit-comp. rel. to LRU(k), whereas

LRU(k) is 0 hit-comp. rel. to FIFO(k), but

LRU($2k - 1$) is 1 comp. rel. to FIFO(k) with constant 0.
$\longrightarrow$ LRU-*may*-analysis can be used for FIFO

Outline

Summary

Relative Competitiveness

…bounds performance of an online policy by that of another one,

…allows to derive guarantees on cache performance,

…can be computed automatically by building quotient system!

Summary

Relative Competitiveness

...bounds performance of an online policy by that of another one,

...allows to derive guarantees on cache performance,

...can be computed automatically by building quotient system!

Thank you for your attention!
Questions?

Siddhartha Chatterjee, Erin Parker, Philip J. Hanlon, and Alvin R. Lebeck.
Exact analysis of the cache behavior of nested loops.
In *PLDI '01: Proceedings of the ACM SIGPLAN 2001 conference on Programming language design and implementation*, pages 286–297, New York, NY, USA, 2001. ACM Press.

Christian Ferdinand, Florian Martin, and Reinhard Wilhelm.
Applying compiler techniques to cache behavior prediction.
In *Proceedings of the ACM SIGPLAN Workshop on Languages, Compilers and Tools for Real-Time Systems*, pages 37–46, Las Vegas, Nevada, June 1997.

Christian Ferdinand and Reinhard Wilhelm.
Efficient and precise cache behavior prediction for real-time systems.
Real-Time Systems, 17(2-3):131–181, 1999.

Somnath Ghosh, Margaret Martonosi, and Sharad Malik.

Precise miss analysis for program transformations with caches of arbitrary associativity.
In *ASPLOS-VIII: Proceedings of the eighth international conference on Architectural support for programming languages and operating systems*, pages 228–239, New York, NY, USA, 1998. ACM Press.

Harini Ramaprasad and Frank Mueller.
Bounding worst-case data cache behavior by analytically deriving cache reference patterns.
In *RTAS '05: Proceedings of the 11th IEEE Real Time on Embedded Technology and Applications Symposium*, pages 148–157, Washington, DC, USA, 2005. IEEE Computer Society.

Randall T. White, Christopher A. Healy, David B. Whalley, Frank Mueller, and Marion G. Harmon.
Timing analysis for data caches and set-associative caches.
In *RTAS '97*, page 192, Washington, DC, USA, 1997. IEEE Computer Society.