Improving System Performance with Compressed Memory

Sumit Roy and Raj Kumar  
Hewlett-Packard Laboratories  
Palo Alto, CA 94304

Milos Prvulovic  
University of Illinois  
Urbana-Champaign, IL 61801
Overview

- Motivation
- System Architecture
- Experimental Results
- Conclusion
Full Spectrum of Machines

- **XL** Linux Graphics Workstation.
  (4 GB max.)

- **C3600** High-End Personal Workstation.
  (8 GB max.)

- **J6000** High-End Rack Mountable Server.
  (16 GB max.)

- **V-class** Data Center Computing.
  (32 GB max.)
What is the Cost of Memory?
Virtual Memory System
Memory Compression

- Hide disk access latencies on page-faults (approx 10 ms).

- Assume Single Dominant Application.
Approaches

- Hardware Compression (IBM’s MXT) - New hardware
- Kernel Modifications - What about Windows?
- Loadable Driver - Applicable to all OSes.
System Architecture

- CPU
- Cache
- Compressed Memory
- Uncompressed Memory
- Device Driver
- Swap Device
Experiments

- Algorithm Selection -
  used 3-D Scene Reconstruction.

- System Evaluation -
  SPEC 2000 CPU benchmarks.

- Scaling Study -
  Successive Over Relaxation.
Setup

- HP Kayak XU800
- Dual Pentium III, 733 MHz, (256 KByte full speed cache).
- 256 MBytes RDRAM
- Linux 2.2.16 (SuSE 7.0).
Algorithm Selection

- Capture snap-shots of Swap file.
- Compress on a per Page basis.
- Generate Histogram.
Decompression Speed

Count

100000
10000
1000
100
10
1

Decompression Rate (10^6 Bytes/sec)

zlib
WKdm
WK4x4
Lzo
Lzrw
System Evaluation

- Select Large Programs.
- Setup Compressed Memory.
- Get Speedup.
## Resident Size of SPEC 2000 Benchmark Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Integer MBytes</th>
<th>Floating Point MBytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>gzip</td>
<td>180</td>
<td>wupwise 176</td>
</tr>
<tr>
<td>vpr</td>
<td>2 – 35</td>
<td>swim 190</td>
</tr>
<tr>
<td>gcc</td>
<td>7 – 31</td>
<td>grid 55</td>
</tr>
<tr>
<td>mcf</td>
<td>11 – 79</td>
<td>applu 180</td>
</tr>
<tr>
<td>parser</td>
<td>26</td>
<td>mesa 8</td>
</tr>
<tr>
<td>perlbmk</td>
<td>1 – 107</td>
<td>art 3</td>
</tr>
<tr>
<td>gap</td>
<td>192</td>
<td>equake 41</td>
</tr>
<tr>
<td>vortex</td>
<td>48 – 122</td>
<td>amp 13</td>
</tr>
<tr>
<td>bzip2</td>
<td>179 – 184</td>
<td>apsi 191</td>
</tr>
</tbody>
</table>

**Table 1:** 256 MByte System with top(1)
Performance under Memory Pressure

Slowdown (base=256) vs Memory Size (MBytes)

- wupwise
- applu
- apsi
- gzip
- gap
- vortex
- bzip2
### Memory Configuration for SPEC 2000 Evaluation

<table>
<thead>
<tr>
<th>Program</th>
<th>Total (MBytes)</th>
<th>Compressed (MBytes)</th>
<th>Effective (MBytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wupwise</td>
<td>96</td>
<td>16</td>
<td>112 (80 + 32)</td>
</tr>
<tr>
<td>apsi</td>
<td>160</td>
<td>48</td>
<td>208 (112 + 96)</td>
</tr>
<tr>
<td>gzip</td>
<td>64</td>
<td>8</td>
<td>72 (56 + 16)</td>
</tr>
<tr>
<td>gap</td>
<td>160</td>
<td>32</td>
<td>192 (128 + 64)</td>
</tr>
<tr>
<td>bzip2</td>
<td>64</td>
<td>16</td>
<td>80 (48 + 32)</td>
</tr>
</tbody>
</table>
Speed-Up with Memory Compression

![Graph showing speed-up for SPEC 2000 Benchmarks]
Scalability Study

- Setup Compressed Memory.
- Increase Size of Application.
- Speedup Compared to Base system.
**Successive Over Relaxation**

- 6.25 % compressed: Red solid line
- 12.5 % compressed: Green dashed line
- 25 % compressed: Blue dotted line
- 50 % compressed: Magenta dash-dotted line
- 75 % compressed: Cyan dashed line

**Y-axis (Speedup):**
- 0
- 0.5
- 1
- 1.5
- 2
- 2.5
- 3
- 3.5
- 4

**X-axis (Application Size (MBytes)):**
- 0
- 20
- 40
- 60
- 80
- 100
- 120
- 140

Source: HP Invent
Conclusion

- Software Based Memory Compression Works.
- Performance Improvement for Free.
- Needs Tuning.
- Patent Pending.