Relative Competitive Analysis of Cache Replacement Policies

Jan Reineke, Daniel Grund

Department of Computer Science
Saarland University

LCTES, June 2008
1 Introduction
   - Motivation
   - Approach

2 Relative Competitiveness
   - Definition
   - Automatic Computation

3 Results
   - Automatically Computed
   - Generalizations

4 Summary
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 $Q$.

$$m_P \leq k \cdot m_Q + c$$
Approach

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

   \[ m_P \leq k \cdot m_Q + c \]

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

   \[ m_Q(T) \]
Approach

1. Determine competitiveness of the policy $P$ relative to policy $Q$.
   \[ m_P \leq k \cdot m_Q + c \]

2. Compute performance of task $T$ for policy $Q$ by cache analysis.
   \[ m_Q(T) \]

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 $Q$.
   \[ m_P \leq k \cdot m_Q + c \Rightarrow m_Q(T) = m_P(T) \]
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]

→ Can provide both!
1 Introduction
   • Motivation
   • Approach

2 Relative Competitiveness
   • Definition
   • Automatic Computation

3 Results
   • Automatically Computed
   • Generalizations

4 Summary
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*
$P$ 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 $Q$.

\[
\Rightarrow m_P \leq \frac{1}{2} \cdot m_Q \text{ on all access sequences.}
\]
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 $Q$.

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

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 $Q$.

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

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

Worst: $P$ is not-miss-competitive (or $\infty$-miss-competitive) relative to $Q$. 
Definition! – Relative Miss-Competitiveness

Notation

\[ m_P(p, s) = \text{number of misses that policy } P \text{ incurs on access sequence } s \in M^* \text{ starting in state } p \in C_P \]
**Definition – Relative Miss-Competitiveness**

**Notation**

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

**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**

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

**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 $Q$.

$$\implies h_P \geq 2 \cdot h_Q$$ on all access sequences.
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 $Q$.

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

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 \( Q \).

\[ \Rightarrow h_P \geq 2 \cdot h_Q \text{ on all access sequences.} \]

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 \( \infty \)-miss-competitiveness.
Definition – Relative Hit-Competitiveness

Notation

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

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

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

$$\Leftrightarrow m_P(p, s) \leq m_Q(q, s)$$
Let $P$ be 1-(miss-)competitive relative to $Q$ with constant 0:

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

$$\iff m_P(p, s) \leq m_Q(q, s)$$

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 \textit{must}-analysis for $Q$ is also a sound \textit{must}-analysis for $P$, and
2. a \textit{may}-analysis for $P$ is also a sound \textit{may}-analysis for $Q$. 

Jan Reineke, Daniel Grund
Relative Competitiveness – Automatic Computation

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

\[
\begin{align*}
\text{Legend} & \quad [abcd]_{\text{FIFO}} \cdot [abcd]_{\text{LRU}} \\
& \quad \text{Cache-set state} \\
& \quad \text{Memory access} \\
& \quad \text{Misses in pairs of cache-set states}
\end{align*}
\]

\[
\begin{align*}
\text{Competitive miss ratio} & = \max \text{ ratio of misses in policy } P \text{ relative to the number of misses in policy } Q \text{ in transition system}
\end{align*}
\]
Problem: The induced transition system is $\infty$ large.
Goal: Construct *finite transition system* with same properties.
Observation: Only the *relative positions* of elements matter:
\(-\)-Equivalent States in Running Example

Jan Reineke, Daniel Grund
Relative Competitiveness
LCTES 2008
15 / 23
Merging $\sim$-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

\[
\begin{align*}
\text{(0, 0)} & \quad \text{(1, 1)} \\
\text{(0, 0)} & \quad \text{(1, 0)} \\
\text{(1, 1)} & \quad \text{(0, 1)}
\end{align*}
\]
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

1 Introduction
   - Motivation
   - Approach

2 Relative Competitiveness
   - Definition
   - Automatic Computation

3 Results
   - Automatically Computed
   - Generalizations

4 Summary
Miss-Competitiveness Results

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

<table>
<thead>
<tr>
<th>Associativity:</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRU vs FIFO</td>
<td>2,1</td>
<td>3,2</td>
<td>4,3</td>
<td>5,4</td>
<td>6,5</td>
<td>7,6</td>
<td>8,7</td>
</tr>
<tr>
<td>FIFO vs LRU</td>
<td>2,1</td>
<td>3,2</td>
<td>4,3</td>
<td>5,4</td>
<td>6,5</td>
<td>7,6</td>
<td>8,7</td>
</tr>
<tr>
<td>LRU vs PLRU</td>
<td>1,0</td>
<td></td>
<td>2,1</td>
<td></td>
<td></td>
<td></td>
<td>5,4</td>
</tr>
<tr>
<td>PLRU vs LRU</td>
<td>1,0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIFO vs PLRU</td>
<td>2,1</td>
<td></td>
<td>4,4</td>
<td></td>
<td></td>
<td></td>
<td>8,8</td>
</tr>
<tr>
<td>PLRU vs FIFO</td>
<td>2,1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Associativity:</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRU vs FIFO</td>
<td>2,1</td>
<td>3,2</td>
<td>4,3</td>
<td>5,4</td>
<td>6,5</td>
<td>7,6</td>
<td>8,7</td>
</tr>
<tr>
<td>FIFO vs LRU</td>
<td>2,1</td>
<td>3,2</td>
<td>4,3</td>
<td>5,4</td>
<td>6,5</td>
<td>7,6</td>
<td>8,7</td>
</tr>
<tr>
<td>LRU vs PLRU</td>
<td>1,0</td>
<td></td>
<td>2,1</td>
<td></td>
<td></td>
<td></td>
<td>5,4</td>
</tr>
<tr>
<td>PLRU vs LRU</td>
<td>1,0</td>
<td></td>
<td>∞</td>
<td></td>
<td></td>
<td></td>
<td>∞</td>
</tr>
<tr>
<td>FIFO vs PLRU</td>
<td>2,1</td>
<td></td>
<td>4,4</td>
<td></td>
<td></td>
<td></td>
<td>8,8</td>
</tr>
<tr>
<td>PLRU vs FIFO</td>
<td>2,1</td>
<td></td>
<td>∞</td>
<td></td>
<td></td>
<td></td>
<td>∞</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Associativity:</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRU vs FIFO</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>FIFO vs LRU</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>LRU vs PLRU</td>
<td>1</td>
<td>0</td>
<td>–</td>
<td>2</td>
<td>1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PLRU vs LRU</td>
<td>1</td>
<td>0</td>
<td>–</td>
<td>$\infty$</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FIFO vs PLRU</td>
<td>2</td>
<td>1</td>
<td>–</td>
<td>4</td>
<td>4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PLRU vs FIFO</td>
<td>2</td>
<td>1</td>
<td>–</td>
<td>$\infty$</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Associativity:</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRU vs FIFO</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
</tr>
<tr>
<td>FIFO vs LRU</td>
<td>$\frac{1}{2}, \frac{1}{2}$</td>
<td>$\frac{1}{2}, 1$</td>
<td>$\frac{1}{2}, \frac{3}{2}$</td>
<td>$\frac{1}{2}, 2$</td>
<td>$\frac{1}{2}, \frac{5}{2}$</td>
<td>$\frac{1}{2}, 3$</td>
<td>$\frac{1}{2}, \frac{7}{2}$</td>
</tr>
<tr>
<td>LRU vs PLRU</td>
<td>1,0</td>
<td>–</td>
<td>$\frac{1}{2}, 1$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>$\frac{1}{8}, \frac{15}{8}$</td>
</tr>
<tr>
<td>PLRU vs LRU</td>
<td>1,0</td>
<td>–</td>
<td>$\frac{1}{2}, 1$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>$\frac{1}{4}, \frac{19}{4}$</td>
</tr>
<tr>
<td>FIFO vs PLRU</td>
<td>$\frac{1}{2}, \frac{1}{2}$</td>
<td>–</td>
<td>$\frac{1}{4}, \frac{5}{4}$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>$\frac{1}{11}, \frac{19}{11}$</td>
</tr>
<tr>
<td>PLRU vs FIFO</td>
<td>0,0</td>
<td>–</td>
<td>0,0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0,0</td>
</tr>
</tbody>
</table>

Jan Reineke, Daniel Grund
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,
  \[\rightarrow\] LRU-\emph{must}-analysis can be used for PLRU
Generalizations

Identified patterns and proved generalizations by hand.

Previously unknown facts:

- \( \text{PLRU}(k) \) is \( 1 \) comp. rel. to \( \text{LRU}(1 + \log_2 k) \) with constant 0,

  \( \rightarrow \) LRU-*must*-analysis can be used for PLRU

- \( \text{FIFO}(k) \) is \( \frac{1}{2} \) hit-comp. rel. to \( \text{LRU}(k) \),
Identified patterns and proved generalizations by hand.

Previously unknown facts:

- PLRU($k$) is $1$ comp. rel. to $\text{LRU}(1 + \log_2 k)$ with constant $0$,
  \[ \rightarrow \text{LRU-must-analysis can be used for PLRU} \]

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

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

Identified patterns and proved generalizations by hand.

Previously unknown facts:

- \( \text{PLRU}(k) \) is 1 comp. rel. to \( \text{LRU}(1 + \log_2 k) \) with constant 0, \( \rightarrow \) \text{LRU-}must-analysis can be used for PLRU

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

- \( \text{LRU}(k) \) is 0 hit-comp. rel. to \( \text{FIFO}(k) \), but

- \( \text{LRU}(2k - 1) \) is 1 comp. rel. to \( \text{FIFO}(k) \) with constant 0, \( \rightarrow \) \text{LRU-}may-analysis can be used for FIFO
Outline

1. Introduction
   - Motivation
   - Approach

2. Relative Competitiveness
   - Definition
   - Automatic Computation

3. Results
   - Automatically Computed
   - Generalizations

4. 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!
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. 

Christian Ferdinand, Florian Martin, and Reinhard Wilhelm. 
Applying compiler techniques to cache behavior prediction. 

Christian Ferdinand and Reinhard Wilhelm. 
Efficient and precise cache behavior prediction for real-time systems. 

Somnath Ghosh, Margaret Martonosi, and Sharad Malik.
Precise miss analysis for program transformations with caches of arbitrary associativity.


Harini Ramaprasad and Frank Mueller.
Bounding worst-case data cache behavior by analytically deriving cache reference patterns.


Timing analysis for data caches and set-associative caches.