Impact of Shutdown Techniques for Energy-Efficient Data Centers

Issam RAIS

Univ. Lyon, Inria, CNRS, ENS de Lyon, UCBL 1, LIP, France.

Green Days, 25 November 2016

An energy driven world ...

- Computing datacenters are not negligible
- DataCenters responsible of 2% of global carbon emissions
- 80 million megawatt-hours of energy annually
- Almost 1.5 the amount of electricity used by New York City

... needs a "greener" usage of infrastructures

- Free cooling
- Low-power processors
- Recover lost energy (ex : reuse water-cooling)
- ShutDown techniques

Shutdown techniques are :

- Extensively studied in literature
- Not used by computing centers administrators

Reasons

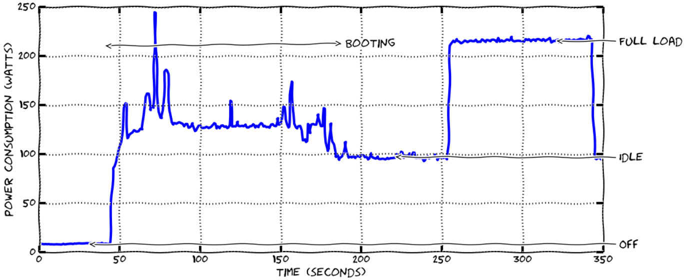
- servers were not designed to be switched off
- difficult to estimate gains versus potential loss

What leverage does it take into account ?

- Available sleep states

Why are we interested in this leverage ?

- Non proportional computing units
- And over provisioning of infrastructures
- Lead to non negligible energy consumption when idle
- One of the most promising leverage



Disadvantage

Turn On or Off is not energy and time free !

In this paper

Contributions :

- Study of different shutdown techniques for computing resources
- Estimate the impact of such techniques on : energy consumption, reactivity of the platform, lifetime of the servers

Validated with :

- Real calibrations
- Real datacenter and supercomputer traces
- Replay of traces

Hardware ability [1]

- Linux specification
- Available states : "Suspend to Idle" to "System shutdown state"
- Used Datacenter : "Suspend to Idle" and "System shutdown state"

Energy aware algorithms

- Pre-determined policies [2]
- Or workload prediction [Orgerie and Lefèvre, 2011]

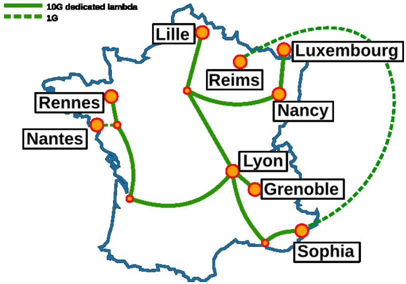
1. http ://www.acpi.info/
2. http ://slurm.schedmd.com/power_save.html

Time Threshold : T_s

- Minimal time for the server for shut down to be energy efficient [Orgerie et al., 2008]
- Takes into account cost in time and energy

ShutDown Policies

- P1 : knowing the future
- P2 : aggressive shutdown
- Ideal : No cost for switching current state



Grid'5000

- Large-scale and versatile testbed
- Experiment-driven research in all areas of computer science
- High heterogeneity in 8 different sites
- 6 years of resource utilization

2 types of traces

E-biothon

- Experimental Cloud platform
- Used by researchers in biology, health and environment
- 4 Blue Gene/P racks
- 4096 4-cores nodes, peak : 56 TeraFlops
- 15 months trace

Switching state calibration :

- Done on Grid'5000 hardware
- External fine grain monitoring
- 3600 measurements per second, 0.125 W precision
- Several nodes, average on 100 measurements

Parameters	Orion	Taurus	Paravance
E_{OffOn} (Joules)	23,386	19,000	19,893
E_{OnOff} (Joules)	2,300	2,000	2,000
T_{OffOn} (seconds)	150	150	167.5
T_{OnOff} (seconds)	10	10	7.5
P_{idle} (Watts)	135	95	150
P_{off} (Watts)	18.5	8.5	4.5
T_s (seconds)	195	227	172

Metrics

On servers consumption

- Comparison with P_{idle} consumption : no shutdown policy applied
- Ideal : No cost for switching state

On servers lifetime

- Maximal ShutDown imposed to server
- Number of On/Off cycles per policy

Simulation

With Described :

- Traces
- Calibration
- Shutdown Policies

Simulation hypotheses

- Homogeneous datacenter
- Replay for every traces
- Replay for every calibrated hardware

	% Energy saved on idle periods		
Calibration	P1	P2	Ideal
Grid'5000 trace, 6 years, 149 nodes on average			
Orion	85.87%	85.59%	86.29%
Taurus	90.56%	90.22%	91.05%
Paravance	96.66%	96.46%	97.00%
E-Biothon trace, 15 months, 4096 nodes			
Orion	85.18%	84.56%	86.29%
Taurus	89.83%	89.07%	91.05%
Paravance	96.03%	95.61%	97.00%

Energy conclusions

- Save important amounts of energy
- *P*1 and *P*2 are very close to *Ideal*
- Workload prediction is not worth the few energy it can save.

	% Shutdown cycles	
Calibration	P1	P2
Grid'5000 trace, 6 years, 149 nodes on average		
Orion	3,080	5,690
Taurus	2,980	5,690
Paravance	3,333	5,690
E-Biothon trace, 15 months, 4096 nodes		
Orion	33	70
Taurus	33	70
Paravance	38	70

Lifetime conclusions

- Up to 2.5 shudown per node for one day
- Far less than the number of start/stop cycles typically allowed [Seagate, 2012, Seagate, 2015]

Conclusion

With :

- Real Calibration on various hardware
- Real consequent traces from various datacenters and supercomputers
- Replay of traces with simulator
- Various shutdown policies

We showed that

- Prediction is not mandatory for ShutDown
- The old adage of shutdown doing too many on/off cycles is tackled here
- Show that Shutdown is a very interesting energy technique saving

Future work

- Apply such logic on close to proportional architecture (ex : BML)
- Integration of failure models when resuming from Off state to study the impact of bad resuming behaviour
- Studying the behaviour of switching nodes to control the impact on cooling system

Orgerie, A.-C. and Lefèvre, L. (2011).
ERIDIS : Energy-Efficient Reservation Infrastructure for Large-scale Distributed Systems.
Parallel Processing Letters, 21(02) :133–154.

Orgerie, A.-C., Lefèvre, L., and Gelas, J.-P. (2008).
Save Watts in Your Grid : Green Strategies for Energy-Aware Framework in Large Scale Distributed Systems.
In *IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, pages 171–178.

Seagate (2012).
Desktop HDD specification sheet.
http://www.seagate.com/staticfiles/docs/pdf/datasheet/disc/desktop-hdd-data-sheet-ds1770-1-1212us.pdf.

Seagate (2015).
NAS HDD specification sheet.
http://www.seagate.com/www-content/product-content/nas-fam/nas-hdd/_shared/docs/nas-hdd-8tb-ds1789-5-1510DS1789-5-1510US-en_US.pdf.