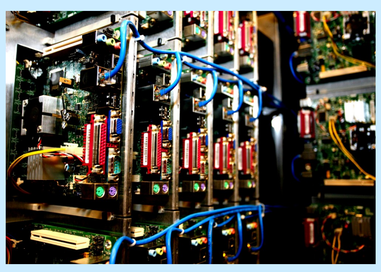
A Case for Energy-Aware Accounting in Large-Scale Computing Facilities

Víctor Jiménez, Francisco J. Cazorla, Roberto Gioiosa, Eren Kursun, Canturk Isci, Alper Buyuktosunoglu, Pradip Bose, Mateo Valero
{victor.javier,francisco.cazorla,roberto.gioiosa,mateo.valero}@bsc.es
{ekursun,canturk,alperb,pbose}@us.ibm.com

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
UPC
IBM
CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

The Energy Wall

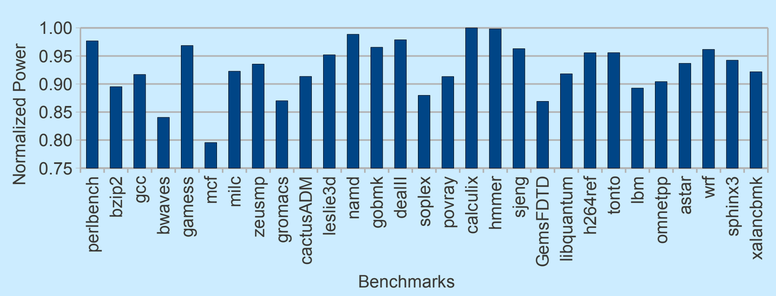




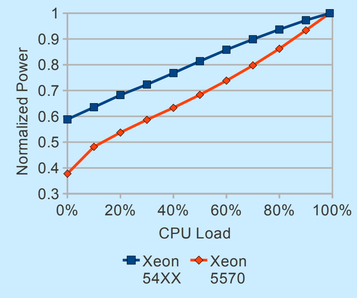
- Large-scale computing facilities (LSCF)
  - Newest facilities consume **up to 20MW**
  - Expensive (up to $30 billion in US)
  - Pollution
- Energy is becoming the most expensive resource
  - It is already up to **40% of the total cost of ownership** (TCO)
  - Cost of HW remains similar while energy price increases

Motivation

- Most of the current LSCF account are based on:
  - Resource size (e.g., number of nodes) and usage time
  - The cost of energy is evenly distributed among users
    - *Based on peak (nameplate) power consumption values*
- However, this does not consider resource utilization...
  - ... and **energy consumption is affected by utilization**!



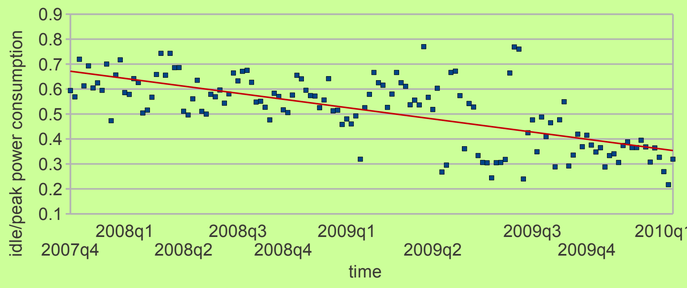
Up to 20% difference among SPEC CPU2006



Up to 40% difference for a server with 50% load

Energy-Proportional Systems

- Energy consumption breakdown
  - **Static**: consumption when the system is idle (e.g., C-states)
  - **Dynamic**: consumption due to activity on the system
- The trend is to **reduce the static part**
  - Towards energy proportional systems



Evolution of idle/peak power consumption for SPECpower submitted results

- Ideally we can reach zero static power consumption
  - Strong motivation for energy-aware accounting

Energy-Aware Accounting

- Energy-aware accounting
  - **Fine-grain tracking of energy consumption** in LSCFs
  - It will allow to accurately track **per-user energy usage**
- General benefits
  - Drive up energy-efficiency in computing facilities
  - Increase energy-awareness within end-user community
  - Strengthen the trend towards energy-proportional systems
  - Ultimately, allow for **greener LSCF**
    - *Without hurting LSCF owner's bottom-line profit margins*
- Technological benefits for LSCF
  - Easier adaption of **adaptive systems**
  - More accurate runtime task and/or cooling resource allocation
  - Safer workload consolidation

Trade-offs

- Granularity vs. Overhead
  - Level at which energy is tracked (node/user/task)
- Fairness
  - Isolate interference of co-scheduled tasks
  - Multiple executions (with the same input) should be ideally accounted the same
- Power vs. Energy
  - Less execution time implies more power
    - *Reduces static consumption significance*
  - More time may help to avoid power peaks
- Accuracy vs. Variation
  - Cooling variation depending on location
  - Variation across server generations

Environments

- **Dedicated** nodes
  - HPC clusters
  - Per-node accounting required
- **Shared** nodes
  - Hardware resources shared via virtualization
  - Intra-node energy accounting required
  - Need to differentiate between static/dynamic energy consumption

Static Power Accounting

- Depending on the component type
  - **Spatial**-sharing (e.g., hard drive)
  - **Temporal**-sharing (e.g., CPU)

M_i: *space used by user i*

$$\sum_{i=1}^{N} M_i = M_{total}$$

S_i: *static consumption incurred by user i*

$$\sum_{i=1}^{N} S_i = S_{total}$$

$$S_i : (M_i / M_{total}) \cdot S_{total}$$

Spatial-sharing

N_k: *number of applications running during interval k*

$$S_{i,k} = S_k / N_k$$

$$S_i = \sum_{k=1}^{N} S_{i,k}$$

Temporal-sharing

Dynamic Power Accounting

- Depending on the workload type
- Request-based workloads
  - **High-level metrics**
    - *CPU utilization*
    - *Requests per unit of time*
  - Reduced complexity and overhead
- CPU-intensive workloads
  - CPU utilization is always close to 100%
  - **Event-based metrics**
    - *Performance counters*
    - *OS statistics*
- Other approaches
  - Instruction mix analysis
  - Program features analysis
- HW/SW support can improve the accuracy

Open Issues

- How to attribute extra energy due to...:
  - Application interference in shared hardware resources
  - Energy consumption due to OS or hypervisor
- Account for VM resource optimizations

Conclusions

- We make the case for Energy Accounting
  - HW/SW solutions to provide accurate energy accounting per task
  - More important as systems become energy-proportional
- Fertile area of research
- The outcome can lead to a greener world

This work was supported by a Collaboration Agreement between IBM and BSC with funds from IBM Research and IBM Deep Computing organizations. It has also been supported by the Ministry of Science and Technology of Spain under contract TIN-2007-60625 and grants AP-2005-3776 and AP-2005-3318, and by the HiPEAC Network of Excellence (IST-004408).