EnaCloud: An Energy-saving Application Live Placement Approach for Cloud Computing Environments

Shayan Mehrazarin, Yasir Alyoubi, and Abdulmajeed Alyoubi

March 25, 2015
Outline

- The Objective of *EnaCloud*
- The Contributions of *EnaCloud*
- The Methodology of *EnaCloud*
- Example of How *EnaCloud* Works
- Energy-Aware Heuristic Algorithm
- Interpretation of the Results
The Objective of *EnaCloud*

- Various solutions have been proposed in the past to address problems of high energy consumption, but there are some issues that have not been properly addressed:
  - Some methods of energy savings, such as turning off the monitor or enabling sleep mode, benefit only a single computer but not the whole cloud platform
  - Releasing some server nodes in large data centers and turning them off ("workloads concentration") requires static configurations & settings
    - In open clouds, applications dynamically arrive & depart
  - These solutions require applications to be able to shut down and then copy them to idle servers
    - However, underutilization of the server is likely as it does not support live application migration
The Contributions of **EnaCloud**

- Major contributions of EnaCloud towards previously proposed solutions:
  - Introduction of an energy-conscious algorithm to gather application schemes with regards to various events that occur (arrival, departure)
  - Designing and implementing an architecture for EnaCloud that is based on a virtual computing environment that works with HaaS (*Hardware-as-a-Service*) and SaaS (*Software-as-a-Service*) cloud services
    - This approach can reduce energy-consumption based on experiments and studies
The Methodology of EnaCloud

- For the purpose of EnaCloud, the authors assume:
  - All computing nodes are similar
  - Each server has a resource capacity of 1 unit
  - All nodes are connected to each other via LAN (high speed)
  - Each computing node contains ≥ 1 virtual machine (VM)

- Additionally, the authors classify nodes as:
  - open box for active server nodes using VM
  - closed box for inactive server nodes not using VM
The Methodology of *EnaCloud* (cont.)

- EnaCloud ensures workloads are calculated in a way that minimizes the amount of *open boxes*
  - Workloads will always depart or arrive dynamically in a typical cloud service

- Over-precision ratio defined as $\alpha \geq 0$ and $\alpha \leq 1$ for energy-aware heuristic algorithm
  - used to check if $\text{size}'(x)$ falls between $(1 - \alpha) \times \text{size}(x)$ and $(1 + \alpha) \times \text{size}(x)$
Example of How *EnaCloud* Works

- Suppose in this example that there is the arrival of a 0.5 unit workload
  - A new box should be opened
  - However, it is possible to avoid opening a close box if this workload can be placed into the first node (out of 2 nodes)
    - It would be required to migrate the first node to the second node
a) **Without migration** - Inserting a new workload requires three open boxes

b) **With migration** - Inserting a new workload while maintaining two open boxes
Example of How *EnaCloud* Works (cont.)

- *Workload resizing* is the event where applications will have resource demands that vary

- Workload resizing includes:
  - *workload inflation*, which impacts the other workloads’ performance within the same node
  - *workload deflation*, which frees some resources and can result in wasting energy along with idling of resources
A common problem is using migration to re-map workloads alongside resource nodes with the arrival, departure, or resizing of workloads.

There are two goals with migration:
- to keep the amount of *open boxes* at a minimum
- to keep migration times at a minimum
Energy-Aware Heuristic Algorithm

- It is based on partitioning workload size from \((0, 1]\) into \(2^M - 2\) subintervals:

\[
\begin{align*}
L_0 &= \left( \frac{M - 1}{M}, 1 \right) \\
L_1 &= \left( \frac{M - 2}{M - 1}, \frac{M - 1}{M} \right) \\
& \quad \ldots \\
L_{M-1} &= \left( \frac{1}{3}, \frac{1}{2} \right) \\
& \quad \ldots \\
L_{2^M-4} &= \left( \frac{1}{M}, \frac{1}{M-1} \right) \\
L_{2^M-3} &= \left( 0, \frac{1}{M} \right)
\end{align*}
\]
Energy-Aware Heuristic Algorithm (cont.)

- Pseudo-code for workload arrival function is shown on the right
- Includes implementation for First-Fit and Best-Fit too

```
Procedure: Insert
Input: x, size of the arrival workload
Output: a placement scheme
1. if level(x)=2M-3 or level(x)=0
2. insert x using First-Fit
3. return the destination node of x
4. foreach node v in pool
5. foreach workload w in node v
6. filter out w where level(w)<level(x)
7. place x to v* using Best-Fit
8. sort each workload w* in v* where level(w*)<level(x)
to \{ w_{1*}, \ldots, w_{n*} \} in ascending order
9. for i = 1 to n
10. if v* can accommodate x
11. break
12. pop w_i* from v* and Insert (w_i*)
```
Energy-Aware Heuristic Algorithm (cont.)

- The *workload departure* function is shown below in pseudo-code:

```
Procedure: Pop
Input: the node x that the workload departs from
Output: migration scheme
1. foreach workload w in node v
2. pop w and invoke Y=Insert(w)
3. Return \( \bigcup Y_i \)
```

- The *workload resize* function is shown below, as well:

```
Procedure: Resize
Input: old size x of workload, new size y of workload
Output: migration scheme
1. \( X=Pop(x) \)
2. \( Y=Insert(y) \)
3. Return \( X \cup Y \)
Interpretation of the Results - Nodes

- With regards to the amount of active nodes, the authors demonstrate using a chart *(shown on the right)* how EnaCloud compares with First Fit and Best Fit data.

- We interpreted from the chart that EnaCloud maintains a decent balance between the increase and decline of active nodes over a period of approximately 500 to 600 minutes.

- Amount of active nodes can range based on experiments from 20 to 40 active nodes.
Interpretation of the Results - Energy

- The authors demonstrate using a chart (shown on the right) how EnaCloud compares with First Fit and Best Fit data with regards to how much energy is consumed.

- Based on our interpretation of the experiment results, it seems that EnaCloud indeed has an energy savings as more time has elapsed.

- However, it would not have too much savings for short periods of time.
Interpretation of the Results - Utilization

- The authors show using a chart (shown on the right) how EnaCloud compares with First Fit and Best Fit data with regards to the percentage of pool utilization.

- Our interpretation of the data suggests that the utilization rate tends to be consistent for the most part with EnaCloud.

- Ranges between 80 to 95 percent

*Pool Utilization (%) vs. time (minutes)*
Any Questions?