Enhancing Cloud Energy Models for Optimizing Datacenters Efficiency

Edouard Outin, Jean-Emile Dartois, Oliver Barais, and Jean-Louis Pazat

Summarized by Warren Connell

19 Oct 15
Overview

• Dealing with data-center optimization, energy consumption must be taken into account
• Energy consumption models may not be accurate
• Uses a genetic algorithm to optimize energy consumption as well as other factors
2 Research Questions

• Do differences exist between the energy simulation based on hardware specifications and the real data that can be observed?
• Could we use machine learning techniques at runtime to improve the simulation accuracy?
Existing Simulation Platforms

- CloudSim
  - CPU utilization -> watts

- GreenCloud
  - Models CPU, RAM, disk, other components
  - Parameters set manually

- SimGrid
  - SURF energy Plug-in
  - CPU only, linear model

- iCanCloud
  - Models components and applications
  - Applications can be problematic to predict for cloud provider
Comparing Simulator Values to Measurements

- CloudSim uses data from [http://spec.org](http://spec.org) for power consumption
Comparing Simulator Values to Measurements

- CloudSim uses data from [http://spec.org](http://spec.org) for power consumption
- Ran physical server and measured consumption
  - Power consumption available from server (ACPI port)
  - Used external Raritan Power Distribution Unit
  - ~10W difference!
Experimental results

Figure 1. DELL PowerEdge R620 benchmark - No hypervisor

Figure 2. DELL PowerEdge R620 benchmark - Hypervisor

Cannot rely on CPU metrics alone!
Must factor in RAM, disk, and other components (network, cooling)
Energy Model

Model

Analyze

Plan

Monitor

Execute

Sensors

Effectors

Managed element: Cloud infrastructure
Energy Model
Energy Model

Diagram showing the process of energy model:
- Analyze
- Plan
- Execute

Detailed elements:
- Compute servers (hosting VMs)
- PDU
- Metrics Collector
- Retrieve energy consumed (Watts) by a given host
- Get host metrics

Element: Cloud infrastructure

Effectors
Energy Model

Heuristics/genetic algorithm to select best model

Managed element: Cloud infrastructure
Energy Model

Incorporate Knowledge via machine learning at runtime
Machine Learning Method

- Multivariate Adaptive Regression Spline (MARS)
  - Non-linear modeling method
  - Does not assume any particular relationship between inputs and output
  - Basic linear and polynomial regression failed to produce accurate results
## Example Input / Output

<table>
<thead>
<tr>
<th>CPU(%)</th>
<th>RAM(%)</th>
<th>Reads</th>
<th>Writes</th>
<th>Recv (kb)</th>
<th>Sent (kb)</th>
<th>Watts</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.5</td>
<td>2.3</td>
<td>2932</td>
<td>64</td>
<td>1</td>
<td>1</td>
<td>131</td>
</tr>
<tr>
<td>91.4</td>
<td>2.3</td>
<td>3188</td>
<td>14628</td>
<td>3473</td>
<td>44</td>
<td>189</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6.2</td>
<td>1.5</td>
<td>259</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>101</td>
</tr>
</tbody>
</table>
Example Input / Output

<table>
<thead>
<tr>
<th>CPU(%)</th>
<th>RAM(%)</th>
<th>Reads</th>
<th>Writes</th>
<th>Recv (kb)</th>
<th>Sent (kb)</th>
<th>Watts</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.5</td>
<td>2.3</td>
<td>2932</td>
<td>64</td>
<td>1</td>
<td>1</td>
<td>131</td>
</tr>
<tr>
<td>91.4</td>
<td>2.3</td>
<td>3188</td>
<td>14628</td>
<td>3473</td>
<td>44</td>
<td>189</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6.2</td>
<td>1.5</td>
<td>259</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>101</td>
</tr>
</tbody>
</table>

\[ E_{\text{host}} = 170.965 + 0.794 \times \max(0, \text{cpu} - 11.900) \]
\[ - 2.625 \times \max(0, 11.900 - \text{cpu}) \]
\[ - 9.997 \times \max(0, 6.800 - \text{ram}) \]
\[ + 0.009 \times \max(0, \text{cpu} - 11.900) \times \max(0, 42.000 - \text{sent}) \]
<table>
<thead>
<tr>
<th>CPU(%)</th>
<th>RAM(%)</th>
<th>Reads</th>
<th>Writes</th>
<th>Recv</th>
<th>Sent</th>
<th>Watts</th>
<th>Pred.</th>
<th>spec.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.5</td>
<td>2.3</td>
<td>2932</td>
<td>64</td>
<td>1</td>
<td>1</td>
<td>131</td>
<td>130.16</td>
<td>84.4</td>
</tr>
<tr>
<td>91.4</td>
<td>2.3</td>
<td>3188</td>
<td>14628</td>
<td>3473</td>
<td>44</td>
<td>189</td>
<td>189.10</td>
<td>225</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6.2</td>
<td>1.5</td>
<td>259</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>101</td>
<td>103.02</td>
<td>67.7</td>
</tr>
</tbody>
</table>
Observations

- Dominant features: CPU / RAM / bytes sent
  - Disk and bytes received not present in final model
- Training set and validation set
  - Average error: 3.8%
- Fast processing time: 0.007 sec for 1000
- Spec.org underestimates at low CPU % and overestimates at high CPU %
Future Work / Unanswered questions

- Disk operations ignored
  - Possibly due to disk operations remaining constant
- Overhead from VM live migrations
- Network power consumption
- Take new machine learning model and use in actual autonomic solution
  - Genetic algorithms mentioned in abstract