Eco4Cloud Scientific Overview

Carlo Mastroianni
mastroianni@eco4cloud.com

Natural Efficiency

Data on data centers

Regular vs. hyperscale data centers

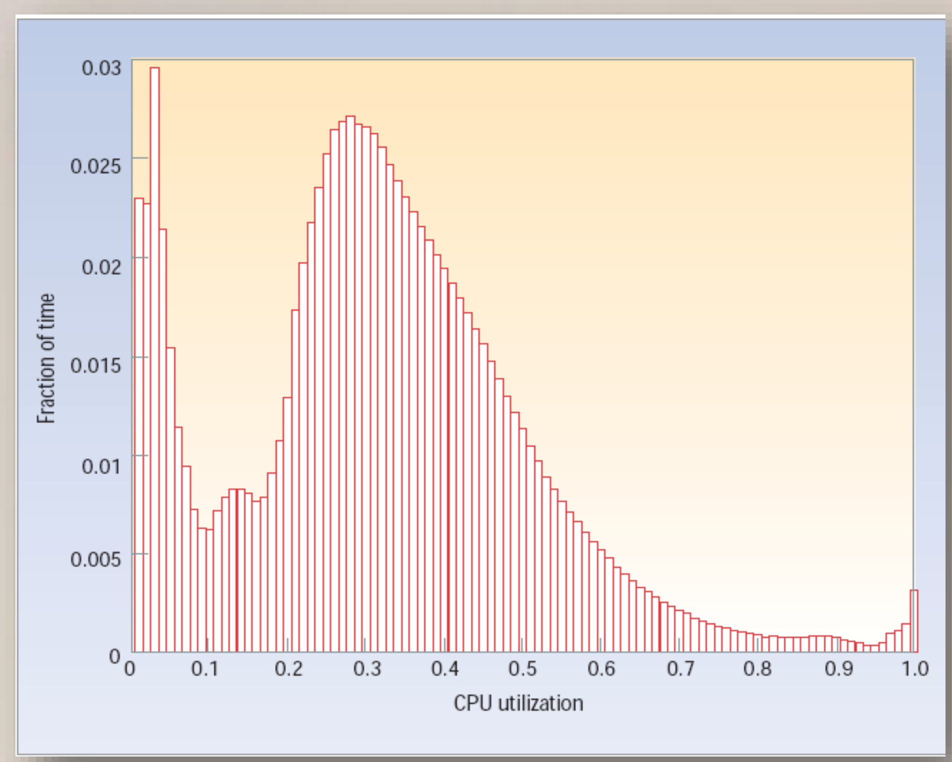
Reliable Datacenter PLC		Colossalcloud inc	
Servers/IT equipment • Branded servers • Bought by customer?	Full price (100%)	Servers/IT equipment • ODM or custom built • Large volume discount	50%
IT efficiency (utilization) • Few VMs per server • Storage utilization low	15%	IT efficiency (utilization) • Many floating VMs per server • Storage managed	30%
Throughput per watt • Moore's Law • Servers swapped out slowly	2x every 2-6 years	Throughput per watt • Moore's Law drives performance • Servers swapped out rigorously	2 x every 2 years or less
Energy price • Must pay local rates	12 cents per KWh	Energy price • Site selection secures low price	3 cents per KWh
Energy overhead (PUE) • Higher tier limits energy savings • Site, customers limit innovation	x 1.7	Energy overhead (PUE) • Engineers drive down energy • Business model allows low tiers	x 1.3
Facility build costs • Higher tiers cost more • City sites cost more, limit space	€12m per MW	Facility build costs • Reduced infrastructure costs less • Remote sites cost less	€6m per MW
Financing costs • Capital may need to be raised • VAT, rates, income tax can be high	3-8% capital & taxes	Financing costs • Cash on balance sheet or cheap • VAT, rates, income tax negotatied	Very low

Source: Andy Lawrence, 451 Research, "Datacenters in a cloud storm"

Inefficient utilization of servers

Two sources of **inefficiency**

- Servers are underutilized (between 15% and 40%)
- An idle server consumes more than 50% of the energy consumed when fully utilized

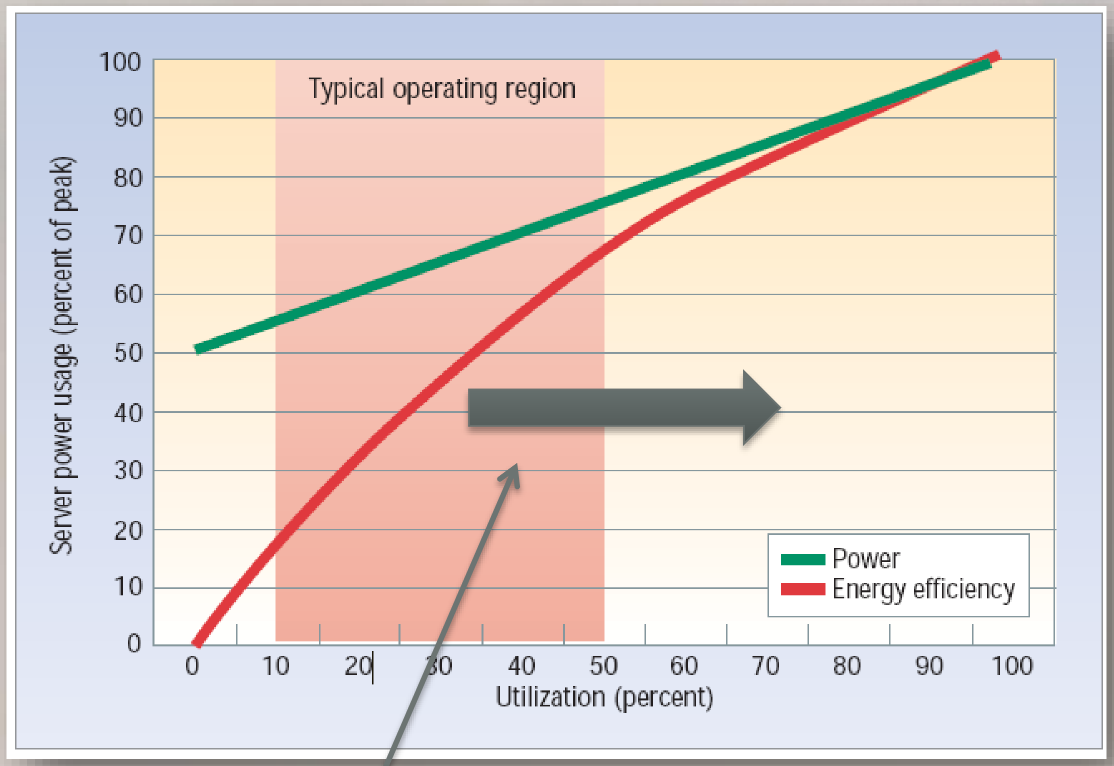


Typical utilization of servers

This means that it is generally possible to **consolidate** the load on **fewer and better utilized servers!**

Source: *L.Barroso, U.Holzle, The case of energy proportional computing, ACM Computer Journal, Volume 40 Issue 12.*

Improving efficiency through consolidation



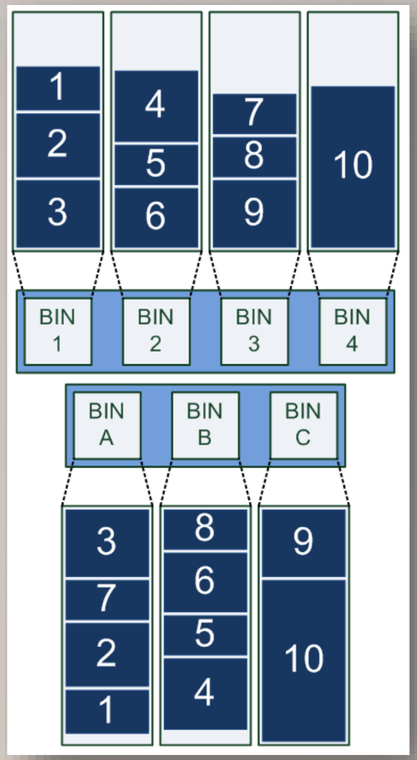
Energy efficiency is utilization divided by power consumption (useful workload/W)

Energy efficiency is **low** in the typical operating region

Consolidation of the workload means shifting the typical operating region to the right, in this way increasing the energy efficiency

Approaching the consolidation problem

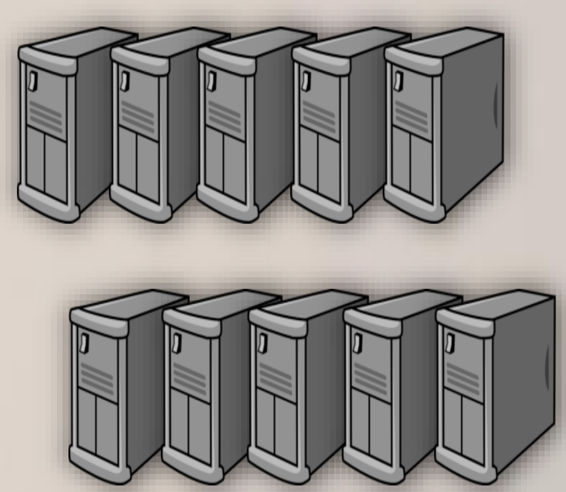
The consolidation problem is a form of **Bin Packing Problem**:



Goal: pack a **collection of VMs** into the **min. number of servers,** so as to hibernate the remaining servers, and save energy.

Issues:

- **NP-Hard** problem: heuristics exist, but their scalability is limited.
- In DCs, this is a **multi-dimensional** problem (CPU, disk, memory, network).
- Load requirements are **highly dynamic:** VMs must be repacked with few and asynchronous migrations
- **Maximize QoS**: prevent overload events even when resources utilization is increased



Known solutions for consolidation

- **Best Fit**: each VM is assigned to the server whose load is the closest to a target (e.g. 90%)

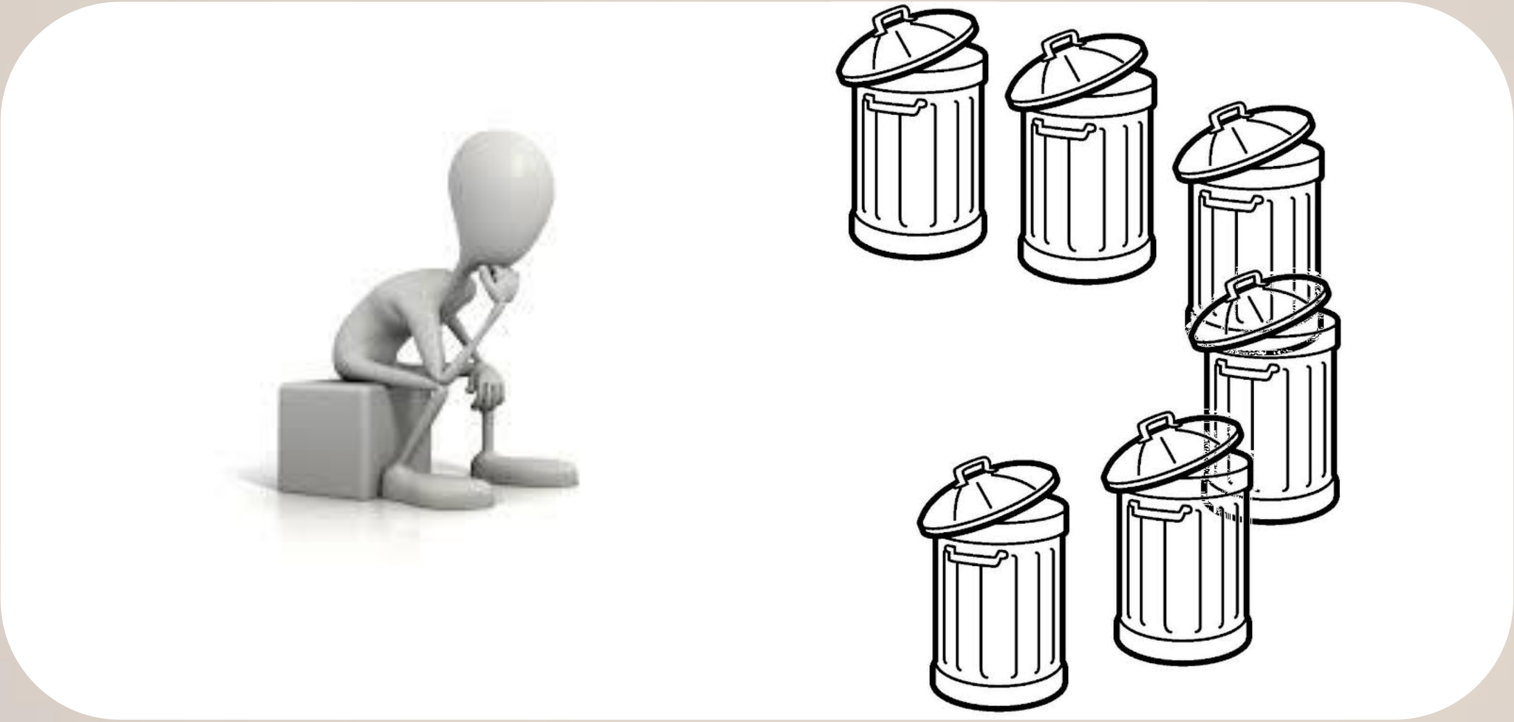
  This only guarantees a performance ratio of 17/10: at most 17 servers are used when the minimum is 10

- **Best Fit Decreasing**: VMs are sorted in decreasing order, then assigned with Best Fit

  Performance ratio is 11/9, but sorting VMs may not be easy in large data centers, and many concurrent migrations are needed

- **DPM** of VMWare adopts a greedy algorithm

  Servers are sorted according to numerous parameters (capacity, power consumption, etc.). DPM scans the list and checks if servers can be unloaded



E4C Superiority over VMware DPM

☺	😐	eco4cloud	VMware DPM
Energy Saving		Eco4Cloud reduces energy consumption (among 30% and 60%) by improving computational efficiency and consolidating the workload.	VMware DPM reduces energy consumption by switching off servers in accordance to a predefined order.
Virtual Machines Consolidation		Eco4Cloud adopts a distributed and self-organizing algorithm to consolidate VMs on the minimum number of servers in real time, so as to achieve energy and cost savings.	DPM is a tool included in the VMware DRS package. The DPM objective is workload consolidation, but the inclusion within DRS is troublesome, as the objective of DRS, load balancing, is opposed to consolidation.
Scalability		The consolidation problem is very complex (NP) when approached in a centralized fashion. Eco4Cloud is scalable because it adopts a self-organizing/probabilistic approach, which decentralizes most of the intelligence to single servers.	DPM is poorly scalable because it adopts a classical centralized approach. Servers are ordered according to a number of parameters and they are switched off according to that order. This procedure is not efficient in data centers with a large number of servers.
Controlled behavior		Eco4Cloud behavior is transparent and users can dynamically select among manual/semi-automatic/automatic modes.	DPM does not allow the user to control and monitor the distribution of the workload. It is our evidence that VMware users generally prefer not to activate DPM at all because its behavior is not clear and it is not controllable/configurable.
Heterogeneous hypervisors support		Eco4Cloud works on top of any virtualization platform: VMware, Microsoft HyperV, KVM etc.	DPM only works in top of the VMware platform.
Inter-Cloud Management		Eco4Cloud has patented an algorithm for the efficient management of workload in geographical data centers. The inter-cloud algorithm combines intra-cloud consolidation with optimal distribution of workload among remote data centers. Objectives may be specialized for specific deployments: inter-cloud load balancing, reduction of costs, consumed energy and carbon emissions, etc.	DPM consolidates the workload only within a single data center.

Eco4Cloud algorithm

ICAR-CNR researchers have devised and developed a **very effective** and **scalable** solution, based on the swarm intelligence paradigm.

PROBLEM INEFFICIENCY OF CONSOLIDATION ALGORITHMS

The solutions available today are **semi-manual**, extremely **complex**, **poorly adaptive**, **not scalable**.

↓

SOLUTION INNOVATIVE BIO-INSPIRED APPROACH

The ICAR-CNR solution uses a bio-inspired probabilistic approach to assign Virtual Machines to servers. The solution is **automatic**, **simple**, **adaptive** and **highly scalable**.

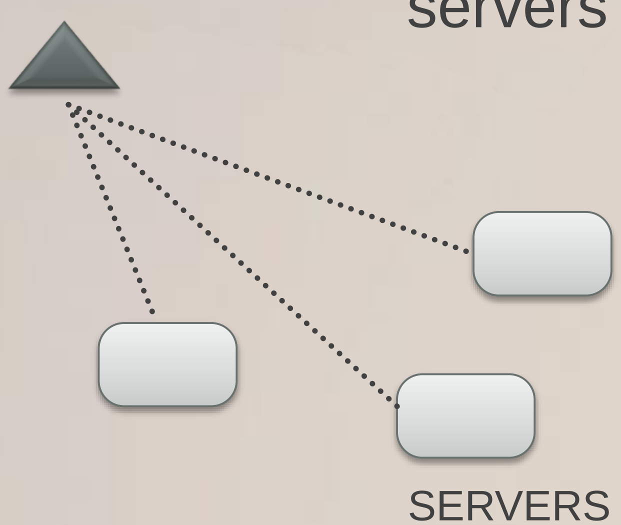
- *C. Mastroianni, M. Meo, G. Papuzzo, "Probabilistic Consolidation of Virtual Machines in Self-Organizing Cloud Data Centers". IEEE Transactions on Cloud Computing, vol. 1, n. 2, pp. 215-228, 2013.*
- ***PCT Patent** "System for Energy Saving in Company Data Centers"*

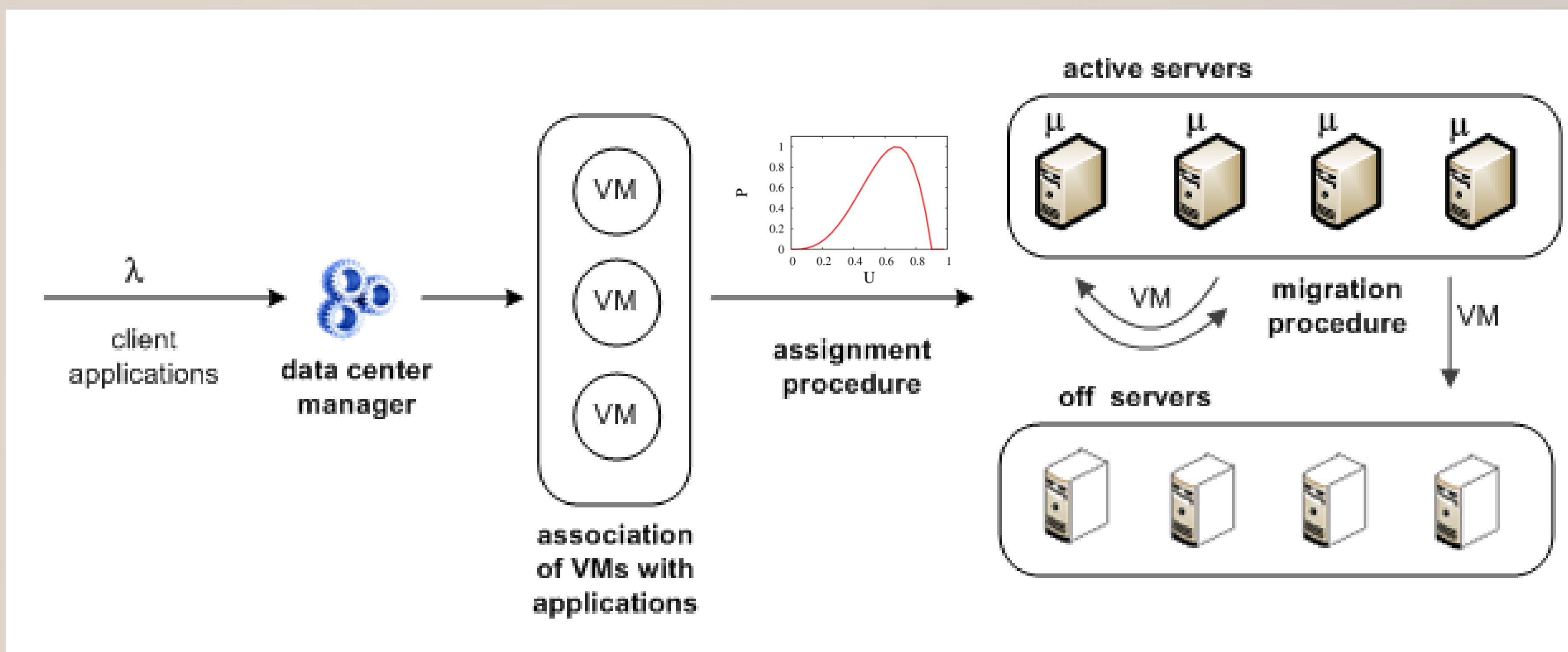
Eco4cloud algorithm in action

The data center manager assigns and migrates VMs to servers based on local probabilistic trials:

- **Lightly loaded** servers tend to **reject** VMs
- **Highly loaded** servers tend to **reject** VMs
- Servers with **intermediate** load tend to **accept** VMs
- Eventually, the workload is distributed to a low number of highly utilized servers

DATA CENTER MANAGER

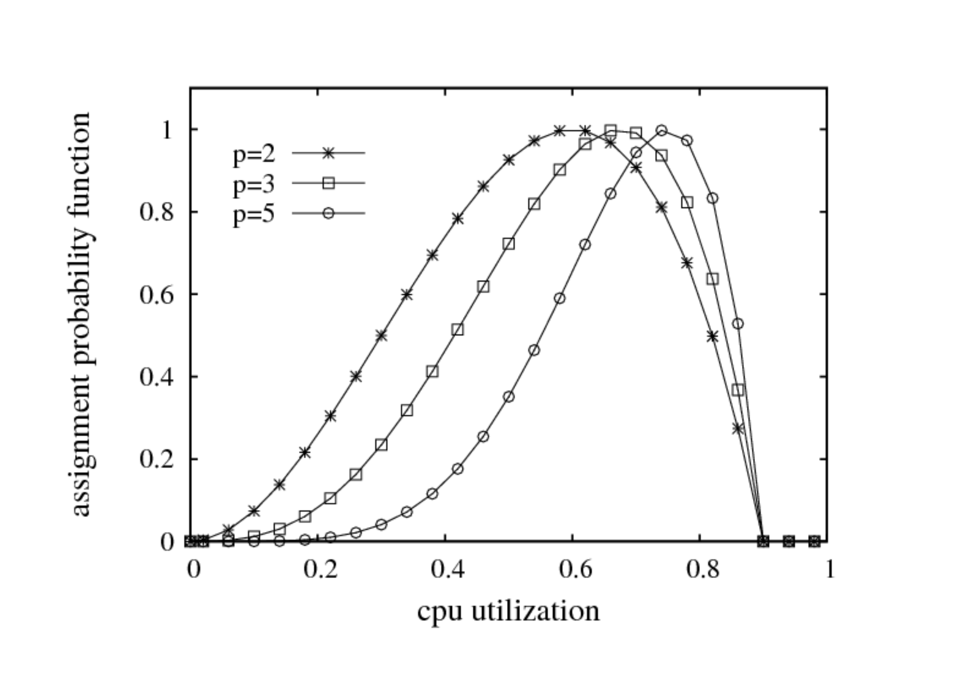




VM assignment/migration

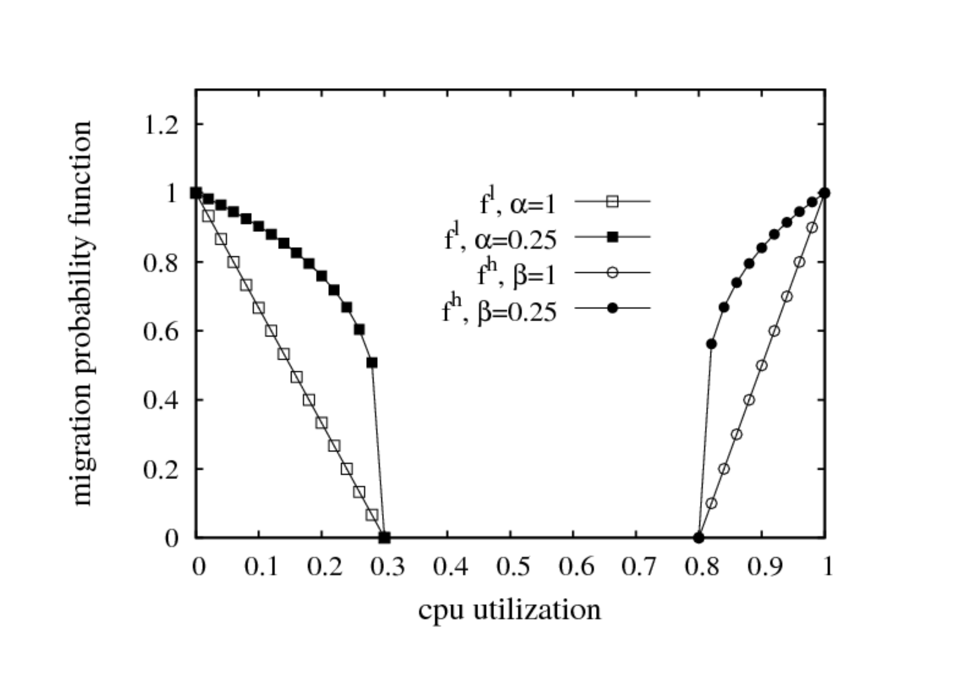
Assignment procedure

1. The manager sends an **invitation** to a subset of servers
2. Each server evaluates the **assignment probability function (Bernoulli trial)** based on the utilization of local resources (e.g. CPU, RAM…) and sends a positive ack if it is available
3. The manager collects positive replies and selects the server that will execute the VM



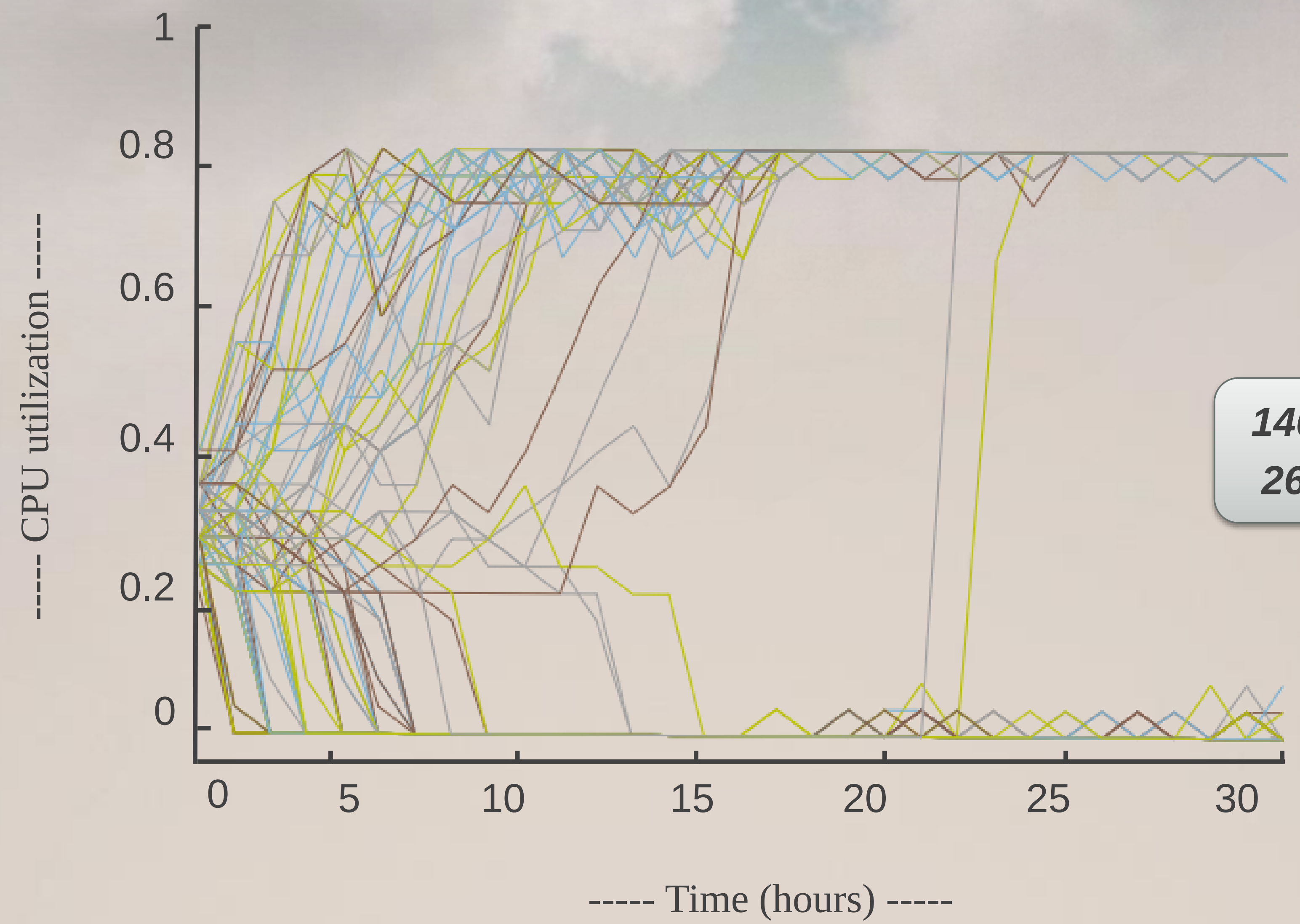
Migration procedure

1. A server checks if its load is in the range between a low and a high threshold
2. When utilization is **too low/high**, the server performs a Bernoulli trial based on the **migration probability function**
3. If the trial is positive, some VMs are migrated
4. Destination servers are determined with a new reassignment procedure



Consolidation Snapshot

(400 servers and 6000 VMs)



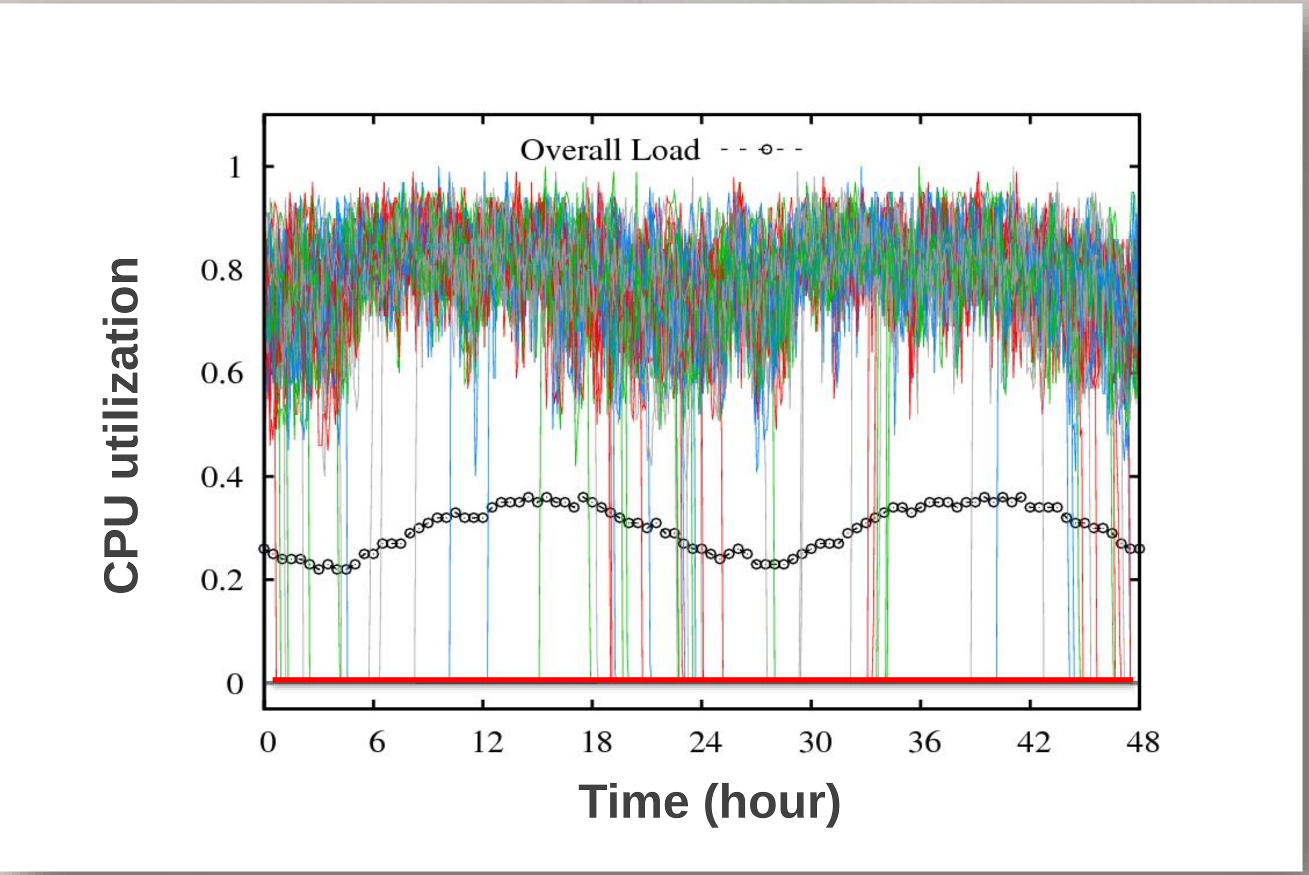
***140** servers take all the load*

***260** servers are hibernated*

- **Energy Savings:** before consolidation, VMs are running at between 20-40% usage. After 15 hours, all VMs are either close to optimal values (80% usage) or hibernated
- **SLAs:** Utilization is not allowed to exceed 85%, providing complete protection of the physical resources and adherence to SLAs

CPU Utilization in steady conditions

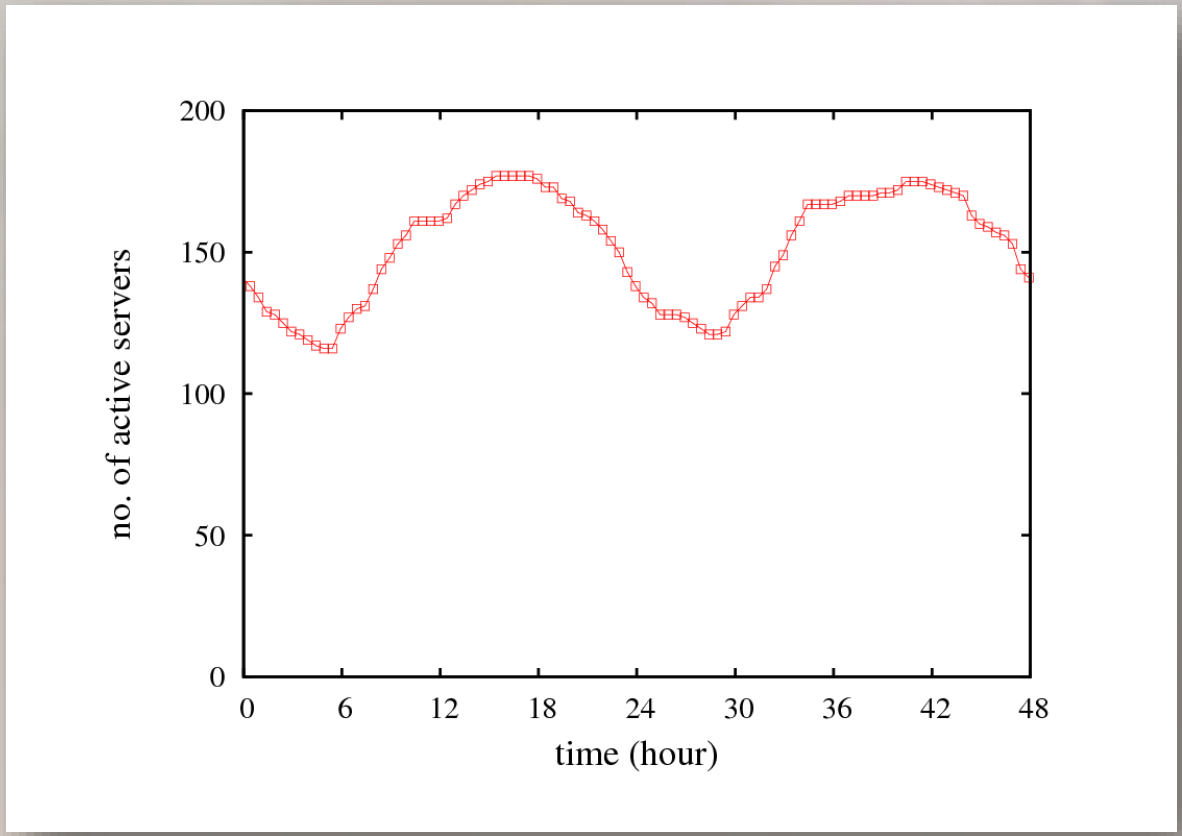
(48 hours: overall load shown as a reference)



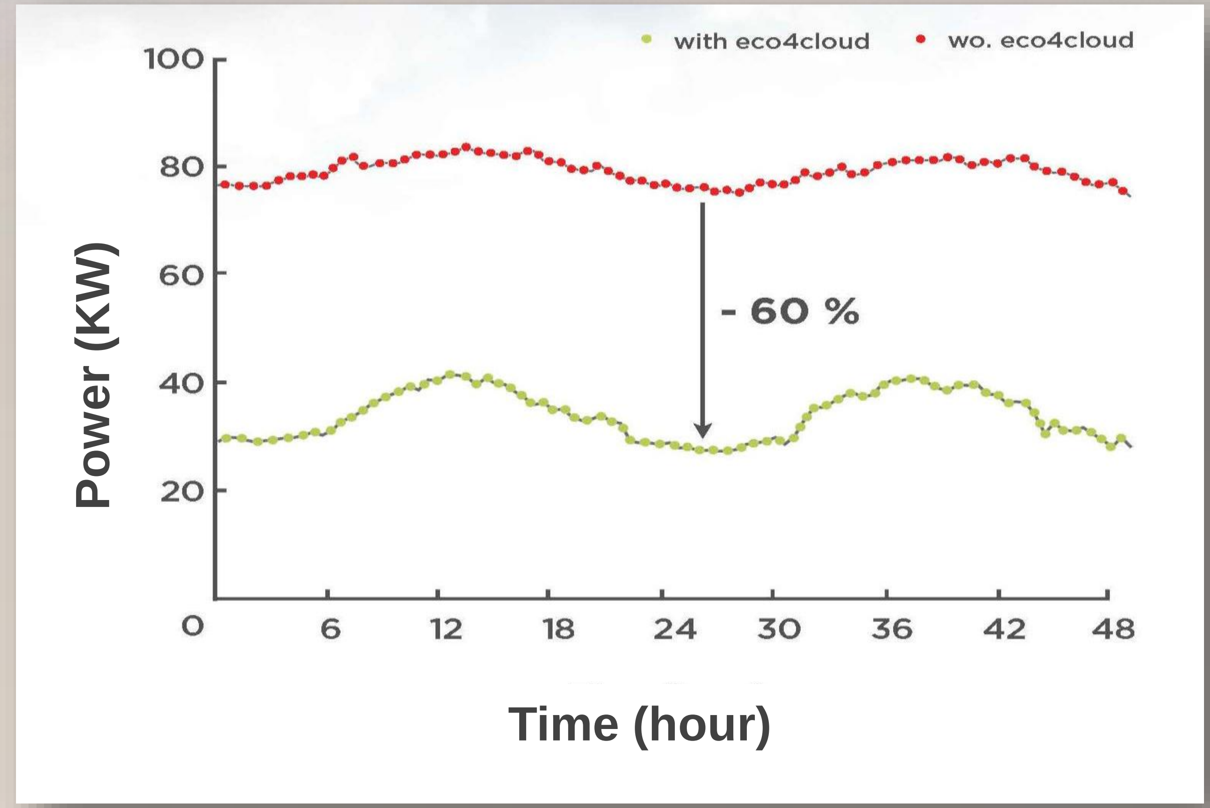
- CPU utilization of active servers is always **between 0.5** and **0.9**
- Many servers are **hibernated** (bottom line)

Active servers and consumed power

Number of active servers



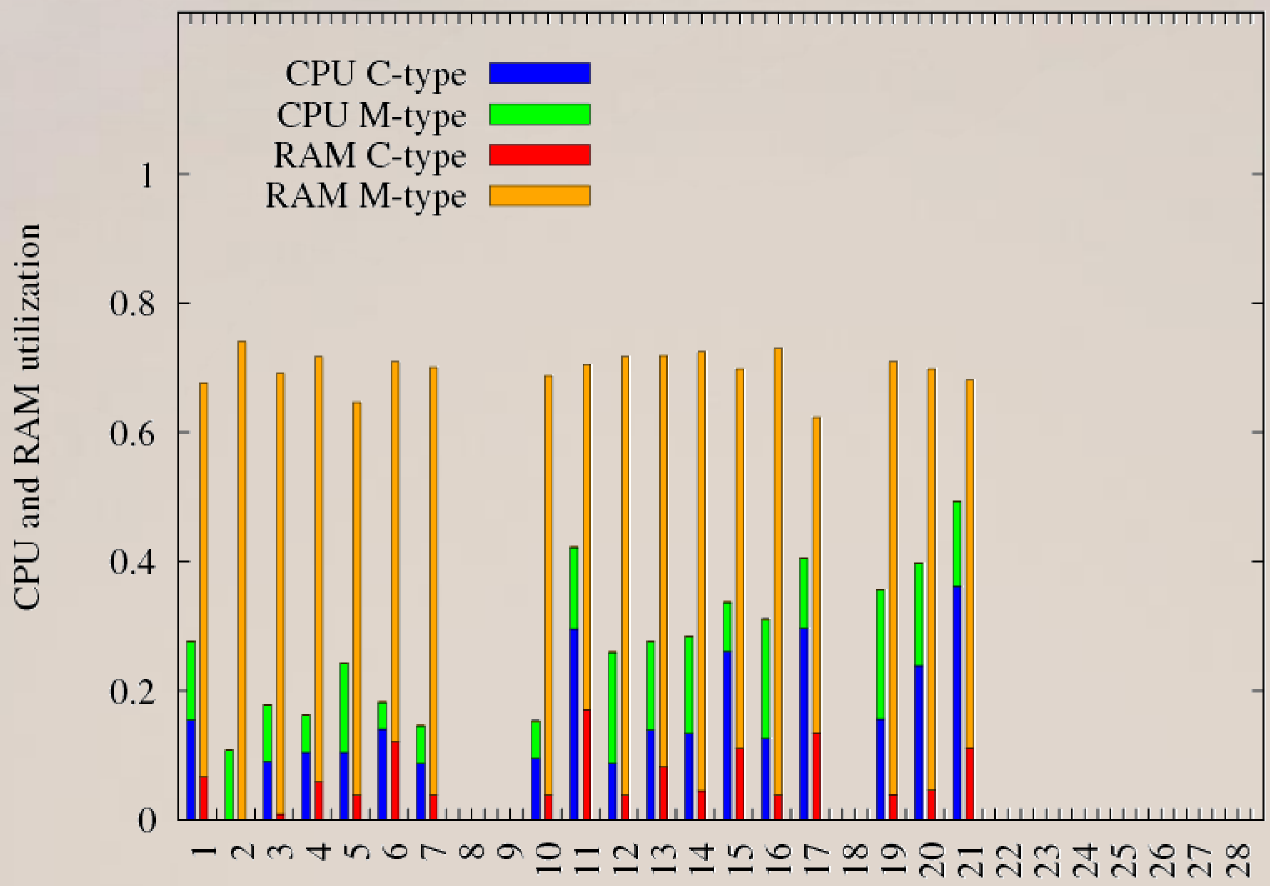
Consumed power



- The number of active servers **follows the overall workload**, and so the power
- Many servers are never activated: they can be safely devoted to other applications
- Power savings up to **60%**!
- More savings are obtained thanks to **decreased cooling needs**

Multi-resource consolidation

- Workload is consolidated on the most utilized resource (RAM in this case)
- VMs with different characteristics (here, CPU-bound and RAM-bound) are **balanced** → hardware resources are exploited efficiently

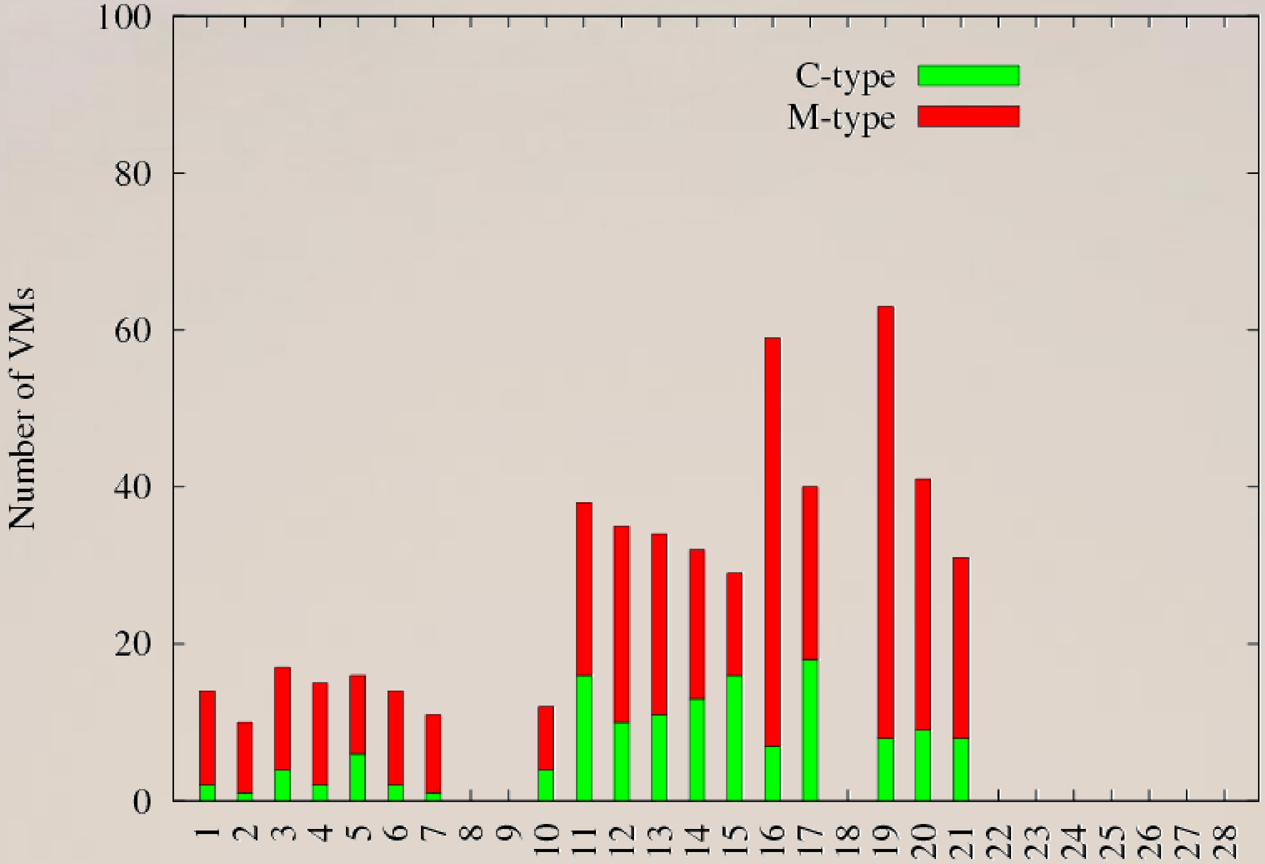


RAM and CPU utilization of 28 servers, separately considered for CPU-bound and RAM-bound VMs

C-type = CPU-bound
M-type = RAM-bound

Workload balancing

- VMs of the two types are distributed among the servers in a proportion that is always comparable to the overall proportion observed in the whole data center



C-type = CPU-bound
M-type = RAM-bound

Number of C-type and M-type VMs running of 28 servers

Mathematical Analysis (single resource)

The assignment process (without migrations) can be modeled with fluid-like differential equations:

$$\boxed{\frac{\partial u_s(t)}{\partial t} = -\mu(t)u_s(t) + \lambda(t)A_s(t)}$$

$$\boxed{s = 0, \cdots, N_s - 1}$$

- **$u_s(t)$** is the CPU utilization of server s
- **$\lambda(t)$** is the arrival rate of VMs
- **$\mu(t)$** is the service rate
- **$A_s(t)$** is portion of VMs that are assigned to server s (to be computed, depends on **f_a**)

The exact computation of **$A_s(t)$** is costly, but the model can be simplified

Mathematical Analysis (single resource)

The portion of VMs assigned to server **s** (**$A_s(t)$**) is assumed to be proportional to the assignment probability evaluated on the server (**$f_a(u_s)$)**)

$$\frac{\partial u_s(t)}{\partial t} = -\mu(t)u_s(t) + \lambda(t)\frac{f_a(u_s(t))}{\sum_{i=0}^{N_s-1}(f_a(u_i(t)))}$$

$$s = 0, \cdots, N_s - 1$$

The rate of incoming VMs is normalized

The equations are useful to:

- better understand system dynamics
- perform parameter sweep and what-if analysis
- validate results obtained with simulations and real testbeds

Mathematical Analysis (multiple resource)

differential equations in the case of two-resource consolidation (CPU and RAM)

$$\frac{\partial u_s^{(C)}(t)}{\partial t} = -N_c\, N_v\, \mu\, u_s^{(C)}(t) + K\, \gamma_C\, \lambda^{(C)}(t)\; f_s(t),$$

$$\frac{\partial u_s^{(M)}(t)}{\partial t} = -N_c\, N_v\, \mu\, u_s^{(M)}(t) + K\, \lambda^{(M)}(t)\; f_s(t),$$

$$\frac{\partial m_s^{(C)}(t)}{\partial t} = -N_c\, N_v\, \mu\, m_s^{(C)}(t) + K\, \lambda^{(C)}(t)\; f_s(t),$$

$$\frac{\partial m_s^{(M)}(t)}{\partial t} = -N_c\, N_v\, \mu\, m_s^{(M)}(t) + K\, \gamma_M\, \lambda^{(M)}(t)\; f_s(t).$$

$$u_s(t) = u_s^{(C)}(t) + u_s^{(M)}(t),$$
$$m_s(t) = m_s^{(C)}(t) + m_s^{(M)}(t).$$

- $u_s^{(C)}, u_s^{(M)}$ fractions of CPU used by CPU- and RAM-bound VMs in server *s*
- $m_s^{(C)}, m_s^{(M)}$ fractions of RAM used by CPU- and RAM-bound VMs in server *s*
- $f_s(t)$ acceptance probability
- $\mu, \lambda^{(C)}, \lambda^{(M)}$ service rate and arrival rates of CPU- and RAM-bound VMs
- N_c, N_v number of cores per server and max number of VMs on a core
- K normalization factor

Benefits of the Eco4Cloud solution

- **Energy saving.** Power consumption reduced between 20% and 60%!
- **Highly scalable.** Thanks to its adaptive/self-organized distributed algorithm, the approach is extremely scalable
- **Capacity Planning.** Optimal occupancy of physical resources and **adaptive** optimization of inherently variable workloads
- **Minimal impact on operations**. Migrations are **gradual** and **asynchronous**
- **Efficient balancing** of heterogeneous applications
- **Meet DC SLAs.** Thanks to the insights and real-time monitoring analytics provided by E4C, data center managers can proactively/predictively prevent SLA violations and increase overall data center reliability
- **Virtualization environment independent**: VMWare vSphere, Microsoft Hyper-V, KVM,…

Scientific Recognitions

Eco4Cloud Tutorial at IEEE/ACM CCGrid 2014
14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, highly-reputed scientific event on Cloud Computing

May 26-29, 2014 - Chicago, IL, USA

Scientific paper published in one of the top international journals in this field, IEEE Transactions on Cloud Computing, and selected by the journal as flag paper for 2014.
Title of the paper: 'Probabilistic Consolidation of Virtual Machines in Self-Organizing Cloud Data Centers'

Whitepaper "Saving energy in datacenters through workload consolidation" co-authored with:

- Institute for High Performance Computing and Networking of the Italian National Research Council
- Dep. of Electronics and Telecommunications at Politecnico di Torino
- eERG – Energy Department at Politecnico di Milano
- PrimeEnergyIT/EfficientDataCenters

Piazza Vermicelli
Rende (CS) Italy
ph +39 0984 49 42 76
--
www.eco4cloud.com
info@eco4cloud.com
www.linkedin.com/company/eco4cloud