Behavioral Model for Cloud aware Load and Power Management

HotTopiCS 2013

Dipl.-Inform. Kiril Schröder
Prof. Dr.-Ing. Wolfgang Nebel
C.v.O. University of Oldenburg, Germany
Outline

• Motivation
• System Overview
• Load and Power Management Extension
• Behavioral Model
• Evaluation
• Conclusion
Motivation

**Cloud Computing**

- Highly flexible
- Penetration is rising

---

**Increasing Penetration of Virtualization and Private Cloud Technologies into x86 Workloads...**

---

**All Public Cloud Options Grow Well**

- A base for cloud computing is (server) virtualization

---

Source: AlphaWise™, Morgan Stanley Research

---

20/04/2012

Kiril Schröder, Wolfgang Nebel
Motivation

Virtualization

• Services encapsulated in virtual machines (VM)
• Consolidation of servers
  — peak-oriented capacity planning
  — low average utilization (20 – 30 %)
• Dynamic consolidation, adapting to the needs
• Energy demand reduction: 40 – 80 %

• Using distributed data centers for
  — further energy, cost reduction
  — greenhouse gas reduction

⇒ Important: effects of migrations

Source: AlphaWise®M, Morgan Stanley Research

Utilization Rates Continuing to Increase

Utilization rate (%)

<table>
<thead>
<tr>
<th></th>
<th>Today</th>
<th>One year</th>
<th>Three years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53</td>
<td>59</td>
<td>63</td>
</tr>
</tbody>
</table>

20/04/2012 Kiril Schröder, Wolfgang Nebel
Motivation

Related Work

• Data center models
  – Abbasi et al. [1], Mukherjee et al. [24], Pakbaznia and Pedram [26], ...
  – covers hardware: servers, cooling (with thermal flow), UPS, ...
  but not the software
   (→ dynamic consolidation)

• Inter-site load management
  – Church et al. [10], Qureshi et al. [28], Zhang et al. [41], ...
  – consider (re)allocation of tasks
    • different optimization problem
    • can be done more fine-granular
System Overview

local

DC

forecasting

LPM

VM pool

agent

model

LPM model

global

arbiter

VM pool

network model

extensions for dynamic environments
LPM Extension

Base Load and Power Management (LPM) from Hoyer et al., 2011

- Dynamic consolidation with QoS
- No additional servers are needed
- Use of a forecast algorithm

- Methodology
  - initial, static distribution (called safe distribution)
    - sufficient resources at any time (assumed)
  - dynamic consolidation leads to dynamic distributions
    - unsafe: not sufficient resources at any time

- Shortcoming
  - not designed for changes (VM set/profiles)
LPM Extension

Problem

• Changing safe distribution considering current dynamic distribution

→ Heuristics created for adding / removing VMs and changing VM profiles

Kiril Schröder, Wolfgang Nebel
Behavioral Model

• Model necessary?
  – Needed information: number of active servers
  – LPM runtime polynomial, too slow for optimizations

• Linear regression model
  – Variables with linear computation complexity
  – Simplification: homogeneous servers, workload ~ only cpu

• Modeling (training) data:
  – 100 scenarios (different selections and number of VMs)
  – 10000 VM traces available
  – 10 simulated days (1 min. resolution)
Behavioral Model

Modeling Steps

Defining the regression model

• Selection of variables
  – Influence of the different variables on the quality

Constraints for use

• Training length
  – $f_{\text{quality}}(\text{training length})$

• Effect of changes
  – $f_{\text{quality}}(\text{training length, changes})$
Behavioral Model

Selection of Variables

- Variables: $x$ to the power of $n$, $n \in \{1,2,3,4\}$
- Forecasted values: $x_t, x_{t+i}, x_{t+2i} \ldots$ $t$: time, $i$: equidistant step

- Two regression models:

$$\#SRV_a(t) = \alpha_0 + \sum_s \left( \alpha_{1,s} \cdot SoL(t + s) + \alpha_{2,s} \cdot \#VM(t + s) + \sum_{i=0}^9 \alpha_{3+i,s} \cdot \#VMC_i(t + s) \right)$$

1: $s \in \{0\}$
2: $s \in \{0,5,10,15,20,25\}$
Behavioral Model

Training Length

- Best results with training length $\geq 24$ hours
Behavioral Model

Effect of Changes

- Only VM pool changes <=50%
Evaluation

Simulation Settings

• Evaluation data: 100 scenarios
  – Dynamic VM pool: initial 150 VMs, at most 300 VMs
  – Randomly adding or removing VMs: every 4 to 8 hours
  – Considering constraints (24h training length, 50 % change limit)
  – Regression model generated at each change
  – Prediction corresponds until next change
Evaluation

Forecast Quality

- Model 2 is only a little better
- Average precision in interval: 95 %
- Average precision point-by-point: 93 %
Evaluation

**Impact of the Heuristics**

- Impact to the safe distributions:
  - 10 % reduction of provided servers

- Impact to the dynamic distributions:
  - nearly none
  → no relation between packing rates in safe and dynamic distribution
Conclusion

• Extension of an existing LPM
  – Now possible: changes in the VM profiles, changing VM selection
  – Heuristics: 10 % reduction of needed servers (safe distribution)

• LPM behavioral model
  – Linear regression model
  – Average precision quality: 93 % (95 %)

• At present: power = f (#servers\textsubscript{active})

• In future:
  – integration into a data center power model
  – targeted generation of loads at each site
  – smart grid integration