Jacques Cartier, November 2012

Dynamic consolidation challenges for virtualized data center

Jean-Marc Menaud

Ascola team EMNantes, INRIA, LINA.

ECOLE DES MINES DE NANTES

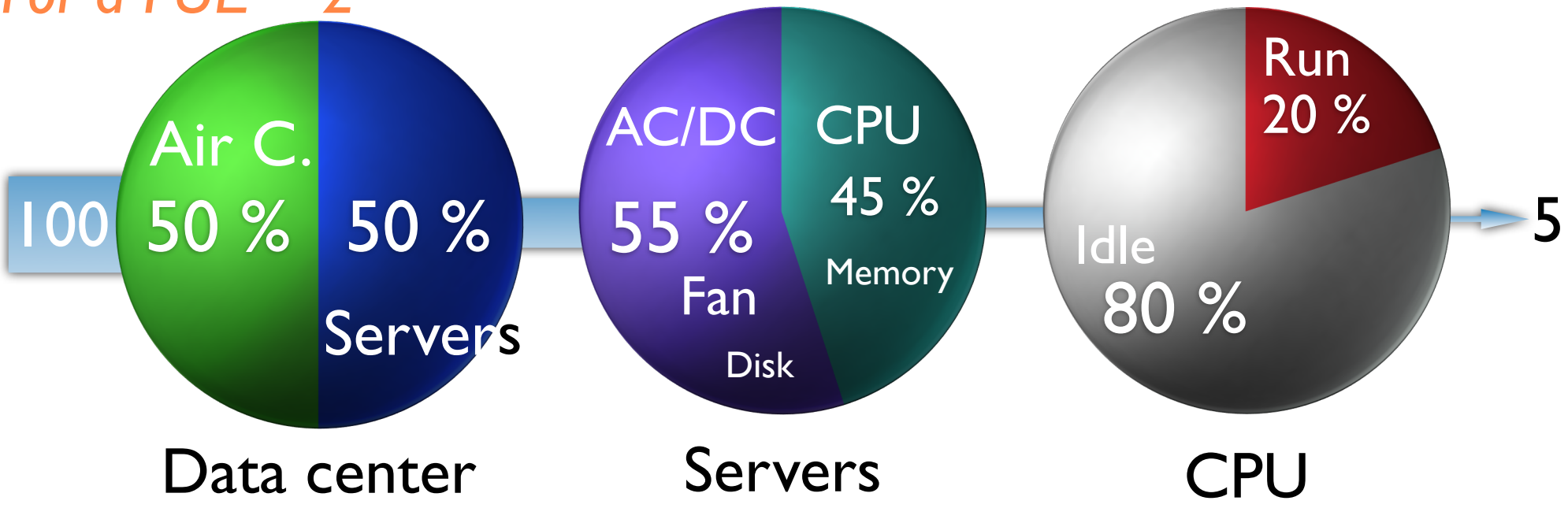
INRIA lina

Motivations

- **Increasing popularity of Cloud Computing solutions**
  - **Data-centers (DCs) are amazingly growing**
    - **DC providers have to face with energy consumption concerns**

Consequences

For a PUE = 2



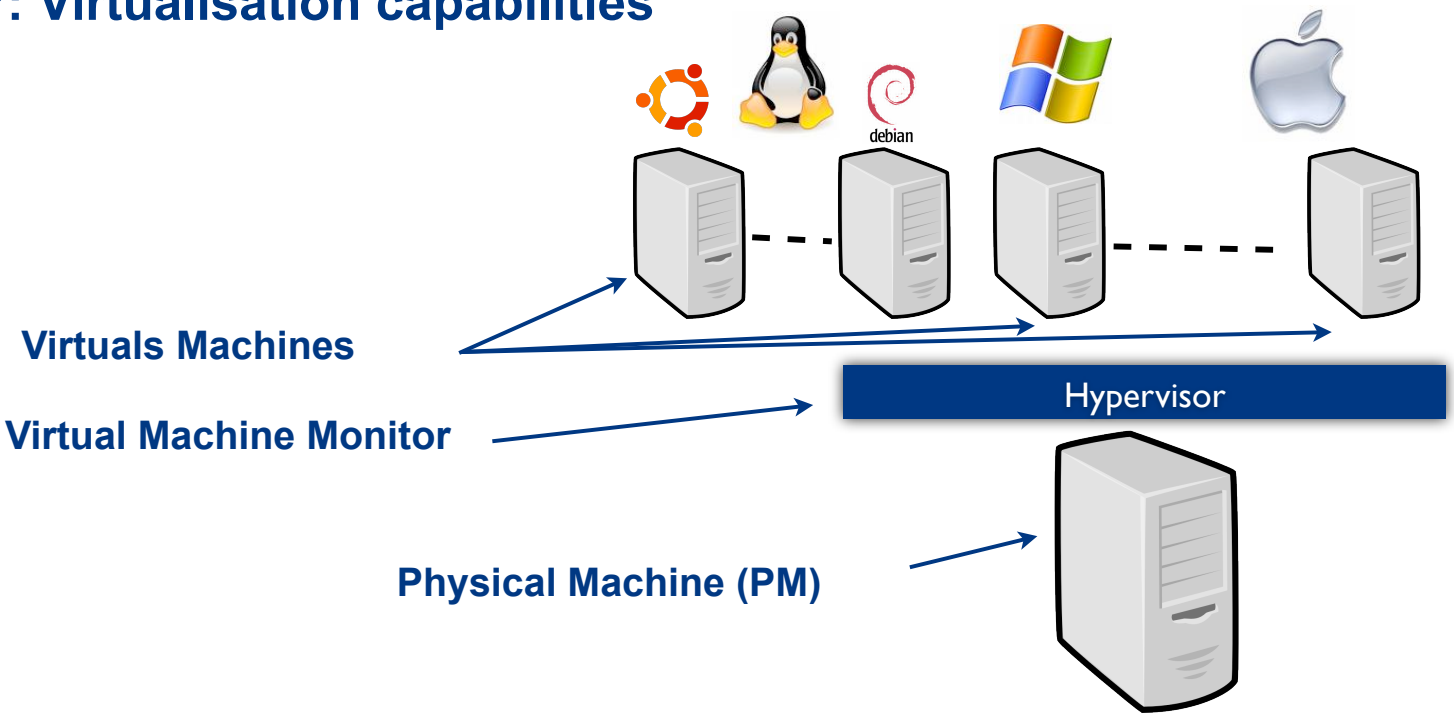
- **Analysis of the cost of a 2 MegaWatts DC (5000 nodes, 400w/h)**
  - PUE of 2, 0.06€/kWh => 2 120 886 €
  - A decrease of 5% enables a gain of 110K€

Managing DC resources finely becomes a major challenge

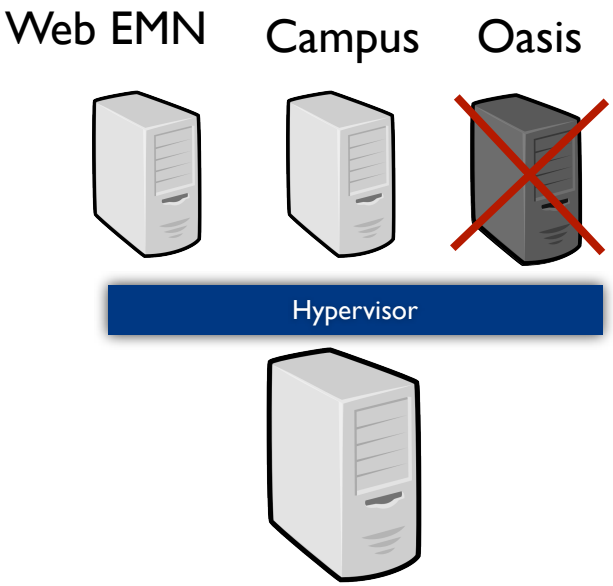
Consolidation

- **Consolidation (virtualization effect) :**
  - Consolidating to virtual machines reduces the number of running nodes

    So ernegy consumption
  - Reduces hardware costs while providing more efficient node
- **How ?: Virtualisation capabilities**



Virtualization capabilities (1/2)



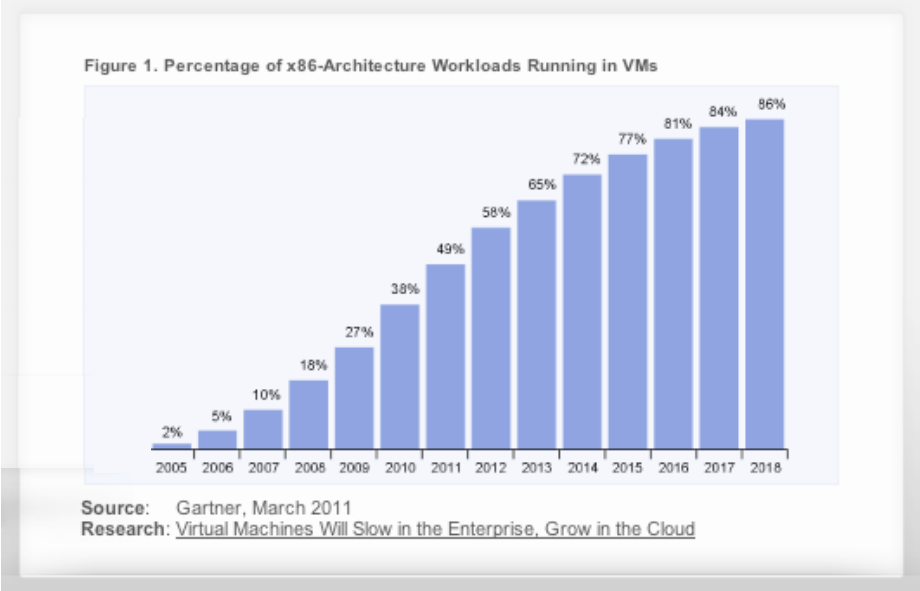
- **Isolation (security between VM)**
- **suspend/resume/reboot (maintenance)**

Virus / Invasion / Crash

Consolidation, some statistics

- **A constant progression**
- **Q3 2011 [2011-07]**
  - virtualization penetration rate: **38.9%**
  - Ratio of virtual machines to physical hosts: **5:1**
  - Primary Hypervisor Usage for Server virtualisation: **ESX 67,5%**

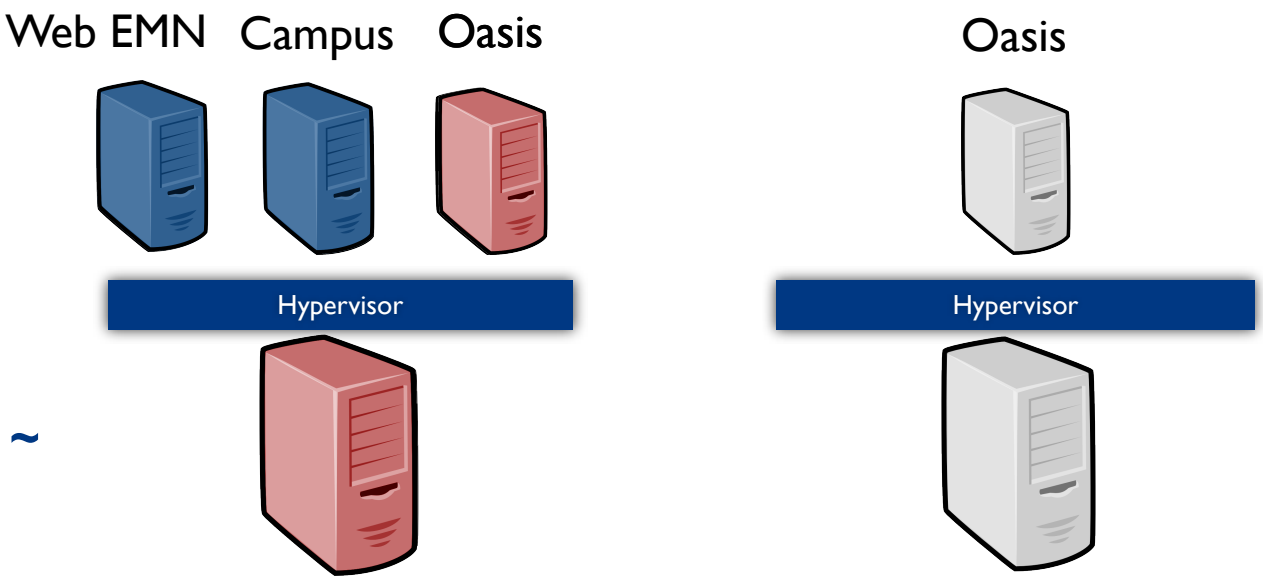
	Avg	UK	FR	DE	US
VMware	67.6	79	65	61	66.5
Citrix	14.4	12	15	20	12.5
Hyper-V	16.4	8	17	16	20.5
Other	1.6	1	3	3	0.5



Gartner March 2011

Dynamic consolidation

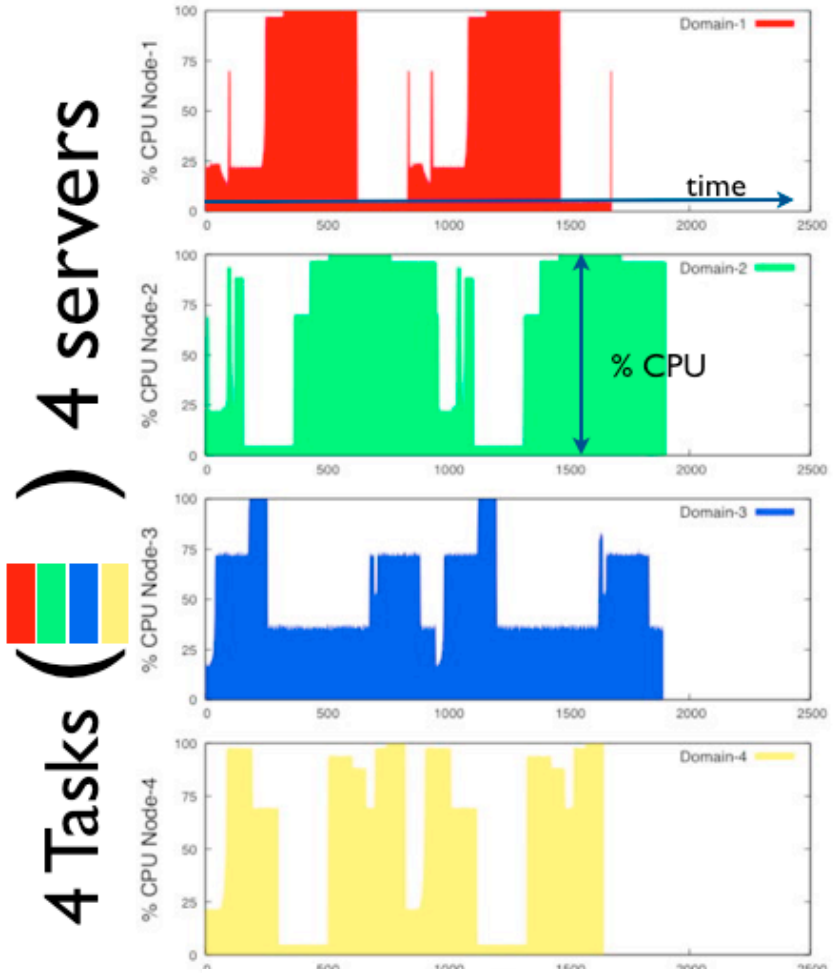
Virtualization capabilities (2/2)



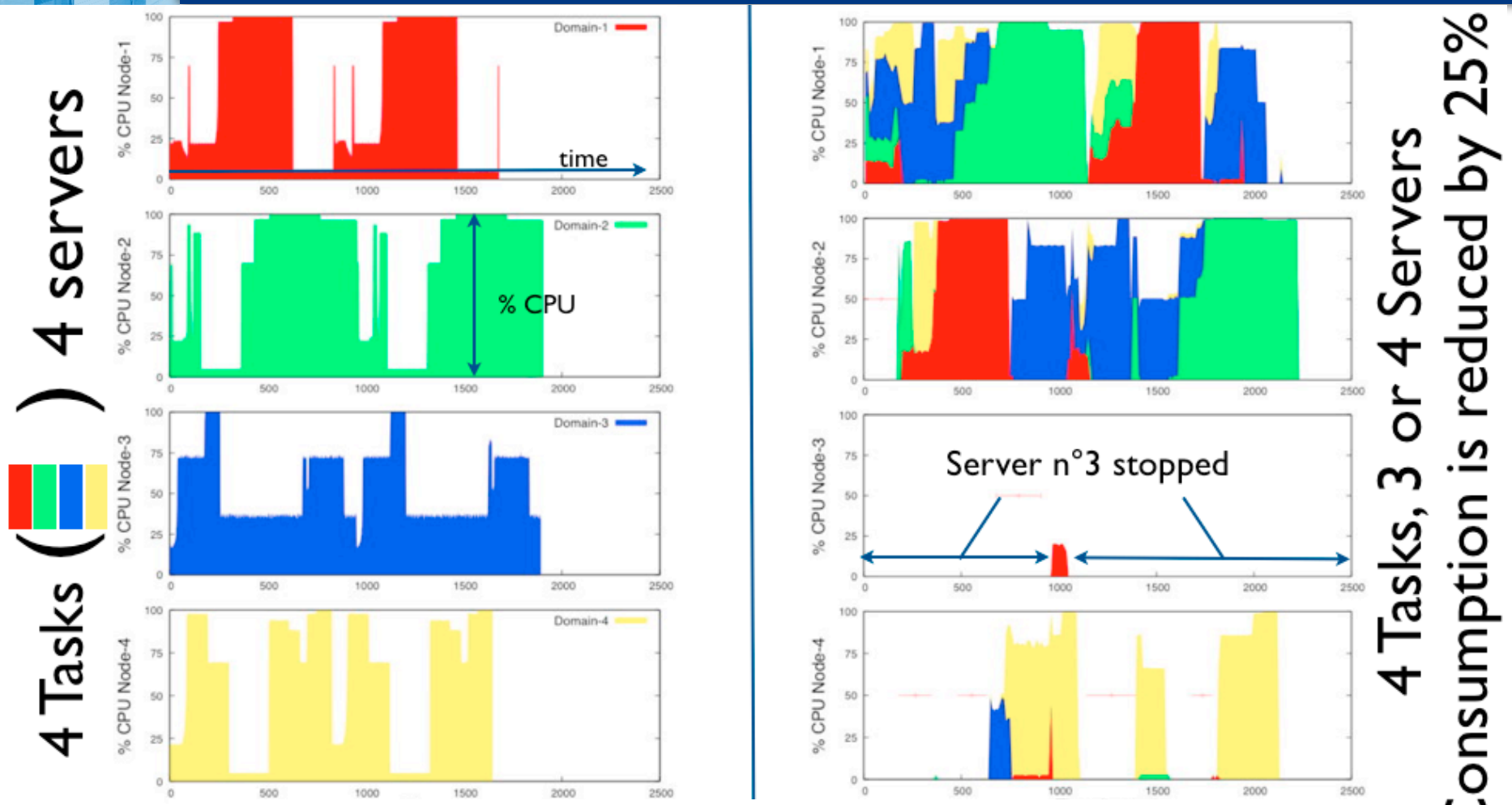
- **Live migration (load-balancing)**
- **High Availability(downtime ~ 60 ms)**
- **Dynamic Consolidation :**
  - The resources are allocated depending on the VM needs
  - VMs are mixed to be hosted on a reduced number of nodes
  - Servers unused can be turned off
  - VMs are remixed when it is necessary

Dynamic consolidation
btrPlace: Principles

[2006-15] *ex-Entropy*



btrPlace: Optimizing the placement of virtual servers



Dynamic consolidation for Energy Management
Some approaches

- **Virtual Machine Placement Problem (VMPP) is similar to the multi-dimensional bin packing problem know to be NP-Hard ... [2007-02]**
- **Heuristic methods**
  - Greedy algorithms Ex: EnaCloud [2009-03]

    Construct a solution by taking local decision without backtrack.
    First-Fit Decrease (FFD), Best-Fit (BF), Worst-Fit (WF), Next-Fit (NF) ... [1997-01]
    Pro: Ease to implement, good worst-case complexity
    Cons: No optimal solution, not realy flexible
  - Metaheuristic Ex: Snooze [2012-04]

    Probailistic algorithms by searching near optimal solution
    Genetic, Tabu, Ant colony, Graps ...
    Pro: Better solution than Greedy algorithms
    Cons: No optimal solution, not realy flexible
- **Exact mehods**
  - Mathematical Ex: Entropy [2009-06]

    Linear or Constraint programming [1986-05]
    Compute optimal solution
    Pro: optimal and flexible
    Cons: Exponantial time solving process

Dynamic consolidation for Energy Management
Some approaches

- **Virtual Machine Placement Problem (VMPP) is similar to the multi-dimensional bin packing problem know to be NP-Hard ... [2007-02]**
- **Heuristic methods**
  - Greedy algorithms Ex: EnaCloud [2009-03]

Mainly based on one or two dimension(s) (CPU,RAM),
on homogenous platform,
focus on one concern

Cons: No optimal solution, not realy flexible

- **Exact mehods**
  - Mathematical Ex: Entropy [2009-06]

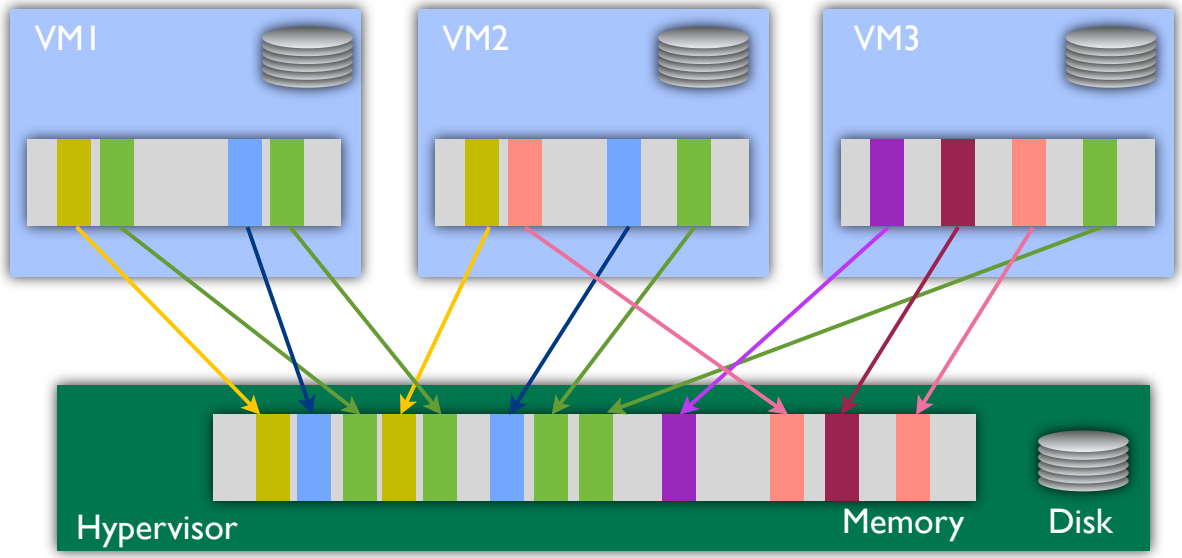
Linear or Constraint programming [1986-05]
Compute optimal solution
Pro: optimal and flexible
Cons: Exponantial time solving process

Which resource take account, and many ?

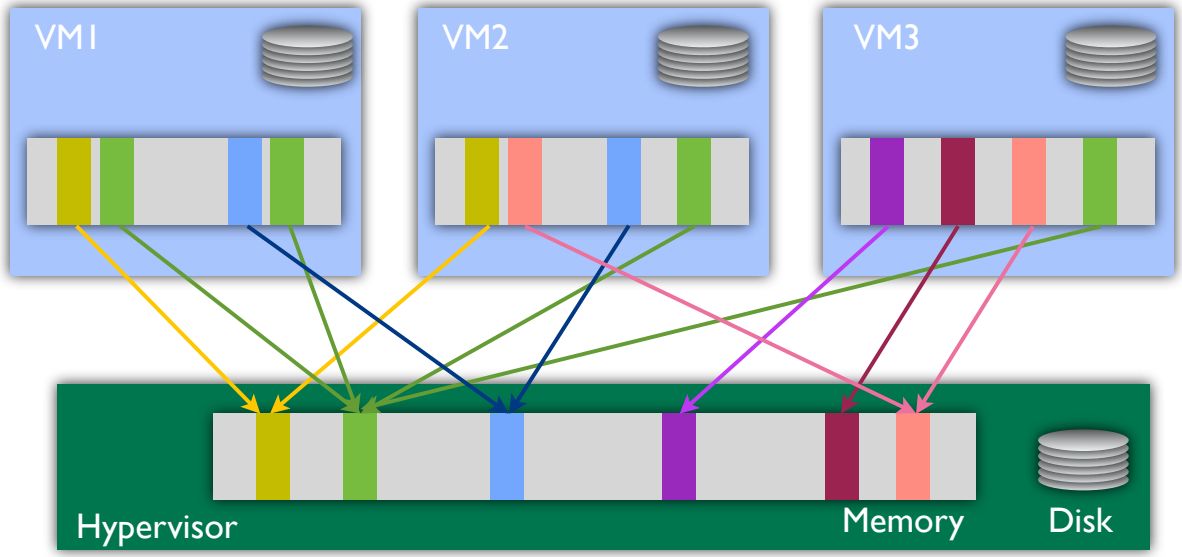
- **CPU are generaly used but :**
  - Memory is the most constrained computing resource in a virtualized data center
    (30% CPU, 80% RAM)
  - Can we use «like this» previous algorithms ?
  - Yes/No, memory overcommitment have specific management system
    Content Based sharing
    Ballooning
    Compressed memory
    Hypervisor swapping
- **These features can be used to defined a better VM placement ?**
  - Exemple: Content base sharing

Understanding Memory Resource Management

Memory overcommitment



Content-Based sharing

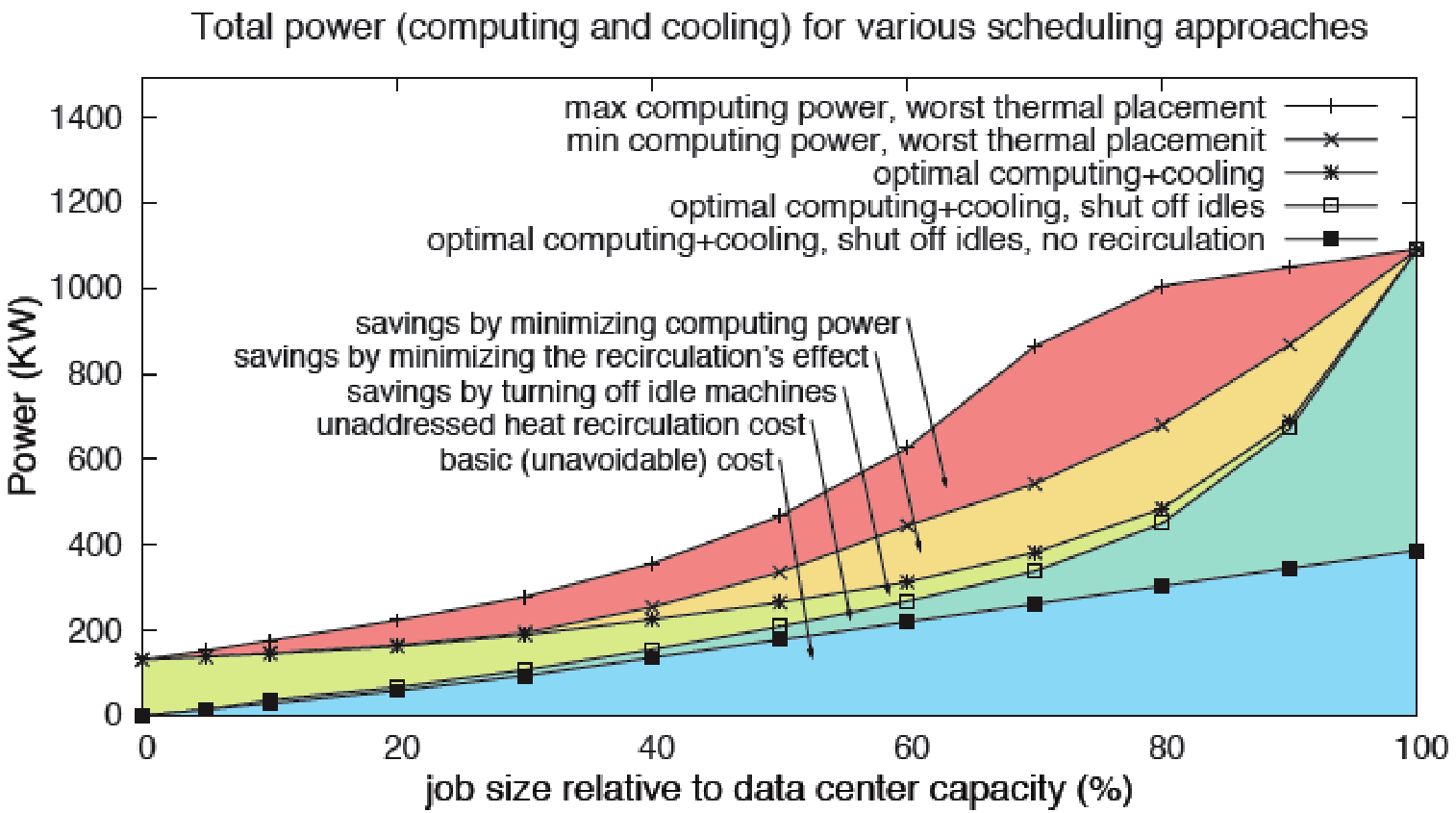


- **The concept of transparent page sharing was first proposed in the Disco system [1997-17]**

Content-Based sharing

- **Effective only if it is complemented by algorithms that ensure that the VMs resident on each physical server contain a significant amount of sharable pages.**
- **Memory Buddies [2010] Goals :**
  - Analyze the memory contents of multiple VMs to determine **sharing potential** then find **more compact VM placement**
  - Evaluation show that “sharing aware” placement has the ***potential*** to significantly improve memory usage (20 VM on 4 servers).
  - ***Invasive*** system (nucleus component into each virtual machine)
- **Sharing-Aware Algorithms for Virtual Machine Colocation [2012]**
  - ***simulation*** with (124 VM on 25 servers) and ***offline***
- ***CBS Challenge :***
  - Transparent Page Sharing with Large Pages, Effects of Memory Randomization, Sanitization and Page Cache on Memory Deduplication ...
- **Dynamic consolidation with resource sharing aware**

Holistic System ?



- **Thermal-Aware Job Scheduling to Minimize Energy Consumption in Virtualized Heterogeneous Data Centers [2009-18]**

ASU iMPACT Lab

Multi-resources

- **Generalization :**
  - Server resources

    CPU, RAM, Disk, Net, Energy
  - Rack Ressources

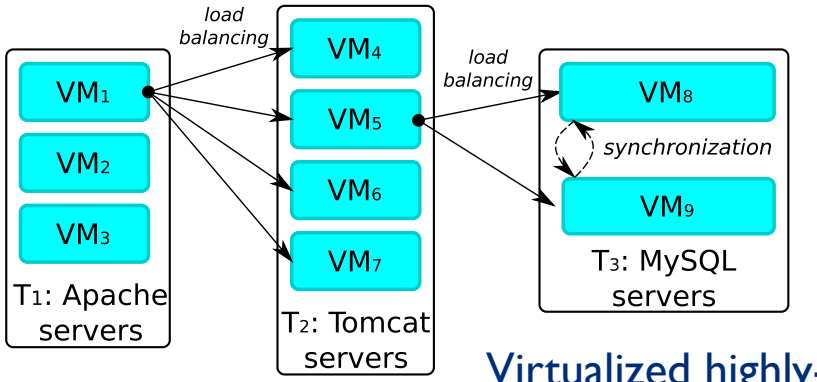
    Net, cooling, space
  - Data center resources

    Cooling, Humidity, Noise, Electrical, Phases, UPS, ...
- **How optimize virtualized datacenter with multiple inter-dependent objectives ?**
  - Ex: you can increase room temperature for reducing the cooling energy consumption, but a collateral effect should be done by a fan speedup (and increase all servers power consumptions).
  - How can express relation between cooling and server consumption

    Server consumption and noise etc.
- **Multi-resources dynamic consolidation**

Flexible for optimization but also to add new concerns

- **Why integrate new concerns?**
  - Fault tolerant, security, availability, energy aware, performance etc.
  - VM can be mutually inter-dependent



Virtualized highly-available Web application

- **These concerns cannot be exhaustively listed ...**
  - new concern emerge regularly depending on the applications' domain, computer science trends, or new technologies
  - VM manager should then support these evolutions as soon as possible

Flexible systems

- **Need for flexible and energy-aware framework for the (re)allocation of virtual machines in a data centre**
  - [2011-10], [2011-11], [2011-12], [2011-13] allow third party developers to implement their own placement constraints
- **[2012-14] propose a flexible and energy-aware framework for the (re)allocation of virtual machines in a data centre**
  - Extend ou previous work on Entropy and add 16 new SLA constraints

    Based on CP Programming
  - Evaluation on 7 servers, limited heterogeneity (2 types), poor performance.
- **Performant Flexible dynamic consolidation**

Conclusion

- **Pack with ressource sharing aware**
  - (DVFS, Core on/off, TurboBoost, CBS etc.)
- **Pack with a holistic view**
  - traditional + many others (cooling, noise, humidity, electrical)
- **Pack with differents concerns**
  - enregy, security, avaibility
- **And lot of other challenges**
  - From black box VM to grey box

    VMM black box unable to provide high-level application QoS guarantees ...
  - VM manager reactivity / scaling

    reactivity : time to compute the solution, time take by the reconfiguration
  - With continious ernergy system to variant (renewable ernegy)

    Transition from dynamic consolidation to scheduling system
  - ...

References

- [1997-01] E. G. Coffman, Jr., M. R. Garey, and D. S. Johnson. Approximation algorithms for bin packing: a survey, pages 46–93. PWS Publishing Co., Boston, MA, USA, 1997. 48, 51
- [2007-02] Bhuvan Urgaonkar, Arnold Rosenberg, and Prashant Shenoy. Application placement on a cluster of servers. Oct 2007. 46, 48, 51, 77, 97
- [2009-03] Bo Li, Jianxin Li, Jinpeng Huai, Tianyu Wo, Qin Li, and Liang Zhong. Enacloud: An energy-saving application live placement approach for cloud computing environ- ments. In CLOUD '09: Proceedings of the 2009 IEEE International Conference on Cloud Computing, pages 17–24, Washington, DC, USA, 2009. IEEE Computer Society. 48, 51
- [2012-04] Eugen Feller, Louis Rilling, and Christine Morin. "Snooze: A Scalable and Autonomic Virtual Machine Management Framework for Private Clouds". The 12th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing (CCGrid 2012), Canada, Ottawa, May 2012
- [1986-05] Alexander Schrijver. Theory of linear and integer programming. John Wiley & Sons, Inc., New York, NY, USA, 1986. 50
- [2009-06] Fabien Hermenier, Xavier Lorca, Jean-Marc Menaud, Gilles Muller, and Julia Lawall. Entropy: a consolidation manager for clusters. In Proceedings of the 2009 ACM SIG- PLAN/SIGOPS international conference on Virtual execution environments, VEE '09, pages 41–50, New York, NY, USA, 2009. ACM. 27, 29, 51, 57, 58
- [2011-07] http://www.v-index.com/
- [2011-10] E. Bin, O. Biran, O. Boni, E. Hadad, E. Kolodner, Y. Moatti, and D. Lorenz. Guaranteeing high availability goals for virtual machine placement. In 31th ICDCS, june 2011

References

- [2011-11] R. Harper, L. Tomek, O. Biran, and E. Hadad. A virtual resource placement service. In 2011 IEEE/IFIP 41st International Conference on Dependable Systems and Networks Workshops (DSN-W) , pages 158–163, june 2011
- [2011-12] F. Hermenier, S. Demassey, and X. Lorca. Bin Repacking Scheduling in Virtualized Datacenters. Principles and Practice of Constraint Programming–CP 2011 , pages 27–41, 2011.
- [2011-13] C. Liu, B. T. Loo, and Y. Mao. Declarative automated cloud resource orchestration. In Proceedings of the 2nd ACM Symposium on Cloud Computing ,SOCC '11, pages 26:1–26:8, New York, NY, USA, 2011. ACM.
- [2012-14] C. Dupont, T. Schulze, G. Giuliani, A. Somov, F. Hermenier. An Energy Aware Framework for Virtual Machine Placement in Cloud Federated Data Centres e-Energy '12
- [2006-15] Fabien Hermenier, Nicolas Loriant, and Jean-Marc Menaud. Power management in grid computing with xen. In Proceedingsof 2006 on XEN in HPC Cluster and Grid Computing Environments (XHPC06), number 4331 in Lecture Notes in Computer Science, pages 407-416, Sorento, Italy, December 2006.
- [2002-16] Carl A. Waldspurger. 2002. Memory resource management in VMware ESX server. SIGOPS Oper. Syst. Rev. 36, SI (December 2002), 181-194. DOI=10.1145/844128.844146 http://doi.acm.org/10.1145/844128.844146
- [1997-17] Edouard Bugnion, Scott Devine, and Mendel Rosenblum. DISCO: Running Commodity Operating Systems on Scalable Multiprocessors. In SOSP, pages 143–156, 1997
- [2009-18] Tridib Mukherjee, Ayan Banerjee, Georgios Varsamopoulos, S. K. S. Gupta, and Sanjay Rungta, Spatio-Temporal Thermal-Aware Thermal-Aware Job Scheduling to Minimize Energy Consumption in Virtualized Heterogeneous Data Centers. (Elsevier) Computer Networks, Special Issue on Virtualized Data Centers(ComNet), accepted (2009)