Multicore OS Benchmarks: We Can Do Better

Ihor Kuz*, Zachary Anderson, Pravin Shinde, Timothy Roscoe
Systems Group, ETH Zurich
* NICTA Australia
Multicore OS benchmarks do not evaluate performance isolation between independent apps.
Mixing Workloads

• Mix must:
  • use system resources
  • not overcommit resources
  • be sensitive to availability of the resources
Mixer Overview

Application candidates $\rightarrow$ Sensitivity analysis $\rightarrow$ Workload Mixer choose optimal mix based on sensitivity

Evaluate Results $\leftarrow$ Run Mixed Workload
# Application Candidates

- Variants
- Application parameters
- Resource constraints

## Example:
- game1: low gfx
- game2: high gfx

<table>
<thead>
<tr>
<th>app</th>
<th>CPU</th>
<th>cache</th>
<th>mem</th>
<th>disk</th>
<th>net</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>game1</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.1</td>
<td>0.1</td>
<td>0.25</td>
</tr>
<tr>
<td>webb1</td>
<td>0.25</td>
<td>0.25</td>
<td>0.1</td>
<td>0.0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>antivN</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.8</td>
<td>0.0</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Sensitivity Analysis

How much resource availability affects goodness score

<table>
<thead>
<tr>
<th>bmark</th>
<th>CPU</th>
<th>cache</th>
<th>mem</th>
<th>disk</th>
<th>net</th>
</tr>
</thead>
<tbody>
<tr>
<td>game</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>webb</td>
<td>0.8</td>
<td>0.7</td>
<td>0.5</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>antiv</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.8</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Mixing

• Optimization problem based on
  • resource usage
  • resource sensitivity

• 2 parts:
  • Choose application variants that use resources they are most sensitive to
  • Constraint: no resource overcommitted

• ILP problem
Run Mixed Workload

• Mix:
  mix = game1, game1, webb1, antivn

• Running the mix
  • unmixed, mixed

• Benchmark result: unmixed - mixed
Evaluate Results

• Low performance difference: Good!
• Comparison between Operating Systems
  • run different optimal mix for each OS
  • compare results
    ➡ how well each OS manages optimal mix
• OS or hardware platform?
Conclusion

- Current status: microbenchmarks
- Real applications
- Bursty applications
- Dynamic workloads
- Extend not Replace