Server Consolidation in Virtualized Data Centers

Seminar *Green Networking*

Leon Richardt

10 January 2020

Osnabrück University

Table of Contents

Principles

Principles

Power Consumption in Data Centers

Principles: Power Consumption in Data Centers

Figure 1: Server racks at CERN [Hug10]

Figure 2: Ventilation unit [Bee15]

Figure 3: UPS in a data center [Cgx08]

Figure 4: Network switch [Cés08]

Principles: Power Consumption in Data Centers

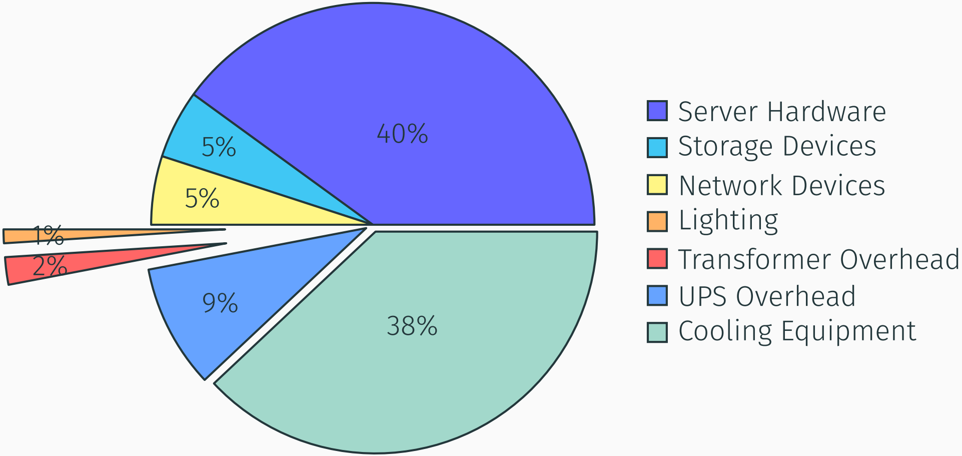

Figure 5: Average distribution of power consumption by different data center components, see [MBS+11].

Principles: Power Consumption in Data Centers

Common features of data center workloads:

High Idle Times	Average server utilization is between 10% and 50% for 80% of the time[1].
Bursty Workloads	Server load often occurs in bursts[2].
Periodicity	For some use cases (e.g. streaming services), workloads may occur with periodic patterns.

[1] BH07.

[2] VG17.

Principles

Virtual Machines

Principles: Virtual Machines

Terminology:

VM	virtual machine
VMM	virtual machine monitor
PM	physical machine
SLA	service-level agreement

Principles: Virtual Machines

- VMs represent "an efficient, isolated duplicate of the real machine"[3].
- For our purposes: VM $\approx$ OS + Applications.
- VMMs provision and manage CPU/memory/device allocation of VMs.
- VMs can be moved between physical machines without losing their state.

[3] PG74.

Principles: Virtual Machines

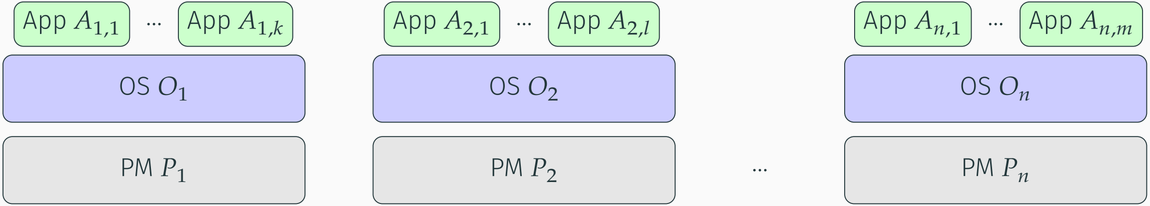

Figure 6: Data center architecture without virtualization

Principles: Virtual Machines

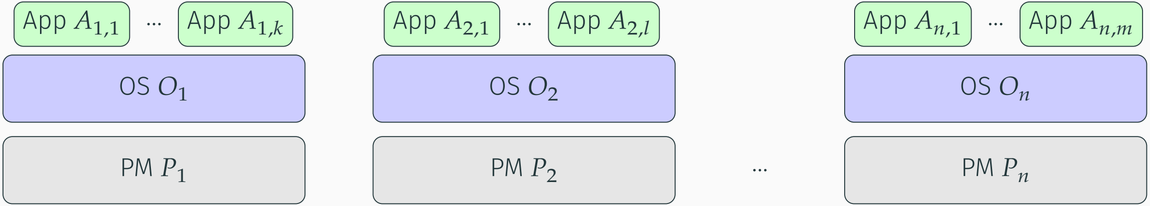

Figure 6: Data center architecture without virtualization

$\Longrightarrow$ All servers must be running to support SLAs, even when current demand for a particular application is low.

$\Longrightarrow$ Some programs may require a specific architecture (e.g. x86, ARM) or specific OS (e.g. Linux, Windows). This may lead to fragmentation.

Principles: Virtual Machines

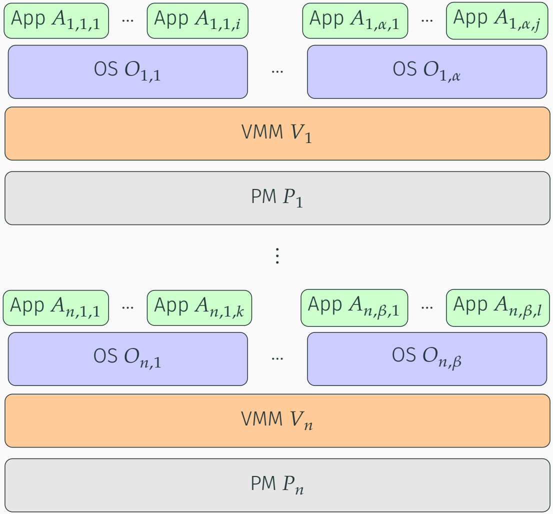

Figure 7: Data center architecture using virtualization

Principles: Virtual Machines

$\Longrightarrow$ Enables usage of uniform host platform (e.g. only CentOS servers running on x86).

$\Longrightarrow$ VMs can be moved between PMs to minimize the number of active PMs: energy savings.

$\Longrightarrow$ VMs can be provisioned on demand: scalability.

Optimization Approaches

Optimization Approaches

What criteria can we consider when deciding which VMs to migrate?

Optimization Approaches

Hardware Utilization

Optimization Approaches: Hardware Utilization

- CPU and memory utilization describe the load observed on a PM.
- High CPU/memory load on a single PM $\implies$ data center is overloaded.
- Low CPU/memory load over the whole system $\implies$ data center is underloaded.

Idea: Migrate VMs whenever the system is under- or overloaded.

Optimization Approaches: Hardware Utilization

Reactive Policy (Migration Controller, *MC*)

Rebalancing is done immediately when a threshold violation is detected.

On an overloaded PM, a VM is determined for migration to the least loaded PM capable of hosting the VM. If no such PM exists, a new server may be booted.

In an underload situation, the VMs on the least loaded PM are transferred to other PMs.

Optimization Approaches: Hardware Utilization

Proactive Policy (Workload Placement Controller, *WP*)

Rebalancing is done in predetermined intervals.

Historical data (resource utilization traces) are used to determine how much capacity is provisioned during each interval. During an interval, no additional VMs will be started, and no existing VMs will be stopped.

Mixed Policies

Historical data is used to predict the expected workload profile in the next interval.

When violations occur during an interval, the same migration strategy as above is used. In another variation, historical data is considered for rebalancing on every threshold violation.

Optimization Approaches: Hardware Utilization

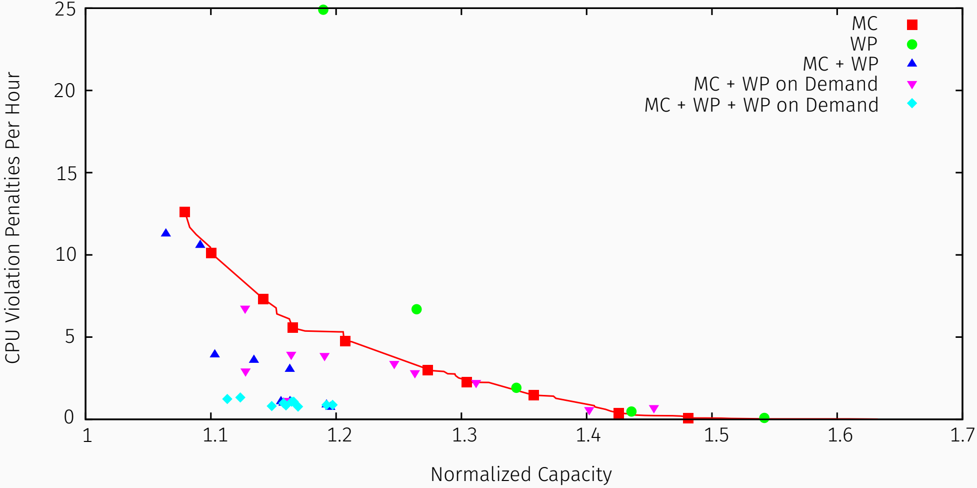

Figure 8: Performance of different provisioning policies [GRCK09]

Optimization Approaches

Network Traffic

Optimization Approaches: Network Traffic

- Many modern data center applications communicate over the network.
- Cost of communication differs based on network placement of VMs.
- Cost of communication also depends on the rate of communication between VMs.

Idea: Place VMs such that total communication cost is minimized.

Optimization Approaches: Network Traffic

$v_1, \dots, v_n$	VMs
$s_1, \dots, s_n$	Slots (a slot represents the ability to place a VM on a PM)
C_{ij}	Communication cost between slots s_i and s_j
D_{ij}	Traffic rate between VMs v_i and v_j
g_i	External communication cost of slot s_i
e_i	External traffic rate of VM v_i

Optimization Approaches: Network Traffic

$v_1, \dots, v_n$	VMs
$s_1, \dots, s_n$	Slots (a slot represents the ability to place a VM on a PM)
C_{ij}	Communication cost between slots s_i and s_j
D_{ij}	Traffic rate between VMs v_i and v_j
g_i	External communication cost of slot s_i
e_i	External traffic rate of VM v_i

Goal: Find a VM-to-slot mapping $\pi\colon [1, \dots, n] \to [1, \dots, n]$ that minimizes

$$\sum_{i,j=1,\dots,n} D_{ij} C_{\pi(i)\pi(j)} + \sum_{i=1}^{n} e_i g_{\pi(i)}.$$

Optimization Approaches: Network Traffic

- Unfortunately, finding an optimal π is NP-hard[4].
- $O(n^4)$ approximation algorithm:
  1. Cluster slots such that slots with low intercommunication rate are close to each other.
  2. Cluster VMs such that VMs with high intercommunication traffic are close to each other. Ensure clusters have the same size.
  3. Recursively map slot clusters to VM clusters.

[4] MPZ10.

Optimization Approaches: Network Traffic

Topology	Algorithm	Objective Function	CPU time
VL2	LOPI	0.9732	22
	SA	1.0000	27
	Cluster-and-Cut	0.9375	11
BCube	LOPI	1.0000	29
	SA	0.9860	35
	Cluster-and-Cut	0.8462	14

(Note: Objective function value has been normalized to 1.)

Figure 9: Comparison of different approximation algorithms on different network topologies. The data is obtained from trace-driven simulations. [MPZ10]

Optimization Approaches: Network Traffic

$\Longrightarrow$ Effectiveness depends on the underlying network topology.

$\Longrightarrow$ Only “sensibly” applicable when many VMs inside the same data center need to communicate with each other.

$\Longrightarrow$ Slightly better objective function value when compared to other approximations. On “favorable” network topologies, the improvement can range up to 15%.

$\Longrightarrow$ … but twice as fast!

Evaluation

Applicability

Not every optimization approach may be applicable (or effective) in every data center.

For example, Cluster-and-Cut can only reasonably be used as the dominant placement strategy when VMs inside the same facility show high intercommunication rates. (Probably not the case for consumer-oriented cloud computing providers but may very well be true for enterprise-oriented solutions.) Some strategies, however, (e.g. temperature-optimizing) may be more generally applicable.

Evaluation

Experimental Data

"New" migration/placement strategies are usually only simulated on traces of past data center workloads. Therefore, it is hard to evaluate how these strategies hold up under real-world conditions.

This is understandable from a business standpoint: No one wants to put their service infrastructure at risk.

Balancing Optimization Aspects

It is unclear how to simultaneously pursue different optimization strategies. For example, a CPU utilization strategy may produce a contradicting VM placement to a network traffic strategy. Balancing those aspects likely requires a good understanding of the local conditions and workload profiles. This, in turn, means more work for the data center controllers.

Conclusion & Outlook

Conclusion & Outlook

To summarize:

- Power consumption in data centers is influenced by a number of factors.
- Server consolidation by virtualization provides an effective tool for influencing these factors. Different strategies aim to optimize different factors.
- Placement strategies are not easy to compare with each other *a priori*.

Conclusion & Outlook

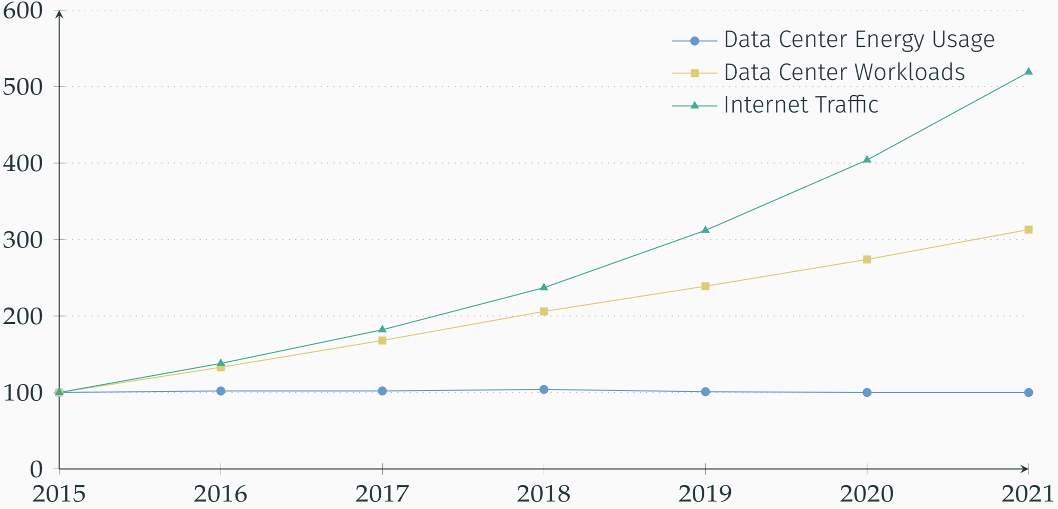

Figure 10: Projected relative data center energy usage [IEA19]

Conclusion & Outlook

In the future:

- As demand for computing power and more reliable online services increases, the field can be expected to stay and become even more relevant in the future.
- Even though Internet traffic will likely continue to grow, improvements in hardware and software are expected to offset increasing data center workloads.
- Containerization is becoming increasingly popular as an alternative to VMs.

Questions.

References

A. v. Beem, *HVAC air handler unit*, Jan. 2, 2015. [Online]. Available: https://commons.wikimedia.org/wiki/File:HVAC_Air_Handler_Unit,_pic2.JPG (visited on 01/04/2020).

L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *Computer*, vol. 40, no. 12, pp. 33–37, Dec. 2007, ISSN: 1558-0814. DOI: 10.1109/MC.2007.443.

References (II)

R. César, *24-port 3com switch*, May 13, 2008. [Online]. Available: https://commons.wikimedia.org/wiki/File:24-port_3Com_switch.JPG (visited on 01/04/2020).

Cgxke, *Large UPS being installed in a datacenter.* Oct. 22, 2008. [Online]. Available: https://commons.wikimedia.org/wiki/File:500kVA-UPS.jpg (visited on 01/04/2020).

D. Gmach, J. Rolia, L. Cherkasova, and A. Kemper, "Resource pool management: Reactive versus proactive or let's be friends," *Computer Networks*, vol. 53, pp. 2905–2922, 2009. DOI: 10.1016/j.comnet.2009.08.011.

References (III)

Hugovanmeijeren, *The CERN datacenter with world wide web and mail servers*. 2010. [Online]. Available: https://commons.wikimedia.org/wiki/File:Cern_datacenter.jpg (visited on 01/03/2020).

IEA. (2019), Data centres and data transmission networks – tracking buildings – analysis, [Online]. Available: https://www.iea.org/reports/tracking-buildings/data-centres-and-data-transmission-networks (visited on 12/15/2019).

References (IV)

E. R. Masanet, R. E. Brown, A. Shehabi, J. G. Koomey, and B. Nordman, "Estimating the energy use and efficiency potential of u.s. data centers," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1440–1453, Aug. 2011, ISSN: 1558-2256. DOI: `10.1109/JPROC.2011.2155610`.

X. Meng, V. Pappas, and L. Zhang, "Improving the scalability of data center networks with traffic-aware virtual machine placement," in *2010 Proceedings IEEE INFOCOM*, ISSN: 0743-166X, Mar. 2010, pp. 1–9. DOI: `10.1109/INFCOM.2010.5461930`.

G. J. Popek and R. P. Goldberg, "Formal requirements for virtualizable third generation architectures," *Commun. ACM*, vol. 17, no. 7, pp. 412–421, Jul. 1974, ISSN: 0001-0782. DOI: `10.1145/361011.361073`.

A. Varasteh and M. Goudarzi, “Server consolidation techniques in virtualized data centers: A survey,” *IEEE Systems Journal*, vol. 11, no. 2, pp. 772–783, Jun. 2017. DOI: 10.1109/JSYST.2015.2458273.