Energy consumption of scheduling policies for HTC jobs in the Cloud

Osama Alrajeh
o.alrajeh1@newcastle.ac.uk

Nigel Thomas
nigel.thomas@newcastle.ac.uk

School of Computing Science
Newcastle University, UK

Presentation for ENERGY-SIM 2015
Outline

- Introduction
- Motivation
- Background
- Policies
- Simulation Environment
- Results
- Conclusions
- Future work
Introduction

- The impact of HTC scheduling policies within a cloud environment have received limited attention in the literature.

- Build on previous work to evaluate the energy and performance impact of scheduling policies for HTC workloads to cloud resources.

- By extending trace-driven simulation to model energy consumption.

- Compare the energy consumption of the same workload on servers whose performance and energy consumption characteristics differ.
Motivation

- Global energy demand has increased from 10 thousand terawatt hours in 1990 to 20 thousand terawatt hours today.

- By 2040, global demand is expected to approach 40 thousand terawatt hours.

- 50% of datacenter total energy consumption goes to servers.

- Existing research has applied policies for scheduling HTC jobs to Cloud resources without consideration for its energy consumption.

- Understand the impact of scheduling policies on energy consumption.

- The need to reduce money and CO$_2$ emissions.
Background

• HTCondor:
  • An open source high throughput computing software.
  • Provides workload management system technology for grid computing jobs.
  • Has a job queuing mechanism, scheduling policy, priority scheme, resource monitoring, and resource management.

• SPECpower_ssj® 2008:
  • Benchmark that evaluates the power and performance characteristics of single server and multi-node servers.
Policies

- McGough et al.\textsuperscript{1} have previously proposed policies governing the scheduling of HTC jobs to Cloud instances.
- Aiming to reduce cost and overheads.
- We seek to quantify the energy impact of these policies on the Cloud provider by using trace-driven simulation.

Policies

- The state diagram for Instances:

Policies

• P1: limiting the maximum number of Cloud instances:
  • Reduce cloud cost.
  • Reduce idle time.
  • Increase overhead.

• P2: merging of different users’ jobs:
  • Reduce cloud cost.
  • Reduce idle time.
  • Reduce overhead.
  • Increase security concerns.
Policies

- **P3: instance keep-alive:**
  - Reduce overhead.
  - Increase idle time.
  - Increase cloud cost.

- **P4: delaying the start of instances:**
  - Reduce idle time.
  - Reduce cloud cost.
  - Increase overhead.
Simulation Environment

• Simulation Scenario:
  • Historical logs for 409,479 completed jobs from the HTCondor cluster located at Newcastle University.
  • Selected server from SPECpower ssj 2008 published results:

<table>
<thead>
<tr>
<th>No</th>
<th>Server Name</th>
<th>Cores</th>
<th>Chips</th>
<th>Peak Power (w)</th>
<th>Idle Power (w)</th>
<th>ssj_ops</th>
<th>Scaling ratio</th>
<th>Test Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PRIMERGY RX2560 M1</td>
<td>36</td>
<td>2</td>
<td>267</td>
<td>40.1</td>
<td>3,256,040</td>
<td>3.66</td>
<td>Mar-15</td>
</tr>
<tr>
<td>2</td>
<td>Express5800/A1080a-E</td>
<td>64</td>
<td>8</td>
<td>1749</td>
<td>993</td>
<td>3,647,000</td>
<td>4.10</td>
<td>Dec-10</td>
</tr>
<tr>
<td>3</td>
<td>ProLiant DL385 G7</td>
<td>48</td>
<td>4</td>
<td>271</td>
<td>101</td>
<td>888,819</td>
<td>1.00</td>
<td>Mar-10</td>
</tr>
<tr>
<td>4</td>
<td>PRIMERGY TX150 S7</td>
<td>4</td>
<td>1</td>
<td>112</td>
<td>24.3</td>
<td>276,514</td>
<td>0.31</td>
<td>Jan-10</td>
</tr>
<tr>
<td>5</td>
<td>Proliant DL580 G5</td>
<td>16</td>
<td>4</td>
<td>387</td>
<td>271</td>
<td>359,523</td>
<td>0.40</td>
<td>Dec-07</td>
</tr>
</tbody>
</table>
Simulation Environment

• Resource model:
  • Energy consumption:
    • SPECpower benchmark.
    • Load level for booting and working jobs is 100%.
  • Performance scaling:
    • Scaling the duration of jobs: \( D_j^{s,s}(s,j) = D_j \times \frac{P_s}{P_b} \)

\( D_j \): the job duration which is the start time job minus end time job.

\( P_s \): the power of one of the other selected servers.

\( P_b \): the baseline server performance.
Simulation Environment

• Metrics:
  • Overhead:
    • The difference between the execution time of the job $D'^s_j$ and the amount of time the job took to run in the simulation.
  • Cloud hours:
    • The number of ‘instance hours’ (provider-side).
    • Cost can easily be calculated from cloud hours.
Results: P1: Limiting the number of Cloud instances.

<table>
<thead>
<tr>
<th>No</th>
<th>Peak Power (w)</th>
<th>Idle Power (w)</th>
<th>ssj_ops</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>267</td>
<td>40.1</td>
<td>3,256,040</td>
</tr>
<tr>
<td>2</td>
<td>1749</td>
<td>993</td>
<td>3,647,000</td>
</tr>
<tr>
<td>3</td>
<td>271</td>
<td>101</td>
<td>888,819</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
<td>24.3</td>
<td>276,514</td>
</tr>
<tr>
<td>5</td>
<td>387</td>
<td>271</td>
<td>359,523</td>
</tr>
</tbody>
</table>
Results: P2: Merging of different users’ jobs.

<table>
<thead>
<tr>
<th>No</th>
<th>Peak Power (w)</th>
<th>Idle Power (w)</th>
<th>ssj_ops</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>267</td>
<td>40.1</td>
<td>3,256,040</td>
</tr>
<tr>
<td>2</td>
<td>1749</td>
<td>993</td>
<td>3,647,000</td>
</tr>
<tr>
<td>3</td>
<td>271</td>
<td>101</td>
<td>888,819</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
<td>24.3</td>
<td>276,514</td>
</tr>
<tr>
<td>5</td>
<td>387</td>
<td>271</td>
<td>359,523</td>
</tr>
</tbody>
</table>
Results: P3: Instance keep-alive.

### Table

<table>
<thead>
<tr>
<th>No</th>
<th>Peak Power (w)</th>
<th>Idle Power (w)</th>
<th>ssj_ops</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>267</td>
<td>40.1</td>
<td>3,256,040</td>
</tr>
<tr>
<td>2</td>
<td>1749</td>
<td>993</td>
<td>3,647,000</td>
</tr>
<tr>
<td>3</td>
<td>271</td>
<td>101</td>
<td>888,819</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
<td>24.3</td>
<td>276,514</td>
</tr>
<tr>
<td>5</td>
<td>387</td>
<td>271</td>
<td>359,523</td>
</tr>
</tbody>
</table>

### Graphs

- Average overhead (seconds) vs. Percentage chance of keep-alive
- Number of instance hours vs. Percentage chance of keep-alive
- Energy consumption (MWh) vs. Percentage chance of keep-alive
Results: P4: Delaying the start of instances.
Conclusions

- The policies exhibit varying impacts on energy consumption and average overheads.

- In all cases the policies have demonstrated the criticality of the trade-off between energy consumption and performance/cost.

- The tool will ultimately support the development and evaluation new algorithms for HTC workload scheduling in cloud environments.
Future work

- Extend the simulation framework to support the modelling of virtualised resources.

- Incorporate and extend our modelling of cloud resources within HTC-Sim$^2$ environment.