Correcting vibration-induced performance degradation in enterprise servers
Christine Chan, Boxiang Pan, Tajana Šimunić Rosing, UCSD
Kenny Gross, Kalyan Vaidyanathan, Oracle

Introduction

Cooling energy challenges
- Combat high peak temperatures during high utilization
- Avoiding over-provisioning
- Monitoring side effects of cooling on application performance

Contributions
- System measurement methodology to characterize disk sensitivities
- Thermal and power management scheme responding to vibration-induced performance issues

Measurement methodology

Instrumentation
- Tri-axial accelerometers
- Programmable shake table
- IPMI (system out-of-band management)

- Amplitude sensitivity curves vary across drives and frequency profiles
- Drives show narrow bands of frequency dependence

Workloads and management policies

- Thermal sensors
- Power sensors
- Application performance
- Disk throughput degradation

- Controller
- Memory page scheduling
- Fan subsystem
- Core scheduling
- Page scheduler
- DIMM

System control framework

- CPU power distribution
- DIMM power distribution
- Target temperature
- Performance constraints

Actuators
- Core scheduling, page migration, fan control

Sensors
- Temperature, power, fan speed, performance feedback

TPC-H: Decision support benchmark
- Decision support benchmark modeling large scale enterprise applications
- Queries can be broken down into phases with varying sensitivity to I/O throughput, e.g. table scans, hash join

Performance improvement

- Average 14%, maximum 30% speedup
- Application level oracle: Possible additional improvement (+5%) but infeasible:
  - Requires app-level integration and workload prediction of when disk is needed
  - Intrusive ongoing I/O throughput measurements
  - Sacrifices batch job performance

Energy savings
- Average 43%, maximum 62% savings
- Reduction in energy consumption due to lower core temperatures, lower fan speeds, shorter uptime

Simulation setup
- DB queries: TPC-H suite
- Compute jobs: SPEC suite
- Core utilization 75%
- Power density, heat dissipation, cooling effect modeled as RC circuit

<table>
<thead>
<tr>
<th>Workload ID</th>
<th>Benchmark composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>5 bzip2, 1 tpchq1</td>
</tr>
<tr>
<td>W2</td>
<td>5 bzip2, 1 tpchq3</td>
</tr>
<tr>
<td>W3</td>
<td>5 bzip2, 1 tpchq10</td>
</tr>
<tr>
<td>W4</td>
<td>5 bzip2, 1 tpchq19</td>
</tr>
<tr>
<td>W5</td>
<td>3 bzip2, 2 hamm, 1 tpchq1</td>
</tr>
<tr>
<td>W6</td>
<td>3 bzip2, 2 hamm, 1 tpchq3</td>
</tr>
<tr>
<td>W7</td>
<td>3 bzip2, 2 hamm, 1 tpchq10</td>
</tr>
<tr>
<td>W8</td>
<td>3 bzip2, 2 hamm, 1 tpchq19</td>
</tr>
</tbody>
</table>

Full System Thermal Management (FSTM)
- Proactive task migration, page migration, and fan control based on a band-limited temperature predictor
- Partition discretized fan settings into temperature or disk performance priority zones based on disk sensitivity slope

Policy results

Publications
- Rajd Zuhair Ayoub, Rajib Nath, Tajana Rosing, “JETC: Joint energy thermal and cooling management for memory and CPU subsystems in servers”. HPSC 2012

Funded by: NSF Project GreenLight Grant 0821155, NSF Grants 0916127 and 1029783, MuSyC, DARPA, Oracle, and UCSD Center for Networked Systems