Energy Efficiency in Virtualized Systems

Segoy Thiel

MIN Faculty
Department of Informatics
University of Hamburg

10. December 2014
Outline

1. Introduction
Outline

1. Introduction

2. Virtualization in Data Centers
   - Architecture
   - Simulation

Segoy Thiel  MIN Faculty - UHH  Energy Efficiency in Virtualized Systems  10.12.14
Outline

1. Introduction

2. Virtualization in Data Centers
   - Architecture
   - Simulation

3. Consequences for Providers and Users
Outline

1. Introduction

2. Virtualization in Data Centers
   - Architecture
   - Simulation

3. Consequences for Providers and Users

4. Conclusion
Motivation

- Data Centers consume large amounts of energy
  - 56% rise from 2005-2010
  - 2010 about 1.3% of total energy use
  - estimated to 2% of CO₂ emissions

- Virtualization allows to create several Virtual Machines (VMs) on a physical server to reduce amount of hardware in use

- VMs can be moved between physical hosts

- Resource usage can be adjusted to current requirements
Focus

- Live migration of VMs to dynamically reallocate VMs to current resource requirements
- Idle physical nodes can be switched off
- Provide reliable Quality of Service (QoS) defined in Service Level Agreements (SLA) characteristics:
  - maximal throughput
  - minimal response time
  - or latency delivered by deployed system
- Thermal Optimization - temperature of physical nodes is considered in reallocation decisions
  - reducing workload of overheated nodes
  - decrease cooling system load
Challenges

How to:

1. Solve trade-off: energy savings vs. delivered performance
2. Determine when which and where to migrate VMs in order to:
   - minimize energy consumption
   - minimize migration overhead
   - ensuring SLA
3. Develop efficient decentralized and scalable algorithms for resource allocation
4. Develop comprehensive solution by combining several allocation policies with different objectives
Infrastructure for large-scale Cloud data center

- Comprises physical nodes characterized by:
  - CPU: performance defined by MIPS
  - RAM
  - Network bandwidth
- Users submit requests for provisioning VMs requiring resources from nodes
- SLA violation occurs when a VM cannot get requested amount of resources
- Tiered comprising: global and local managers
Local and Global Managers

Local managers:

- Reside on nodes as part of VMM
- Choose VMs to be migrated in cases:
  - utilization of resource is close to 100% - risk of SLA Violation
  - utilization is low - all VMs reallocated to another node
  - VMs have intensive network communication allocated to different hosts
  - temp exceeds limit - VMs are migrated to cool node down
- Send information and VMs to migrate to global manager
- Issue VMresizing, appiclate DVFS, turn on/off idle nodes

Global manager:

- Attached to set of nodes and process data from local managers
The System Model

Figure 4.1
The System Model

http://beloglazov.info/thesis.pdf Figure 4.1

1 New requests for VM provisioning
The System Model

1. New requests for VM provisioning
2. Data about utilization of resources and VMs chosen to migrate
The System Model

1. New requests for VM provisioning
2. Data about utilization of resources and VMs chosen to migrate
3. Migration commands
The System Model

1. New requests for VM provisioning
2. Data about utilization of resources and VMs chosen to migrate
3. Migration commands
4. Commands for VM resizing and adjusting of power states
The System Model

1. New requests for VM provisioning
2. Data about utilization of resources and VMs chosen to migrate
3. Migration commands
4. Commands for VM resizing and adjusting of power states
5. VM resizing, scheduling and migration actions
Three stages of VM placement optimization

1. Reallocation according current utilization of multiple system resources
   - Main idea: set upper and lower utilization thresholds and keep utilization of CPU created by VMs sharing node inbetween

2. Optimization of virtual network topologies established between VMs
   - place communicating VMs in a way to minimize overhead of data transfer over network

3. VM reallocation considering thermal state of resources
   - monitor nodes thermal states, reallocate workload from overheated nodes to allow natural cooling
Allocation Policies

Algorithms have to meet following requirements:

- Decentralization and parallelism
- High performance
- Guaranteed QoS
- Independence of the workload type

VM reallocation problem can be divided in two:

1. Selection of VMs to migrate
   - has to be considered separately for each stage

2. Determining new placement of these on physical hosts
   - solved by heuristic for multidimensional bin-packing
Example Simulation of a Data Center

- Data Center consists of 100 heterogeneous physical nodes
- Each node have one CPU core, performance: 1000, 2000 or 3000 MIPS, 8Gb RAM, 1TB storage
- Users submit requests for provisioning 290 heterogeneous VMs that fill full capacity of Data Center
- Simulated policies:
  - Non Power Aware (NPA)
  - DVFS
  - Single Threshold (ST)
  - Two-threshold policy aimed at Minimization of Migrations (MM)
## Example Simulation Results

<table>
<thead>
<tr>
<th>Policy</th>
<th>Energy</th>
<th>SLA</th>
<th>Migr.</th>
<th>Avg. SLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPA</td>
<td>9.15 KWh</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DVFS</td>
<td>4.40 KWh</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ST 50%</td>
<td>2.03 KWh</td>
<td>5.41%</td>
<td>35 226</td>
<td>81%</td>
</tr>
<tr>
<td>ST 60%</td>
<td>1.50 KWh</td>
<td>9.04%</td>
<td>34 231</td>
<td>89%</td>
</tr>
<tr>
<td>MM 30-70%</td>
<td>1.48 KWh</td>
<td>1.11%</td>
<td>3 359</td>
<td>56%</td>
</tr>
<tr>
<td>MM 40-80%</td>
<td>1.27 KWh</td>
<td>2.75%</td>
<td>3 241</td>
<td>65%</td>
</tr>
<tr>
<td>MM 50-90%</td>
<td>1.14 KWh</td>
<td>6.69%</td>
<td>3 120</td>
<td>76%</td>
</tr>
</tbody>
</table>

http://seelab.ucsd.edu/virtualefficiency/related_papers/26_4039a826.pdf
Cloud Computing & Virtualization

Cloud Computing leverage virtualization

- Provides ability to provision resources on demand
- Organizations can outsource computation needs
- Eliminate necessity to maintain computing infrastructure

Cloud Computing leads to energy efficiency providing following characteristics:

- Economy of scale and elimination of redundancies
- Increased utilization of computing resources
- Location independence
- Scaling up/down and in/out
Emissions


Xen Hypervisor

- P- and C-states
- Supports offline and live migration of VMs

Kernel-based Virtual Machine (KVM)

- Open source as module for Linux kernel
- Linux as hypervisor

VM-Ware

- Similar to Xen, supports host-level power management via DVFS
Data Centers

- Data Center goals:
  - Save energy to reduce costs
  - Provide reliable QoS for heterogenous workloads

- Mostly achieved by DVFS
- Different Virtualization techniques achieve better results by:
  - Switching off idle nodes
  - Reallocate workload to provide better utilization of resources
- Provided Fault and performance isolation between VMs
- Problems handle arbitrary workloads with reliable QoS
Virtualization for Providers

- VM migration causes performance and energy overhead
- Provider is not aware of workloads
  - QoS must be defined workload independent
- Oversubscribe system resources (CPU)
  - most benefits from dynamic VM consolidation
  - risky from QoS perspective: performance degradation
- Elimination of single points of failure
  - if a compute or controller node fails system stays operable
Users

Average Users

- Create low amount of traffic
- Have little loss of performance when put on VM
- Can save much energy for providers when virtualized

High performance Users

- Create high amount of traffic
- Utilize big parts of machines
- Virtualization overhead can decrease performance significantly
- Decrease of performance when sharing CPU
Consequences for Users

- For average users:
  - disadvantages are overseeable
  - unsignificant decrease of perfomance
  - good virtualization techniques no difference will be noticed
  - note: maybe interesting: security of data when migrated to multiple servers

- For high performance users:
  - dependend on how much perfomance is needed
  - migration should be avoided
  - sharing physical host can be critical
  - virtualization overhead can decrease perfomance
Conclusion

- For saving energy on data centers virtualization useful
- Benefits from cloud computing
- Through modern virtualization techniques
  - overhead and performance loss are reduced
  - becomes efficient and applicable
- Problems with high performance users
- Optimization Algorithms long runtime not live applicable
Anton Beloglazov

*Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing*
The University of Melbourne, 2013

Anton Beloglazov

*Energy Efficient Resource Management in Virtualized Cloud Data Centers*
The University of Melbourne, 2010
http://seelab.ucsd.edu/virtualefficiency/related_papers/26_4039a826.pdf

http://www.environmentalleader.com
on December 9th 2014