Cache Related Preemption Delay for Set-Associative Caches

Resilience Analysis

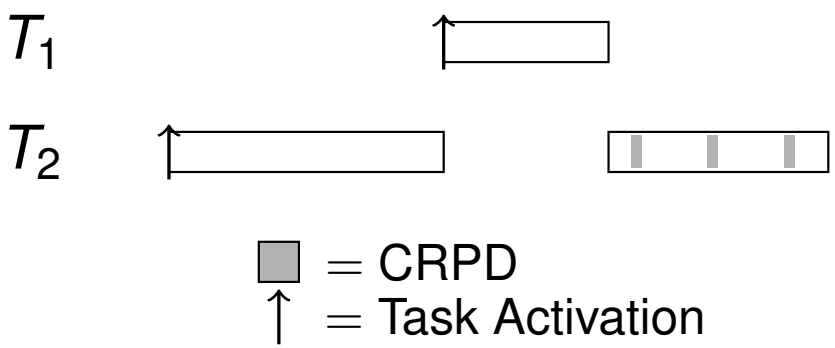
Sebastian Altmeyer, Claire Burguière, Jan Reineke

Predator Meeting, Pisa 2010

SAARLAND UNIVERSITY
COMPUTER SCIENCE

Context

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



CRPD for set-associative caches - LRU

SAARLAND UNIVERSITY
COMPUTER SCIENCE

- CRPD computation:
  - preempted task: Useful Cache Blocks (UCB)
  - preempting 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]

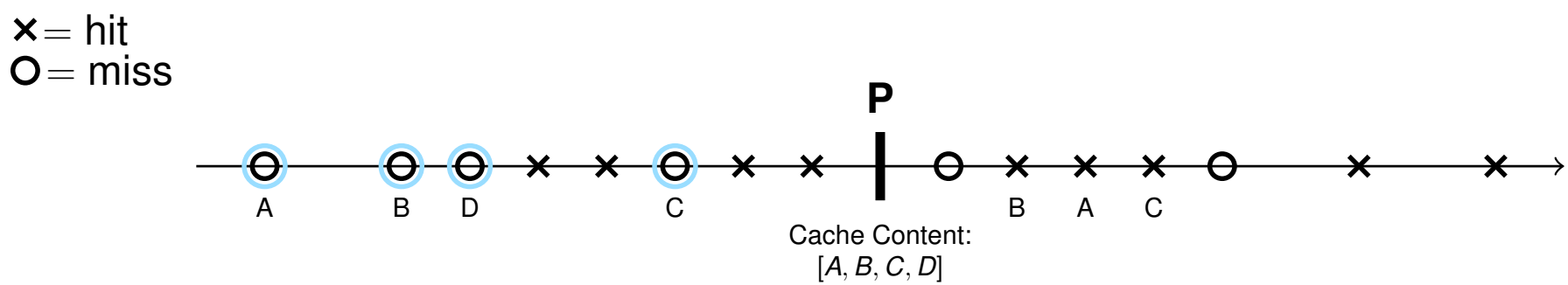
SAARLAND UNIVERSITY
COMPUTER SCIENCE

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.*



$$\mathrm{CRPD}_{\mathrm{UCB}} = \sum_{s=1}^{c} \mathrm{CRPD}^{s}_{\mathrm{UCB}}$$

$$\mathrm{CRPD}^{s}_{\mathrm{UCB}} = \mathrm{BRT} \times \mathit{min}(|\mathrm{UCB}(s)|, n)$$

n = associativity
BRT = Block Reload Time

Useful Cache Block - [Lee et al., 1996]

SAARLAND UNIVERSITY
COMPUTER SCIENCE

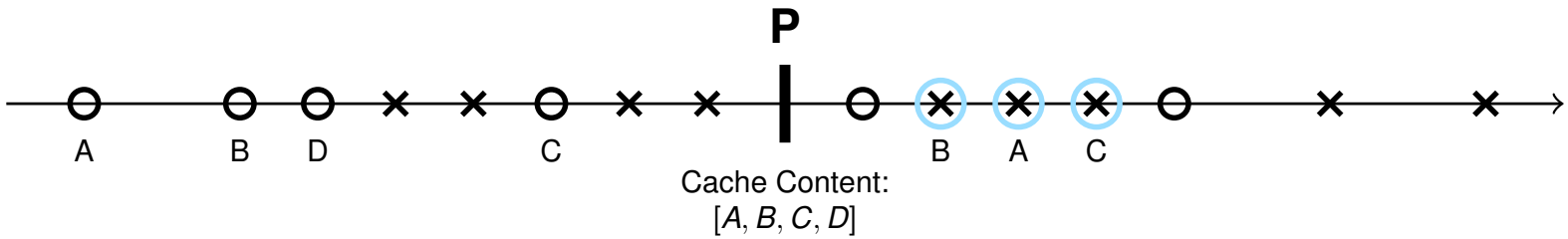
Definition (Useful Cache Block)

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.*

× = hit
O = miss



$$\mathrm{CRPD}_{\mathrm{UCB}} = \sum_{s=1}^{c} \mathrm{CRPD}^{s}_{\mathrm{UCB}}$$

$$\mathrm{CRPD}^{s}_{\mathrm{UCB}} = \mathrm{BRT} \times \mathit{min}(|\mathrm{UCB}(s)|, n)$$

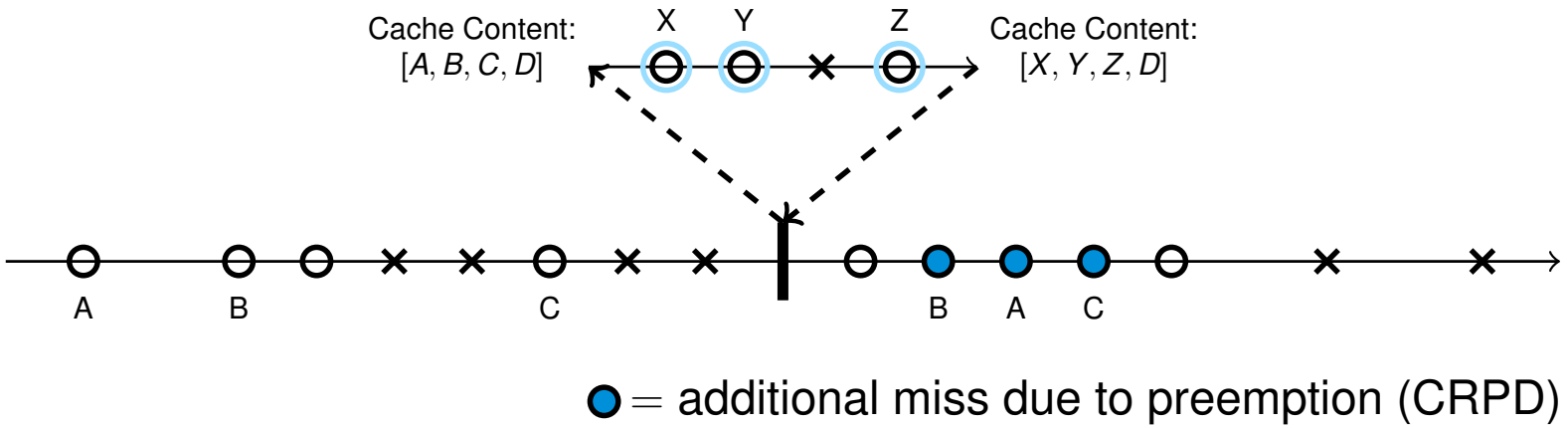
n = associativity
BRT = Block Reload Time

Evicting Cache Blocks [Tomiyama & Dutt, 2000]

SAARLAND UNIVERSITY
COMPUTER SCIENCE

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.



$$\mathrm{CRPD}^{s}_{\mathrm{ECB}} = \begin{cases} 0 & \text{if } \mathrm{ECB}(s) = \emptyset \\ \mathrm{BRT} \times n & \text{otherwise} \end{cases}$$

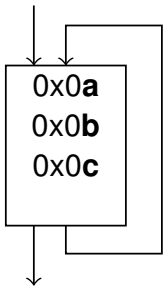
Impact of the preempting task on the preempted task

SAARLAND UNIVERSITY
COMPUTER SCIENCE

CRPD (using UCB and ECB)

$$CRPD_{UCB\&ECB} = \sum_{s=1}^{c} min(CRPD^{s}_{UCB}, CRPD^{s}_{ECB})$$

Impact of the preempting task on the preempted task (example)



ECBs $= \{\mathbf{e}\}$

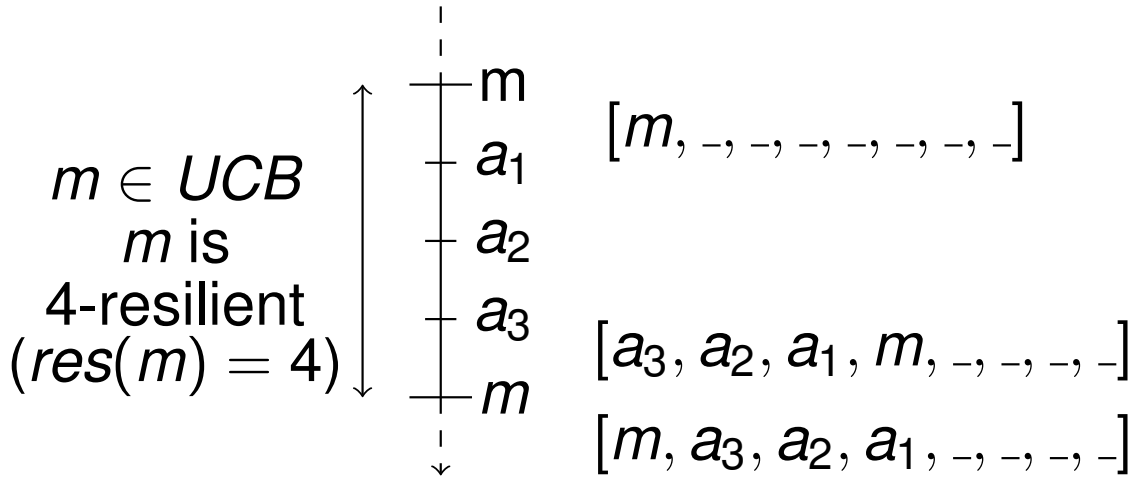
$[c, b, a, x] \xrightarrow{a} [a, c, b, x] \xrightarrow{b} [b, a, c, x] \xrightarrow{c} [c, b, a, x]$ no miss

$[\mathbf{e}, c, b, a] \xrightarrow{a} [a, \mathbf{e}, c, b] \xrightarrow{b} [b, a, \mathbf{e}, c] \xrightarrow{c} [c, b, a, \mathbf{e}]$ no miss

- $\text{CRPD}_{\text{UCB}} \Rightarrow |\text{UCB}| = 3$
- $\text{CRPD}_{\text{ECB}} \Rightarrow n = 4$
- $\text{CRPD}_{\text{UCB\&ECB}} = min(\text{CRPD}_{\text{UCB}}, \text{CRPD}_{\text{ECB}}) \Rightarrow 3$
  - Overestimation: number of additional misses$= 0 < 3$
- Why?
  - $|\text{ECB}|$ to evict a UCB $= 2$
  - but, $|\text{ECB}| = 1$
  - One single ECB is not sufficient to evict a UCB

Refinement

Determining $max|_{\text{ECB}}|$ s.t. no additional cache miss occur



Resilience analysis

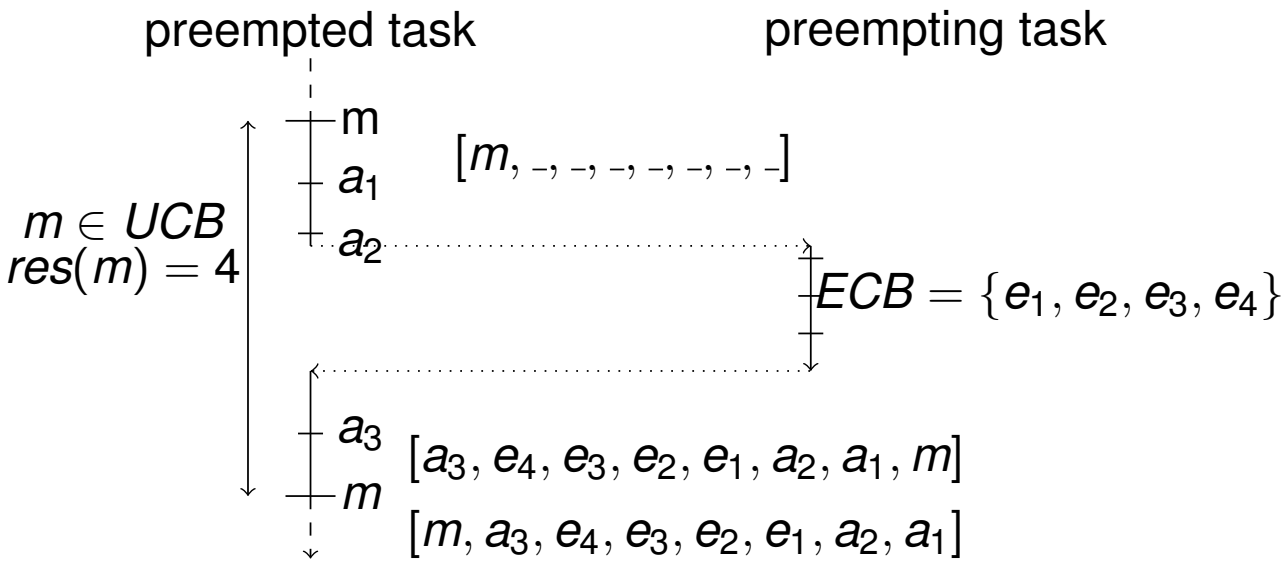
SAARLAND UNIVERSITY
COMPUTER SCIENCE

Definition (Resilience)

The resilience $res_P(m)$ 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.*



if $|\text{ECB}| \leq l$ then the UCB is not evicted

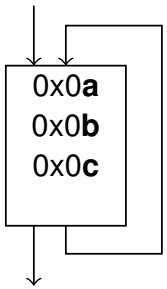
CRPD using resilience

SAARLAND UNIVERSITY
COMPUTER SCIENCE

CRPD (combining UCB and ECB by using resilience)

$$CRPD \leq BRT \times |\overbrace{UCB}^{\text{useful}} \setminus \overbrace{\{m \mid res(m) = |ECB|\}}^{\text{remain useful}}| \quad \text{(blocks contributing to CRPD)}$$

CRPD using resilience - example



ECBs $= \{\mathbf{e}\}$

$[c, b, a, x] \xrightarrow{a} [a, c, b, x] \xrightarrow{b} [b, a, c, x] \xrightarrow{c} [c, b, a, x]$ no miss

$[\mathbf{e}, c, b, a] \xrightarrow{a} [a, \mathbf{e}, c, b] \xrightarrow{b} [b, a, \mathbf{e}, c] \xrightarrow{c} [c, b, a, \mathbf{e}]$ no miss

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

Evaluation Setting

SAARLAND UNIVERSITY
COMPUTER SCIENCE

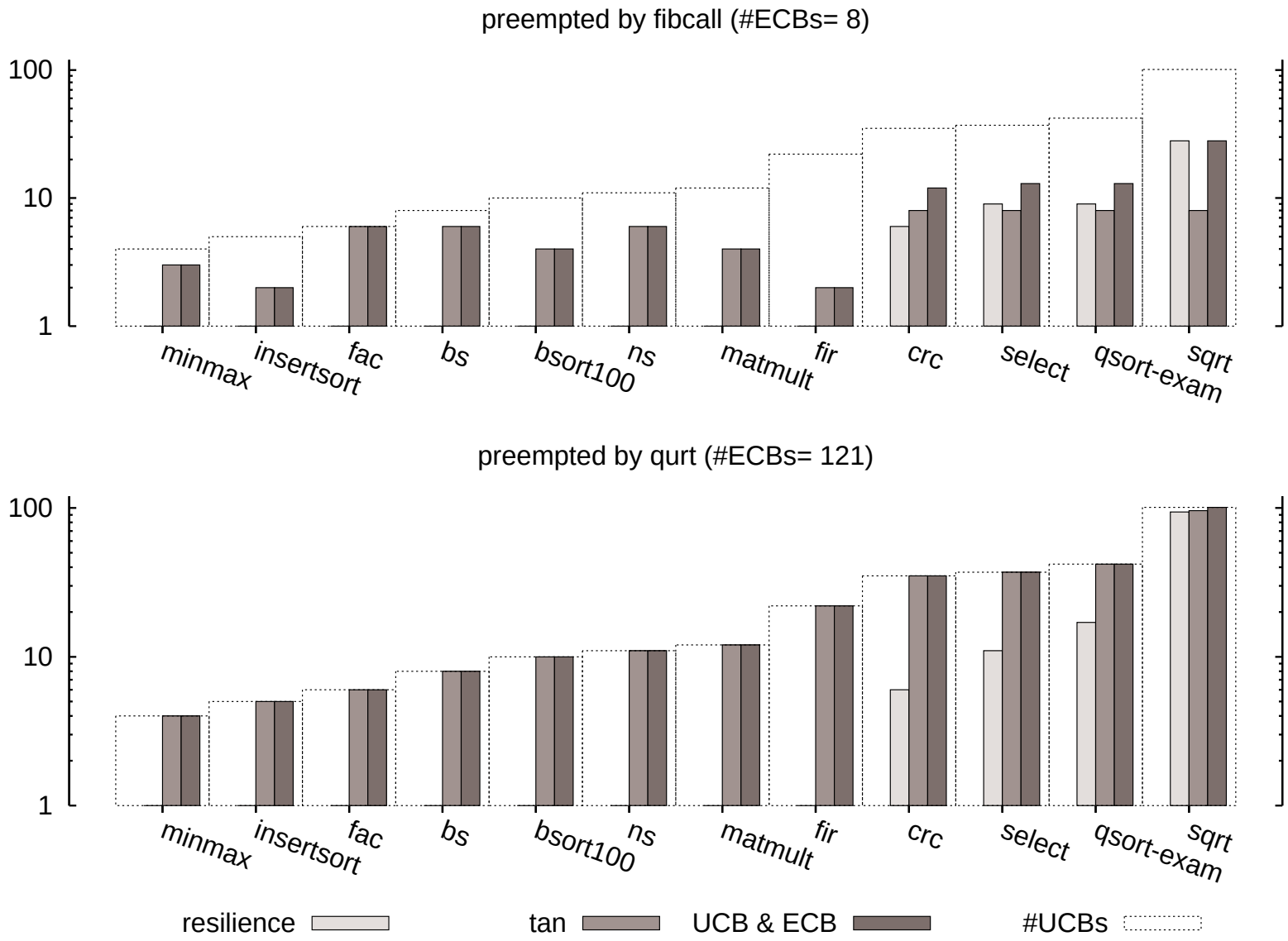
- Cachesize 8KB
- 8 ways
- 32 sets
- linesize 32 bytes
- LRU caches
- Testcases: Mälardalen benchmark suite:

Evaluation Benchmarks

SAARLAND UNIVERSITY
COMPUTER SCIENCE

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

Evaluation


preempted by fibcall (#ECBs= 8)
100
10
1
minmax
insertsort
fac
bs
bsort100
ns
matmult
fir
crc
select
qsort-exam
sqrt
preempted by qurt (#ECBs= 121)
100
10
1
minmax
insertsort
fac
bs
bsort100
ns
matmult
fir
crc
select
qsort-exam
sqrt
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 Proceedings of the 21st Euromicro Conference on Real-Time Systems (ECRTS '09) pp. 109–118, IEEE Computer Society.

Lee, C.-G., Hahn, J., Min, S. L., Ha, R., Hong, S., Park, C. Y., Lee, M. & Kim, C. S. (1996).
In RTSS'96 p. 264, IEEE Computer Society.

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

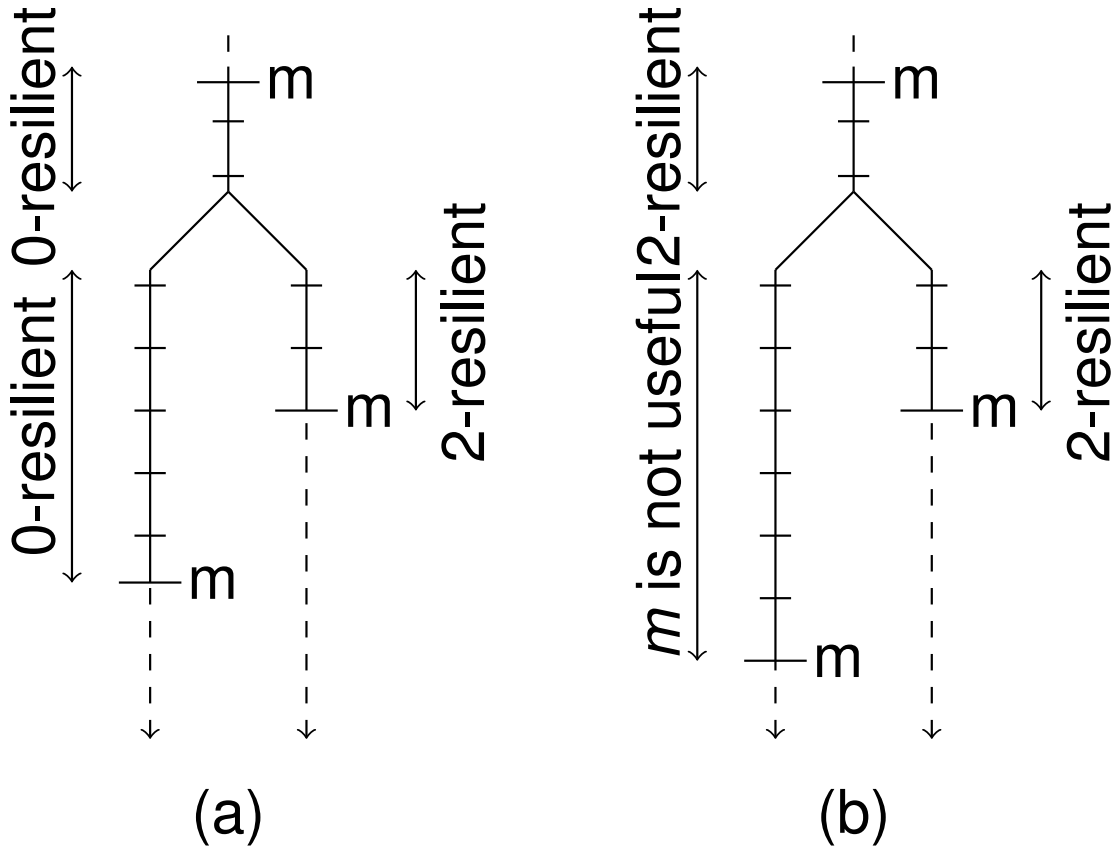
Reineke, J. (2008).
Caches in WCET Analysis.
PhD thesis, Universität des Saarlandes, Saarbrücken.

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

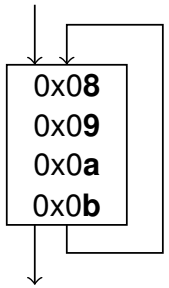
Tan, Y. & Mooney, V. (2004).
In SCOPES'04 pp. 182–199,.

Tomiyama, H. & Dutt, N. D. (2000).
In CODES'00 ACM.

/-resilience analysis



CPRD using ECB: Pitfall



ECBs $= \{\mathbf{e}\}$

$[b, a, 9, 8] \xrightarrow{8} [8, b, a, 9] \xrightarrow{9} [9, 8, b, a] \xrightarrow{a} [a, 9, 8, b] \xrightarrow{b} [b, a, 9, 8]$ 0 misses

$[\mathbf{e}, b, a, 9] \xrightarrow{8^*} [8, \mathbf{e}, b, a] \xrightarrow{9^*} [9, 8, \mathbf{e}, b] \xrightarrow{a^*} [a, 9, 8, \mathbf{e}] \xrightarrow{b^*} [b, a, 9, 8]$ 4 misses

- $|\mathrm{UCB}(s)| = 4$
- $|\mathrm{ECB}(s)| = 1$
- $n = 4$
- number of additional misses$= 4$

Upper-bound on the CRPD - direct-mapped caches

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

$$\mathrm{CRPD}_{\mathrm{UCB}} = \mathrm{BRT} \cdot |\{s_i \mid \exists m \in \mathrm{UCB} : m \bmod c = s_i\}|$$

- using ECB [Tomiyama & Dutt, 2000]:

$$\mathrm{CRPD}_{\mathrm{ECB}} = \mathrm{BRT} \cdot |\{s_i \mid \exists m \in \mathrm{ECB} : m \bmod c = s_i\}|$$

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

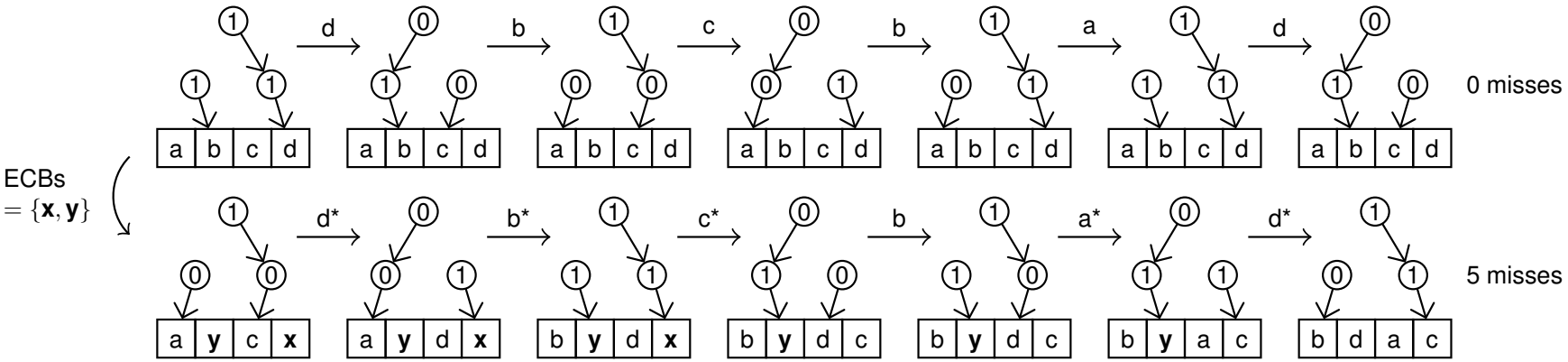
$$\begin{aligned}\mathrm{CRPD}_{\mathrm{UCB\&ECB}} = \mathrm{BRT} \cdot |\{s_i \mid\ & \exists m \in \mathrm{UCB} : m \bmod c = s_i \\ & \wedge \exists m' \in \mathrm{ECB} : m' \bmod c = s_i\}|\end{aligned}$$

CRPD for FIFO: Pitfalls

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

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

CRPD for PLRU: Pitfalls



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