Cache Related Preemption Delay for Set-Associative Caches
Resilience Analysis

Sebastian Altmeyer, Claire Burguière, Jan Reineke

Predator Meeting, Pisa 2010
Context

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

\[ T_1 \quad \uparrow \quad T_2 \]

\[ \text{□} = \text{CRPD} \]

\[ \uparrow = \text{Task Activation} \]
CRPD for set-associative caches - LRU

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

- CRPD from UCB and ECB:
  - Previous combination overestimates
  \[\Rightarrow\] Some UCBs remain useful under preemption
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.

\[
\begin{align*}
\times &= \text{hit} \\
\bigcirc &= \text{miss}
\end{align*}
\]

Cache Content: \([A, B, C, D]\)

\[
\begin{align*}
\text{CRPD}_{\text{UCB}} &= \sum_{s=1}^{c} \text{CRPD}_{\text{UCB}}^{s} \\
\text{CRPD}_{\text{UCB}}^{s} &= \text{BRT} \times \min(|\text{UCB}(s)|, n)
\end{align*}
\]

\( n = \text{associativity} \)

\( \text{BRT} = \text{Block Reload Time} \)
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.

\[ \text{CRPD}_{\text{UCB}} = \sum_{s=1}^{c} \text{CRPD}_{\text{UCB}}^{s} \]

\[ \text{CRPD}_{\text{UCB}}^{s} = \text{BRT} \times \min(|\text{UCB}(s)|, n) \]

$\times$ = hit

$\bigcirc$ = miss

$P$

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

$n$ = associativity

BRT = Block Reload Time
Evicting Cache Blocks
[Tomiyama & 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.

\[
\text{Cache Content: } [A, B, C, D] \quad \Rightarrow \quad \text{Cache Content: } [X, Y, Z, D]
\]

- \( \bullet \) = additional miss due to preemption (CRPD)

\[
\text{CRPD}^s_{\text{ECB}} = \begin{cases} 
0 & \text{if } \text{ECB}(s) = \emptyset \\
\text{BRT} \times n & \text{otherwise}
\end{cases}
\]
Impact of the preempting task on the preempted task

CRPD (using UCB and ECB)

$$\text{CRPD}_{\text{UCB}\&\text{ECB}} = \sum_{s=1}^{c} \min(\text{CRPD}_{\text{UCB}}^{s}, \text{CRPD}_{\text{ECB}}^{s})$$
Impact of the preemtting task on the preempted task (example)

ECBs = \{e\}

\begin{align*}
[c, b, a, x] & \xrightarrow{a} [a, c, b, x] \xrightarrow{b} [b, a, c, x] \xrightarrow{c} [c, b, a, x] \\
[e, c, b, a] & \xrightarrow{a} [a, e, c, b] \xrightarrow{b} [b, a, e, c] \xrightarrow{c} [c, b, a, e]
\end{align*}

no miss

\begin{align*}
[c, b, a, x] & \xrightarrow{a} [a, c, b, x] \xrightarrow{b} [b, a, c, x] \xrightarrow{c} [c, b, a, x] \\
[e, c, b, a] & \xrightarrow{a} [a, e, c, b] \xrightarrow{b} [b, a, e, c] \xrightarrow{c} [c, b, a, e]
\end{align*}

no miss

- CRPD_{UCB} \Rightarrow |UCB| = 3
- CRPD_{ECB} \Rightarrow n = 4
- CRPD_{UCB\&ECB} = \min(CRPD_{UCB}, CRPD_{ECB}) \Rightarrow 3
  - Overestimation: number of additional misses\(= 0 < 3\)

Why?
- |ECB| to evict a UCB = 2
- but, |ECB| = 1
- One single ECB is not sufficient to evict a UCB
Refinement

Determining $\max_{\text{ECB}}$ s.t. no additional cache miss occur

$m \in UCB$

$m$ is 4-resilient ($\text{res}(m) = 4$)

\[
\begin{align*}
    m &\quad [m, \_, \_, \_, \_, \_, \_, \_]
\end{align*}
\]

\[
\begin{align*}
    a_1 &\quad [a_3, a_2, a_1, m, \_, \_, \_, \_]
\end{align*}
\]

\[
\begin{align*}
    a_2 &\quad [m, a_3, a_2, a_1, \_, \_, \_, \_]
\end{align*}
\]

\[
\begin{align*}
    a_3 &\quad [m, a_3, a_2, a_1, \_, \_, \_, \_]
\end{align*}
\]
Resilience analysis

Definition (Resilience)

The resilience \( \text{res}_P(m) \) of memory block \( m \) at program point \( P \) is the greatest \( l \), such that all possible next accesses to \( m \),

a) that would be hits without preemption,

b) would still be hits in case of a preemption with \( l \) accesses at \( P \).

\[ m \in \text{UCB} \]
\[ \text{res}(m) = 4 \]

If \( |\text{ECB}| \leq l \) then the UCB is not evicted.
CRPD using resilience

CRPD (combining UCB and ECB by using resilience)

\[ CRPD \leq BRT \times \left| \left\{ m \mid \text{res}(m) = \left| \text{ECB} \right| \right\} \right| \]

blocks contributing to CRPD

UCB \setminus \{m \mid \text{useful} \}

remain useful
CRPD using resilience - example

- $|ECB| = 1$
- $res(a) = res(b) = res(c) = 1$
- $CRPD_{\text{UCB}&\text{ECB}}^{\text{res}} = \text{BRT} \times |UCB \setminus \{m \mid res(m) = |ECB| \}| = 0$
- Instead of: $CRPD_{\text{UCB}&\text{ECB}} = min(CRPD_{\text{UCB}}, CRPD_{\text{ECB}}) = 3 \times \text{BRT}$
Evaluation Setting

- Cachesize 8KB
- 8 ways
- 32 sets
- linesize 32 bytes
- LRU caches
- Testcases: Mälardalen benchmark suite:
### Evaluation Benchmarks

<table>
<thead>
<tr>
<th>Task</th>
<th>Code Size</th>
<th>Cache Util.</th>
<th>UCB</th>
</tr>
</thead>
<tbody>
<tr>
<td>minmax</td>
<td>608B</td>
<td>7.4%</td>
<td>4</td>
</tr>
<tr>
<td>insertsort</td>
<td>384B</td>
<td>4.7%</td>
<td>5</td>
</tr>
<tr>
<td>fibcall</td>
<td>256B</td>
<td>3.1%</td>
<td>5</td>
</tr>
<tr>
<td>fac</td>
<td>256B</td>
<td>3.1%</td>
<td>6</td>
</tr>
<tr>
<td>bs</td>
<td>320B</td>
<td>3.9%</td>
<td>8</td>
</tr>
<tr>
<td>bsort100</td>
<td>544B</td>
<td>6.6%</td>
<td>10</td>
</tr>
<tr>
<td>ns</td>
<td>576B</td>
<td>7%</td>
<td>11</td>
</tr>
<tr>
<td>matmult</td>
<td>864B</td>
<td>10.5%</td>
<td>12</td>
</tr>
<tr>
<td>fir</td>
<td>928B</td>
<td>11.3%</td>
<td>22</td>
</tr>
<tr>
<td>crc</td>
<td>1216B</td>
<td>14.8%</td>
<td>35</td>
</tr>
<tr>
<td>select</td>
<td>1280B</td>
<td>15.6%</td>
<td>37</td>
</tr>
<tr>
<td>qsort-exam</td>
<td>1440B</td>
<td>17.6%</td>
<td>42</td>
</tr>
<tr>
<td>sqrt</td>
<td>3680B</td>
<td>44.9%</td>
<td>101</td>
</tr>
<tr>
<td>qurt</td>
<td>4160B</td>
<td>50.8%</td>
<td>118</td>
</tr>
</tbody>
</table>
Evaluation

preempted by fibcall (#ECBs = 8)

preempted by qurt (#ECBs = 121)

resilience  tan  UCB & ECB  #UCBs
Conclusions

- UCB and ECB analyses:
  - pessimistic overapproximation of the CRPD

- Resilience analysis:
  - regain some precision
  - reduce pessimism

- Resilience analysis:
  - simple data-flow analyses
  - similar to UCB analysis for LRU
**Further reading**

- **Altmeyer, S. & Burguière, C. (2009).**

  In RTSS’96 p. 264, IEEE Computer Society.

- **Negi, H. S., Mitra, T. & Roychoudhury, A. (2003).**
  In CODES+ISSS’03 ACM.

- **Reineke, J. (2008).**
  Caches in WCET Analysis.

- **Staschulat, J. & Ernst, R. (2007).**
  ACM Trans. on Embedded Computing Sys. 6, 25.

- **Tan, Y. & Mooney, V. (2004).**
  In SCOPES’04 pp. 182–199.

- **Tomiyama, H. & Dutt, N. D. (2000).**
  In CODES’00 ACM.
/-resilience analysis

(a) 0-resilient

(b) m is not useful 2-resilient

m

m

m

m
CPRD using ECB: Pitfall

\[ [b, a, 9, 8] \xrightarrow{8} [8, b, a, 9] \xrightarrow{9} [9, 8, b, a] \xrightarrow{a} [a, 9, 8, b] \xrightarrow{b} [b, a, 9, 8] \]

- 0 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 misses

- \(|\text{UCB}(s)| = 4\)
- \(|\text{ECB}(s)| = 1\)
- \(n = 4\)
- number of additional misses = 4
Upper-bound on the CRPD - direct-mapped caches

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

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

- using ECB [Tomiyama & Dutt, 2000]:

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

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

\[ \text{CRPD}_{\text{UCB}\&\text{ECB}} = BRT \cdot \left| \{ s_i \mid \exists m \in \text{UCB} : m \mod c = s_i \right. \\
\left. \land \exists m' \in \text{ECB} : m' \mod c = s_i \} \right| \]
CRPD for FIFO: Pitfalls

ECBs

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

- \(|UCB(s)| = 2\)
- \(|ECB(s)| = 1\)
- \(n = 2\)
- But: number of additional misses = 3
CRPD for PLRU: Pitfalls

ECBs = \{x, y\}

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