Cache Related Preemption Delay Computation for Set-Associative Caches
Pitfalls and Solutions

Claire Burguière, Jan Reineke, Sebastian Altmeyer

Workshop on WCET Analysis, Dublin 2009
Preemptive scheduling

Cache related preemption delay (CRPD):
- Impact of preemption on the cache content
- Overall cost of additional reloads due to preemption

\[ T_1 \quad \uparrow \quad CRPD \]

\[ T_2 \quad \uparrow \quad Task\ Activation \]
CRPD computation for set-associative caches

- CRPD computation:
  - Useful Cache Blocks (UCB)
  - Evicting Cache Blocks (ECB)

- CRPD for set-associative caches:
  ⇒ Pitfalls:
  - LRU: CRPD not bounded by the number of ECBs
  - FIFO and PLRU: CRPD not bounded
Useful Cache Block - [Lee et al., 1996]

Definition (Useful Cache Block)

A memory block $m$ at program point $P$ is called a useful cache block, if

a) $m$ may be cached at $P$

b) $m$ may be reused at program point $P'$ that may be reached from $P$ with no eviction of $m$ on this path.

\[ \begin{array}{cccccc}
A & | & B & | & D & | & C \\
\times & | & \circ & | & \times & | & \times \\
& & & & & & \\
\end{array} \]

Cache Content: $[A, B, C, D]$

$x$ = hit  \hspace{1cm}  o = miss

\[ \begin{align*}
\text{CRPD} & = \sum_{s} \text{s} \\
\text{CRPD} & = \text{LCU} \\
\text{CRPD} & = \text{BRT} \cdot \min(|\text{LCU}|, n)
\end{align*} \]

Burguière, Reineke, Altmeyer  
CRPD for set-associative caches  
WCET, Dublin 2009  
4 / 13
Useful Cache Block - [Lee et al., 1996]

Definition (Useful Cache Block)

A memory block $m$ at program point $P$ is called a useful cache block, if

a) $m$ may be cached at $P$

b) $m$ may be reused at program point $P'$ that may be reached from $P$ with no eviction of $m$ on this path.

$\times =$ hit
$\circ =$ miss

Cache Content: $[A, B, C, D]$
Useful Cache Block - [Lee et al., 1996]

**Definition (Useful Cache Block)**

A memory block $m$ at program point $P$ is called a useful cache block, if

a) $m$ may be cached at $P$

b) $m$ may be reused at program point $P'$ that may be reached from $P$ with no eviction of $m$ on this path.

$\text{CRPD}_{\text{LRU}}(s) = \sum_{s=1}^{c} \text{CRPD}_{\text{LRU}}(s)$

$\text{CRPD}_{\text{UCB}}(s) = \text{BRT} \cdot \min(|\text{UCB}(s)|, n)$

$\text{BRT} = \text{Block Reload Time}$

$\text{hit}$

$\text{miss}$
Evicting Cache Blocks
[Tomiyama and Dutt, 2000]

Definition (Evicting Cache Blocks (ECB))

A memory block of the preempting task is called an evicting cache block, if it may be accessed during the execution of the preempting task.

![Diagram showing cache content and additional miss due to preemption (CRPD)]
Evicting Cache Blocks
[Tomiyama and Dutt, 2000]

Definition (Evicting Cache Blocks (ECB))

A memory block of the preempting task is called an evicting cache block, if it may be accessed during the execution of the preempting task.

\[
\text{CRPD}^{\text{LRU}}_{\text{ECB}}(s) \equiv \text{BRT} \cdot \min(|\text{ECB}(s)|, n)
\]
CRPD computation for LRU using ECB: Pitfall

\[ [b, a, 9, 8] \xrightarrow{8} [8, b, a, 9] \xrightarrow{9} [9, 8, b, a] \xrightarrow{a} [a, 9, 8, b] \xrightarrow{b} [b, a, 9, 8] \quad 0 \text{ misses} \]
CRPD computation for LRU using ECB: Pitfall

ECBs = \{e\}

\[\begin{align*}
[b, a, 9, 8] & \xrightarrow{8} [8, b, a, 9] \xrightarrow{9} [9, 8, b, a] \xrightarrow{a} [a, 9, 8, b] \xrightarrow{b} [b, a, 9, 8] & 0 \text{ misses} \\
[e, b, a, 9] & \xrightarrow{8^*} [8, e, b, a] \xrightarrow{9^*} [9, 8, e, b] \xrightarrow{a^*} [a, 9, 8, e] \xrightarrow{b^*} [b, a, 9, 8] & 4 \text{ misses}
\end{align*}\]

\[\begin{align*}
|\text{UCB}(s)| &= 4 \\
|\text{ECB}(s)| &= 1 \\
n &= 4 \\
\text{number of additional misses} &= 4
\]
LRU: new computation of CRPD using ECB

- ECB derivation used only to know if the set is used by the preempting task:

\[
\text{CRPD}_{\text{ECB}}^{\text{LRU}}(s) = \begin{cases} 
0 & \text{if } \text{ECB}(s) = \emptyset \\
\text{BRT} \cdot n & \text{otherwise}
\end{cases}
\]
CRPD for FIFO: Pitfalls

\[
[b, a] \xrightarrow{a} [b, a] \xrightarrow{e^*} [e, b] \xrightarrow{b} [e, b] \xrightarrow{c^*} [c, e] \xrightarrow{e} [c, e] \quad 2 \text{ misses}
\]
CRPD for FIFO: Pitfalls

\[
\begin{align*}
\text{ECBs} &= \{x\} \\
\text{[b, a]} &\xrightarrow{a} [b, a] \xrightarrow{e^*} [e, b] \xrightarrow{b} [e, b] \xrightarrow{c^*} [c, e] \xrightarrow{e} [c, e] & \text{2 misses} \\
\text{[x, b]} &\xrightarrow{a^*} [a, x] \xrightarrow{e^*} [e, a] \xrightarrow{b^*} [b, e] \xrightarrow{c^*} [c, b] \xrightarrow{e^*} [e, c] & \text{5 misses}
\end{align*}
\]

- \(|\text{UCB}(s)| = 2|
- \(|\text{ECB}(s)| = 1|
- \(n = 2|
- \text{But: number of additional misses} = 3|
**Notation**

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

**Definition (Relative miss competitiveness)**

Policy \( P \) is \((k, c)\)-miss-competitive relative to policy \( Q \), 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 \( p, q \).
Definition – Relative Miss-Competitiveness

Notation

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

Definition (Relative miss competitiveness)

Policy \( P \) is \((k, c)\)-miss-competitive relative to policy \( Q \), 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 \( p, q \).

- PLRU\((n)\) is \((1, 0)\)-miss-competitive relative to LRU\((1 + \log_2 n)\).
- FIFO\((n)\) is \((\frac{n}{n-r+1}, r)\)-miss-competitive relative to LRU\((r)\).
A sequence of memory accesses

- **Notation:**
  - \( m \) = number of misses
  - \( \overline{m} \) = number of misses in the case of preemption

\[
\begin{align*}
  m_{\text{pre}} &= 4 \\
  m_{\text{post}} &= 2 \\
  \overline{m}_{\text{pre}} &= m_{\text{pre}} = 4 \\
  m_{\text{post}} &= m_{\text{post}} + m_{\text{CRPD}} = 5
\end{align*}
\]
A sequence of memory accesses

- Notation:
  - $m$ = number of misses
  - $\bar{m}$ = number of misses in the case of preemption

Relative miss competitiveness:

$$\bar{m}^{(t)} = \bar{m}_{pre}^{(t)} + \bar{m}_{post}^{(t)}$$
A sequence of memory accesses

- **Notation:**
  - \( m \) = number of misses
  - \( \overline{m} \) = number of misses in the case of preemption

\[
\overline{m} = m_{\text{pre}} + m_{\text{post}}
\]

\[
\overline{m} = m_{\text{pre}} + m_{\text{post}} + m_{\text{CRPD}} = 5
\]

\[
\overline{m} = m_{\text{pre}} = 4
\]

\[
\overline{m} = m_{\text{pre}} = 4
\]

- **Relative miss competitiveness:**

\[
\overline{m}^{(t)} = \overline{m}_{\text{pre}}^{(t)} + \overline{m}_{\text{post}}^{(t)}
\]

\[
\leq [k \cdot m_{\text{pre}}^{LRU(s)} + c] + [k \cdot (m_{\text{post}}^{LRU(s)} + m_{\text{CRPD}}^{LRU(s)}) + c]
\]
A sequence of memory accesses

- Notation:
  - \( m \) = number of misses
  - \( \overline{m} \) = number of misses in the case of preemption

\[
\begin{align*}
\overline{m}_{pre} &= 4 \\
\overline{m}_{post} &= 2 \\
\overline{m}_{pre} &= \overline{m}_{pre} = 4 \\
\overline{m}_{post} &= \overline{m}_{post} + \overline{m}_{CRPD} = 5
\end{align*}
\]

- Relative miss competitiveness:
\[
\overline{m}^{(t)} = \overline{m}^{(t)}_{pre} + \overline{m}^{(t)}_{post} 
\leq [k \cdot m^{LRU(s)}_{pre} + c] + [k \cdot (m^{LRU(s)}_{post} + m^{LRU(s)}_{CRPD}) + c] 
= [k \cdot (m^{LRU(s)}_{pre} + m^{LRU(s)}_{post}) + c] + [k \cdot m^{LRU(s)}_{CRPD} + c]
\]
A sequence of memory accesses

Notation:

- $m$ = number of misses
- $\overline{m}$ = number of misses in the case of preemption

\[ \begin{align*}
\overline{m}_\text{pre} &= 4 \\
\overline{m}_\text{post} &= 2
\end{align*} \]

Relative miss competitiveness:

\[ \overline{m}^{(t)} = \overline{m}_\text{pre}^{(t)} + \overline{m}_\text{post}^{(t)} \]
\[ \leq \left[ k \cdot m^{\text{LRU}(s)}_\text{pre} + c \right] + \left[ k \cdot \left( m^{\text{LRU}(s)}_\text{post} + m^{\text{LRU}(s)}_{\text{CRPD}} \right) + c \right] \]
\[ = \left[ k \cdot \left( m^{\text{LRU}(s)}_\text{pre} + m^{\text{LRU}(s)}_\text{post} \right) + c \right] + \left[ k \cdot m^{\text{LRU}(s)}_{\text{CRPD}} + c \right] \]
\[ = \left[ k \cdot m^{\text{LRU}(s)} + c \right] + \left[ k \cdot m^{\text{LRU}(s)}_{\text{CRPD}} + c \right] \]
Relative Competitiveness – Application

- PLRU(8) using LRU(4):

\[ m_{\text{PLRU}(8)} \leq m_{\text{LRU}(4)} + m_{\text{CRPD}} \]

- FIFO(8) using LRU(5):

\[ m_{\text{FIFO}(8)} \leq (2 \cdot m_{\text{LRU}(5)} + 5) + (2 \cdot m_{\text{CRPD}} + 5) \]
## Conclusions

### Pitfalls

<table>
<thead>
<tr>
<th>Pitfalls</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LRU:</strong></td>
<td><strong>LRU:</strong></td>
</tr>
<tr>
<td>$</td>
<td>ECBs</td>
</tr>
<tr>
<td><strong>FIFO and PLRU:</strong></td>
<td>$\Rightarrow \text{the set is not used}$</td>
</tr>
<tr>
<td>$</td>
<td>UCBs</td>
</tr>
<tr>
<td>$\text{do not bound the number of additional misses}$</td>
<td>$\text{using relative competitiveness and LRU bounds}$</td>
</tr>
</tbody>
</table>
Further reading

A New Notion of Useful Cache Block to Improve the Bounds of Cache-Related Preemption Delay.

Analysis of cache-related preemption delay in fixed-priority preemptive scheduling.
In RTSS’96 p. 264, IEEE Computer Society.

Accurate estimation of cache-related preemption delay.
In CODES+ISSS’03 ACM.

Caches in WCET Analysis.

Scalable precision cache analysis for real-time software.
ACM TECS 6, 25.

Integrated intra- and inter-task cache analysis for preemptive multi-tasking real-time systems.
In SCOPES’04 pp. 182–199.

Upper-bound on the CRPD - direct-mapped caches

- using UCB [Lee et al., 1996]:

\[
\text{CRPD}_{\text{UCB}} = \text{BRT} \cdot |\{s_i \mid \exists m \in \text{UCB} : m \mod c = s_i\}|
\]

- using ECB [Tomiyama and Dutt, 2000]:

\[
\text{CRPD}_{\text{ECB}} = \text{BRT} \cdot |\{s_i \mid \exists m \in \text{ECB} : m \mod c = s_i\}|
\]

- using UCB and ECB [Negi et al., 2003, Tan and Mooney, 2004]:

\[
\text{CRPD}_{\text{UCB}\&\text{ECB}} = \text{BRT} \cdot |\{s_i \mid \exists m \in \text{UCB} : m \mod c = s_i \\
\land \exists m' \in \text{ECB} : m' \mod c = s_i\}|
\]
Relative Competitiveness – Example

\[
\bar{m}_{pre}^{(t)} \leq k \cdot \bar{m}_{pre}^{LRU(s)} + c = k \cdot m_{pre}^{LRU(s)} + c
\]

\[
\bar{m}_{post}^{(t)} \leq k \cdot \bar{m}_{post}^{LRU(s)} + c = k \cdot (m_{post}^{LRU(s)} + m_{CRPD}^{LRU(s)}) + c
\]

\[
\bar{m}^{(t)} = \bar{m}_{pre}^{(t)} + \bar{m}_{post}^{(t)} \leq k \cdot m_{pre}^{LRU(s)} + c + k \cdot (m_{post}^{LRU(s)} + m_{CRPD}^{LRU(s)}) + c
\]

\[
= (k \cdot (m_{pre}^{LRU(s)} + m_{post}^{LRU(s)}) + c) + (k \cdot m_{CRPD}^{LRU(s)} + c)
\]

\[
= (k \cdot m^{LRU(s)} + c) + (k \cdot m_{CRPD}^{LRU(s)} + c).
\]
Relative Competitiveness – Application

- **PLRU\((n)\) using LRU\((1 + \log_2 n)\):**
  - \(k = 1, c = 0\)

\[
\overline{m}^{\text{PLRU}(n)} \leq \overline{m}^{\text{LRU}(1+\log_2 n)} + m^{\text{LRU}(1+\log_2 n)}
\]

  - Example (n=8):

\[
\overline{m}^{\text{PLRU}(8)} \leq \overline{m}^{\text{LRU}(4)} + m^{\text{LRU}(4)}
\]

- **FIFO\((n)\) using LRU\((r)\):**
  - \(k = \frac{n}{n-r+1}, c = r\)

\[
\overline{m}^{\text{FIFO}(n)} \leq \left(\frac{n}{n-r+1} \cdot m^{\text{LRU}(r)} + r\right) + \left(\frac{n}{n-r+1} \cdot m^{\text{LRU}(r)} + r\right)
\]

  - Example (n=8,r=5):

\[
\overline{m}^{\text{FIFO}(8)} \leq \left(2 \cdot m^{\text{LRU}(5)} + 5\right) + \left(2 \cdot m^{\text{LRU}(5)} + 5\right)
\]
CRPD for PLRU: Pitfalls

- |UCB(s)| = 4
- |ECB(s)| = 2
- \( n = 4 \)
- But: number of additional misses = 5