Energy: A new criteria for performances in large scale systems

Jean-Marc Pierson

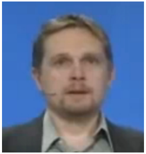
IRIT, Toulouse University

IRIT CNRS - INPT - UPS - UT1 - UTM

Université Paul Sabatier TOULOUSE III

cost www.cost804.org Action IC0804

From data dissemination to mobility prediction to energy


www.cost804.org
Action IC0804

.... Energy consumption vs. latency ... Avoidance and Energy Efficient Routing .... for WSN ....

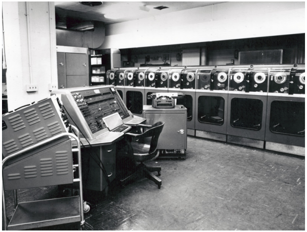
Congratulations Günter!!

Green IT

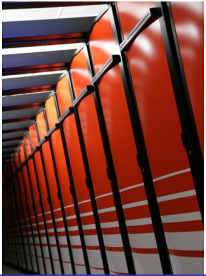
- ICT consumes a lot :
  - Estimated to 10% of global consumption of electricity (90 TW)
  - Carbon footprint equivalent to aeronautics
  - More than $30 billion in 2008 "worldwide cost"
  - Increased by 10 times over the past ten years

What is it all about?

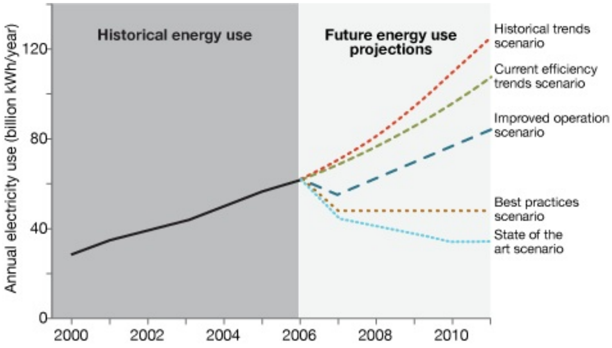
UNIVAC I (UNIVersal Automatic Computer) machine in the 1950's was consuming 125 kW for 1905 operations per second.



Jaguar: 1.7 Petaflops at a cost of 6950 kW.



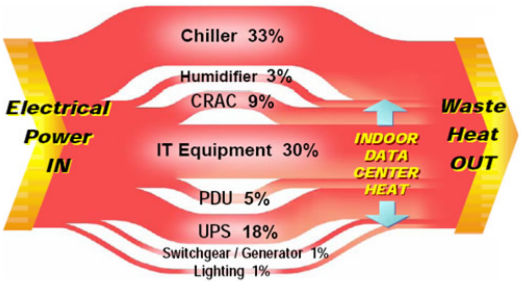
Energy trends for data centers



Source: The Report to Congress on Server and Data Center Energy Efficiency, 2007

Power loss in data centers

Figure 1 – Power flow in a typical data center



(c) APC

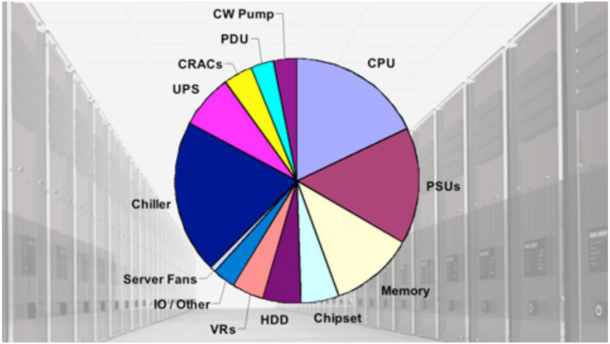
A first metric: PUE

Power Usage Effectiveness

- a standard developed by the Green Grid consortium
- power performance evaluation at the data center level
- aims at getting the effectiveness of the environment of the IT equipment : the optimal is 1. Today best is about 1.1, average is at 1.9.

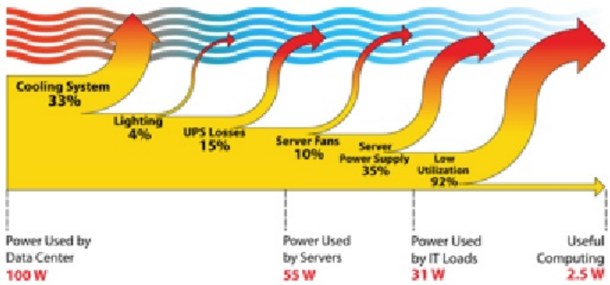
$$PUE = \frac{TotalFacilityPower}{ITEquipmentPower}$$

Power shares



(c) Podtech

Energy loss for computing



(c) Jeff Kandyba and RMI

Power is lost during idle periods: Switch off computers, consolidate? Slow down components?

But, actually, what is energy?

$$energy = power * time$$

expressed as a number of Watt hour, Watt second or Joules.

Compute energy?

- Determine the power used
- Determine the time to solution

But:

- Which power to take into account? Max, Average? Of which part of the infrastructure?
- Power may not be easy to measure. Power-meter as oscilloscopes and/or external power-meters?
- Need to distinguish between static and dynamic power.
- Infinite jobs don't care about TTS.

Benchmarks

- Green500: for HPC. computes *Flops/watt* using LinPack public Top500 lists.

  Example: Green500's best : 773.38 MFlops/Watt.
- SpecPower: For java servers. computes server-side-java operations/*watt*, including the idle time on specific workloads. Example: 2927 ssj ops/watt for a 304 cores Fujitsu machine (best among 172 tested servers)
- TPC-Energy: for transactional applications. It computes *watts/operations*
  Example: 5.84 watts/transaction per seconds for an typical online transaction processing workload on HP servers. (only few tested)
- + vendor related benchmarks...

Integrating energy concerns - 1/2

- Hardware level: a long time believed to be the only way for reducing energy.
  - Energy efficient hardware, Faster hardware reduce both Power and Time, thus Energy.
  - P-C states in CPU: the power can drop to 2 watts when idle. (automatic and by software).
  - Cores switched off, disk spin off, network cards sleep modes, ...
- Network level
  - Little influence of the traffic on the energy consumption of core network equipments
  - Works on the most energy efficient routes, router configuration adaptation

Integrating energy concerns - 2/2

- Middleware level: the mammoth of the research works
  - Adapt the hardware dynamically as a function of the infrastructure load: Dynamic Voltage Frequency Scaling. remember that $power = voltage^2 \times frequency \times alpha$.
  - Server consolidation, mainly using energy-aware jobs (virtual machines) placement, migration, scheduling: switch off hosts, find most efficient hosts for a given task.
  - Many heuristic based works, few theoretical / analytical models
- Software level
  - Mathematical modeling of consumption of applications by observing behavior (performance counters, electrical demands, communications, ...)
  - SLA description and enforcement
  - Dynamic software adaptation with the actual load of the system

Our approach: Yield per Watt

Task allocation in virtual clusters

- Allows to allocate fractions of resources
- Allows to migrate jobs without loss of connectivity

Energy-Aware Resource Allocation. Damien Borgetto, Georges Da Costa, Jean-Marc Pierson, Amal Sayah. E2GC2 Workshop, IEEE Grid'2009

Energy-Aware Service Allocation. Damien Borgetto, Henri Casanova, Georges Da Costa, Jean-Marc Pierson. submitted to Elsevier FGCS

Modeling as a MILP problem

$$\forall i,h \qquad e_{ih} \in \{0,1\} \qquad (1)$$
$$\forall i,h \qquad \alpha_{ih} \in Q \qquad (2)$$
$$\forall h \qquad p_h \in \{0,1\} \qquad (3)$$
$$\forall i \qquad \sum_h e_{ih} = 1 \qquad (4)$$
$$\forall i,h \qquad 0 \leq \alpha_{ih} \leq e_{ih} \qquad (5)$$
$$\forall i,h \qquad p_h \geq \alpha_{ih} \qquad (6)$$
$$\forall h \qquad p_h \leq \sum_i e_{ih} \qquad (7)$$
$$\forall h \qquad \sum_i \alpha_{ih} \leq p_h \qquad (8)$$
$$\forall h \qquad \sum_i e_{ih} m_i \leq p_h \qquad (9)$$
$$\forall i \qquad \sum_h \alpha_{ih} \leq \alpha_i \qquad (10)$$

Solving the problems

- Optimal solution for small instances of the problem (computation costs)
- Bound on optimal for large instances, using relaxed problem
- Propose heuristics and compare them: Greedy-like and Binpacking-like (EA-ResAlloc)
- Problems for Energy bound, Yield bound or tradeoff between both

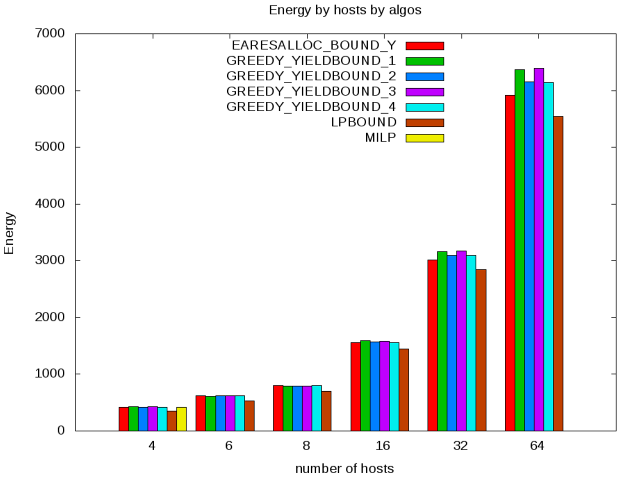
EA-ResAlloc: Placement and Consolidation

1. Placement: Favors an energy efficient host.
2. Consolidation: Favors an already loaded host.
3. Tradeoff between placement and consolidation (λ)
4. Tradeoff between energy savings and performance (k)

An heuristic based on a cost function to maximize. This cost function represents the yield of a job:

$$\frac{Y^{(1-k)}}{E^k} = \frac{\left[\sum_{j=1}^{H}\left(\frac{\alpha_{ij}}{\alpha_i}\right)\right]^{1-k}}{\left[\lambda(C_j^{max} - C_j^{min}) \times \alpha_{ij} + (1-\lambda)\left[A_j(1 - \sum_{i'=1,i'\neq i}^{J}(\alpha_{i'j}))\right]\right]^k}$$

Results with a bound on Yield



Lessons learned

- Simplified modeling: no migration, no job dependencies / communications, simplistic model for interdependence of job allocation
- Needs to improve the model with real monitored data (for consumption of jobs, of nodes, ...)
- Difficult to model the tradeoff between energy and performance: Which performance? Fairness? Utilization? Time To Solution? ...

An ongoing work

- Metrics, benchmarks, best practices, are application dependent?
- To Include ecological concerns and not only electrical concerns?
- more info at: http://www.irit.fr/˜Jean-Marc.Pierson or pierson AT irit.fr

cost
www.cost804.org
Action IC0804

Next meeting: November 8-9th in Coimbra, Portugal.
You are welcome !!