Locality Approximation Using Time

Xipeng Shen
College of William and Mary
Williamsburg, VA, USA

Jonathan Shaw
Shaw Technologies
Portland, OR, USA

Brian Meeker and Chen Ding
University of Rochester
Rochester, NY, USA
Introduction

- Heterogeneous cache sharing

![Diagram of cache sharing structure]
Introduction

- Heterogeneous cache sharing

Locality analysis should be flexible to handle different caches and contentions
Introduction

- Contentions cause dynamic changes of available cache resources.
- Applications’ needs also change dynamically.
Contentions cause dynamic changes of available cache resources.
Applications’ needs also change dynamically.

Locality analysis should be efficient to measure, ideally being applicable during runtime.
Reuse Distance: A Flexible Locality Model

◆ **Definition** [Mattson et. al. 1970, Ding+ 2003]
  - the number of distinct elements between this and the previous access to the same data

◆ **Comparison**
Reuse Distance: A Flexible Locality Model

Definition [Mattson et. al. 1970, Ding+ 2003]

- the number of distinct elements between this and the previous access to the same data

\[ Rd = 2 \]

Comparison
**Reuse Distance: A Flexible Locality Model**

- **Definition** [Mattson et. al. 1970, Ding+ 2003]
  - the number of distinct elements between this and the previous access to the same data

\[
\text{Rd} = 2
\]

- **Comparison**
  - **Reuse Distance**
    - hardware independent
    - defined on each point, no need for windows
    - a distance at each access
  - **Cache Miss Rate**
    - hardware dependent
    - interval-based, defined on windows
    - hit or miss at each access
Reuse Distance Histogram

Percent of references

Reuse distance (cache blocks)

Xipeng Shen, College of William and Mary
Reuse Distance Histogram

Percent of references

Reuse distance (cache blocks)
Reuse Distance Histogram

Xipeng Shen, College of William and Mary
Reuse Distance Histogram

The diagram shows the reuse distance histogram for cache misses. The x-axis represents the reuse distance (cache blocks), and the y-axis represents the percent of references. The histogram is overlaid with a blue curve indicating cache misses. The graph highlights the distribution of references and their corresponding reuse distances.
Many Uses of Reuse Distance in Research

- Study cache reuses \([Ding+:SC04,Huang+:ASPLOS05]\)
- Guide and evaluate program transformation \([Almasi+:MSP02,Zhong+:TOC07]\)
- Discover locality-improving refactoring \([Beyls+:HPCC06]\)
- Model cache sharing \([Chandra+:HPCA05]\)
- Insert cache hints \([Beyls+:JSA05]\)
- Manage superpages \([Cascaval+:PACT05]\)
- Guide memory disambiguation \([Fang+:PACT05]\)
- Predict program performance \([Marin+:SIGMETRICS04]\)
- Model reference affinity \([Zhong+:PLDI04]\)
- Detect locality phases \([Shen+:ASPLOS04]\)

Xipeng Shen, College of William and Mary
Few if Any Uses in Practice
The Big Hurdle

- High cost of measurement
  - \( T \): execution length; \( N \): data size

1970 (Mattson+)
1975 (Bennett+)
1981 (Olken)
1991 (Kim+)
1993 (Sugumar+)
2002 (Almasi+)
2003 (Ding+)

\[ O(T^*N) \]

\[ O(T^*\log\log N) \]
The Big Hurdle

- High cost of measurement
  - T: execution length; N: data size

1970 (Mattson+)
1975 (Bennett+)
1981 (Olken)
1991 (Kim+)
1993 (Sugumar+)
2002 (Almasi+)
2003 (Ding+)

O (T*N)
O (T*\(\log\log N\))

But still, measuring a 1-minute run takes several hours!
Our Solution

◆ Previous methods
  ▶ Essentially implement the definition of reuse distance:
    “Counting” the # of distinct data
◆ Rationale of our method:
  ▶ Use some “cheap” program behavior to statistically approximate reuse distance
The “Cheap” Behavior: Time Distance

- Time distance definition: the number of elements between this and the previous access to the same data.
The “Cheap” Behavior: Time Distance

- Time distance definition: the number of elements between this and the previous access to the same data.

\[ \text{Td} = 5 \]

bcaacdb
The “Cheap” Behavior: Time Distance

- Time distance definition: the number of elements between this and the previous access to the same data.
  \[ Td = 5 \]
  bcaacb

- Reuse distance definition: the number of distinct elements between this and the previous access to the same data.
  \[ Rd = 2 \]
  bcaacb
From Time Distance to Reuse Distance

◆ Is it possible?
From Time Distance to Reuse Distance

Is it possible?

$b \ldots b$

$\text{Td}=5 \quad \text{Rd}= 0, 1, 2, 3, \text{ or } 4?$
From Time Distance to Reuse Distance

- Is it possible?

Td=5 \rightarrow Rd= 0, 1, 2, 3, or 4?

No idea.
**From Time Distance to Reuse Distance**

◆ Is it possible?

Td=5  \[\rightarrow\] Rd= 0, 1, 2, 3, or 4?

◆ What if given the time distance histogram?
  
  ◆ E.g. totally 4 reuses; one time distance is 5, three are all 1.

*No idea.*
From Time Distance to Reuse Distance

◆ Is it possible?

◆ What if given the time distance histogram?
  ❖ E.g. totally 4 reuses; one time distance is 5, three are all 1.

Td=5 \rightarrow Rd = 0, 1, 2, 3, or 4?

No idea.
From Time Distance to Reuse Distance

◆ Is it possible?

```
  b....b
```

Td=5  ➔   Rd= 0, 1, 2, 3, or 4?

No idea.

◆ What if given the time distance histogram?
  ❖ E.g. totally 4 reuses; one time distance is 5, three are all 1.

```
  bxxxxxb
```

Td=5  ➔   Rd= 1.
From Time Distance to Reuse Distance

- Is it possible?

  \[ Td=5 \rightarrow Rd = 0, 1, 2, 3, \text{ or } 4? \]

- What if given the time distance histogram?
  - E.g. totally 4 reuses; one time distance is 5, three are all 1.

  \[ Td=5 \rightarrow Rd = 1. \]
General Form of the Problem

Time distance histogram  Reuse distance histogram
Outline of the Statistical Model

interval length $\Delta$

$P_3(\Delta)$: the expected probability for a variable to appear in the interval.

- Given $P_3(\Delta)$, the problem becomes a Bernoulli process (like tossing $N$ coins.)
Outline of the Statistical Model

interval length $\Delta$

$P_3(\Delta)$: the expected probability for a variable to appear in the interval.

- Given $P_3(\Delta)$, the problem becomes a Bernoulli process (like tossing $N$ coins.)
  - The probability of having $k$ distinct variables in a $\Delta$-long interval:

$$P(k, \Delta) = \binom{N}{k} (P_3(\Delta))^k (1 - P_3(\Delta))^{N-k}$$
Outline of the Statistical Model

interval length $\Delta$

$P_3(\Delta)$: the expected probability for a variable to appear in the interval.

- Given $P_3(\Delta)$, the problem becomes a Bernoulli process (like tossing $N$ coins.)

- The probability of having $k$ distinct variables in a $\Delta$-long interval:

$$P(k, \Delta) = \binom{N}{k} (P_3(\Delta))^k (1 - P_3(\Delta))^{N-k}$$

Enough to calculate the reuse distance histogram.
Outline of the Statistical Model

- $P_3(\Delta)$: the expected probability for a variable to appear in $\Delta$. 

Diagram:
- Interval length $\Delta$
- Time $t$
Outline of the Statistical Model

- $P_3(\Delta)$: the expected probability for a variable to appear in $\Delta$. 

\[ P_3(\Delta) = \text{the expected probability for a variable to appear in } \Delta. \] 

(interval length $\Delta$)

\[ t \]

\[ t - \tau \]
Outline of the Statistical Model

- $P_3(\Delta)$: the expected probability for a variable to appear in $\Delta$.

- $P_2(\tau)$: the probability for a variable’s last access before $t$ to be at time $(t - \tau)$.

\[ P_3(\Delta) = \sum_{\tau=1}^{\Delta} P_2(\tau) \]
Outline of the Statistical Model

◆ $P_3(\Delta)$: the expected probability for a variable to appear in $\Delta$.

◆ $P_2(\tau)$: the probability for a variable’s last access before $t$ to be at time $(t-\tau)$.

\[
\therefore P_3(\Delta) = \sum_{\tau=1}^{\Delta} P_2(\tau)
\]

◆ The following is proved:

\[
P_2(\tau) = \sum_{\varepsilon=\tau+1}^{T} \frac{1}{N} P_T(\varepsilon),
\]

where $P_T(\varepsilon)$ is the value of the time distance histogram when $Td = \varepsilon$.
Methodology

◆ Implementation
  ▶ PIN 3.4 for instrumentation
  ▶ GCC -O3
  ▶ Intel Xeon 2GHz, Fedora Core 3 Linux

◆ Comparison
  ▶ Ding and Zhong’s technique [PLDI’03]
    ▶ Asymptotically the fastest tool for measuring reuse distance
Evaluation (Pulse-like reuse distributions)

Time distance histogram

Reuse distance histogram
(linear scale)
Evaluation (Pulse-like reuse distributions)

Reuse distance histogram (log scale)

Xipeng Shen, College of William and Mary
## Evaluation (SPEC CPU2000)

<table>
<thead>
<tr>
<th>Benchmarks</th>
<th>Speedup</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CINT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>crafty</td>
<td>15.6X</td>
<td>100</td>
</tr>
<tr>
<td>gcc</td>
<td>10.1X</td>
<td>99.9</td>
</tr>
<tr>
<td>gzip</td>
<td>12.3X</td>
<td>99.6</td>
</tr>
<tr>
<td>mcf</td>
<td>19.2X</td>
<td>98.4</td>
</tr>
<tr>
<td>twolf</td>
<td>17.4X</td>
<td>97.5</td>
</tr>
<tr>
<td>vortex</td>
<td>15.6X</td>
<td>100</td>
</tr>
<tr>
<td>CFP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ammp</td>
<td>19.5X</td>
<td>99.5</td>
</tr>
<tr>
<td>applu</td>
<td>19.9X</td>
<td>99.4</td>
</tr>
<tr>
<td>equake</td>
<td>21.1X</td>
<td>98.5</td>
</tr>
<tr>
<td>mesa</td>
<td>18.7X</td>
<td>100</td>
</tr>
<tr>
<td>mgrid</td>
<td>21.5X</td>
<td>99.8</td>
</tr>
<tr>
<td>wupwise</td>
<td>15.6X</td>
<td>99.6</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>17.2X</strong></td>
<td><strong>99.4</strong></td>
</tr>
</tbody>
</table>
Summary

- Demonstrate the connection between time and locality
- Propose a probabilistic model for fast accurate reuse distance approximation
- A step toward practical uses with potential for runtime uses
Thanks!