Optimal Sleeping

EDGE Lab OpenHouse
04/28/2011

Ioannis Kamitsos, Sangtae Ha, Jasika Bawa

Princeton University SCENIC
Social and Clean Edge Network via Inter-disciplinary Collaboration

The Princeton EDGE Lab

Acknowledgements

- Lachlan Andrew (Swinburne, Australia)
- Hongseok Kim (Bell Labs)
- Paschalis Tsiaflakis (KU Leuven, Belgium)

Motivation

- Reduction of energy consumption of computing infrastructure is motivated by several reasons.
  - Economic
  - Environmental
  - Concern of overheating
- Greening ICT sector will be beneficial to other sectors of the economy as well.

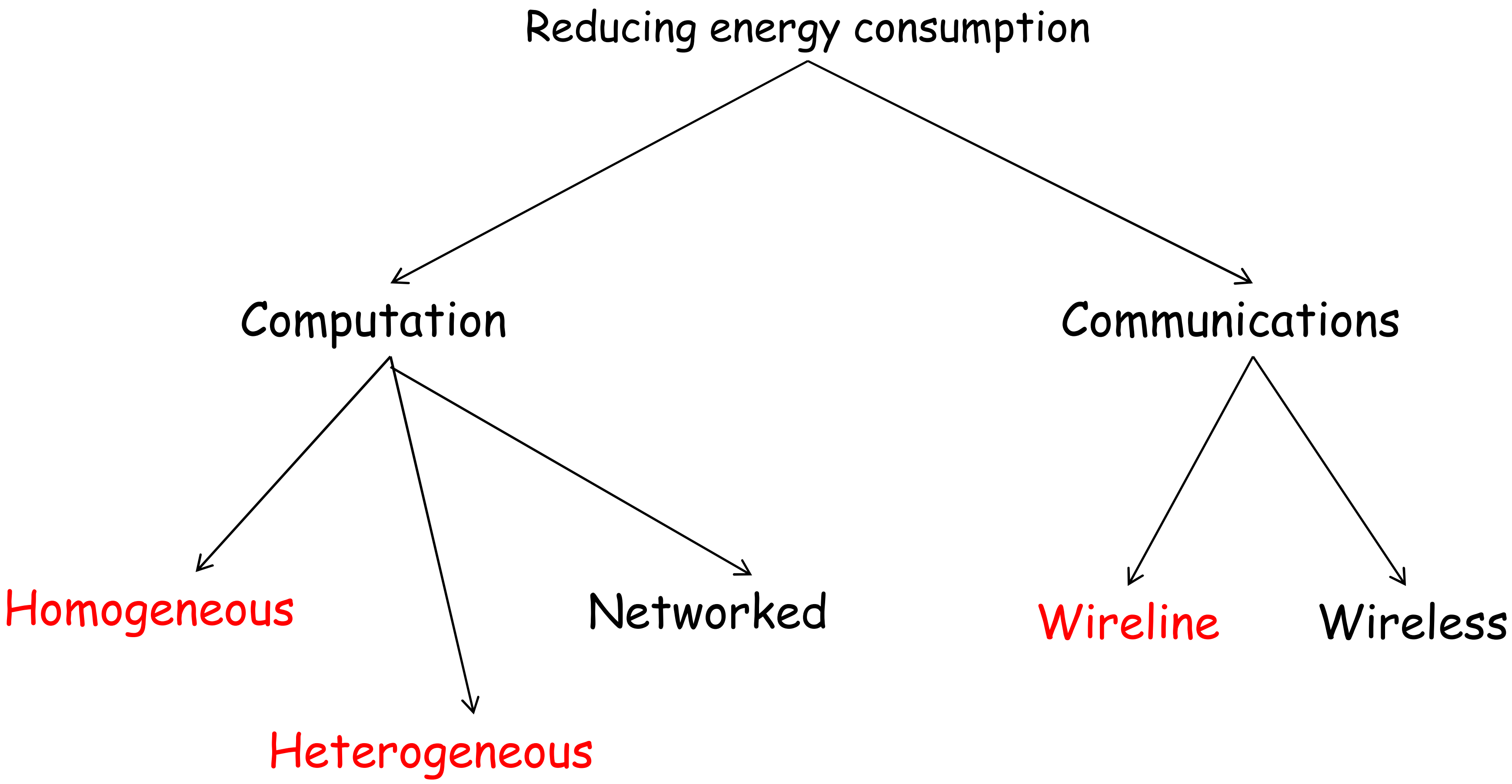
Problem

We seek to find the Pareto optimal tradeoffs between energy consumption and performance when sleeping mode operation is enabled on computing/communication infrastructure.

→ Saving energy is not trivial, since it takes energy and time to enter or leave sleep mode.

→ "Greening" price to pay: Delay/Stability.

Problem taxonomy



Sleeping in Data Centers

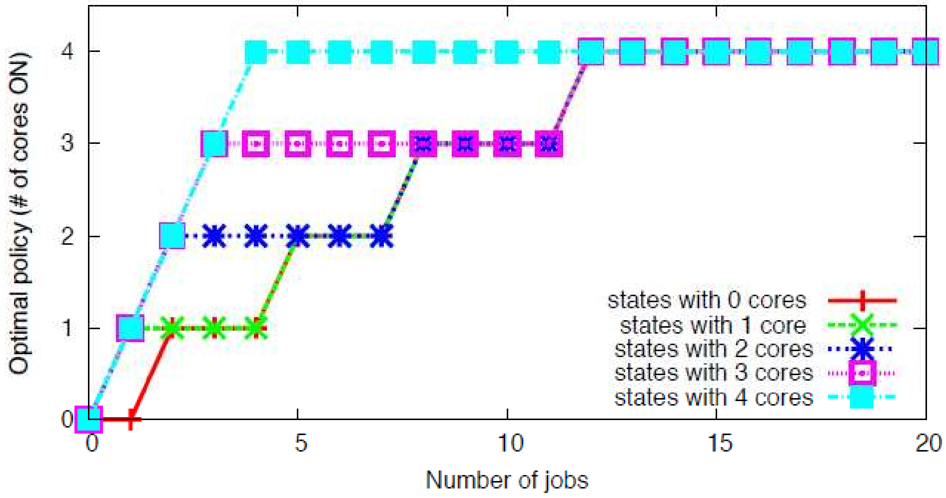
- Cloud computing has become significantly important.

- Average data center utilization 20-30%
  - Significant amount of energy wasted on idle mode.

- Designing an optimal sleeping policy is non trivial.
  - Best energy-performance tradeoff point depends on relative importance between energy savings and fast job processing.

  - Entering and leaving sleeping mode requires energy and time.

Optimal sleeping policy

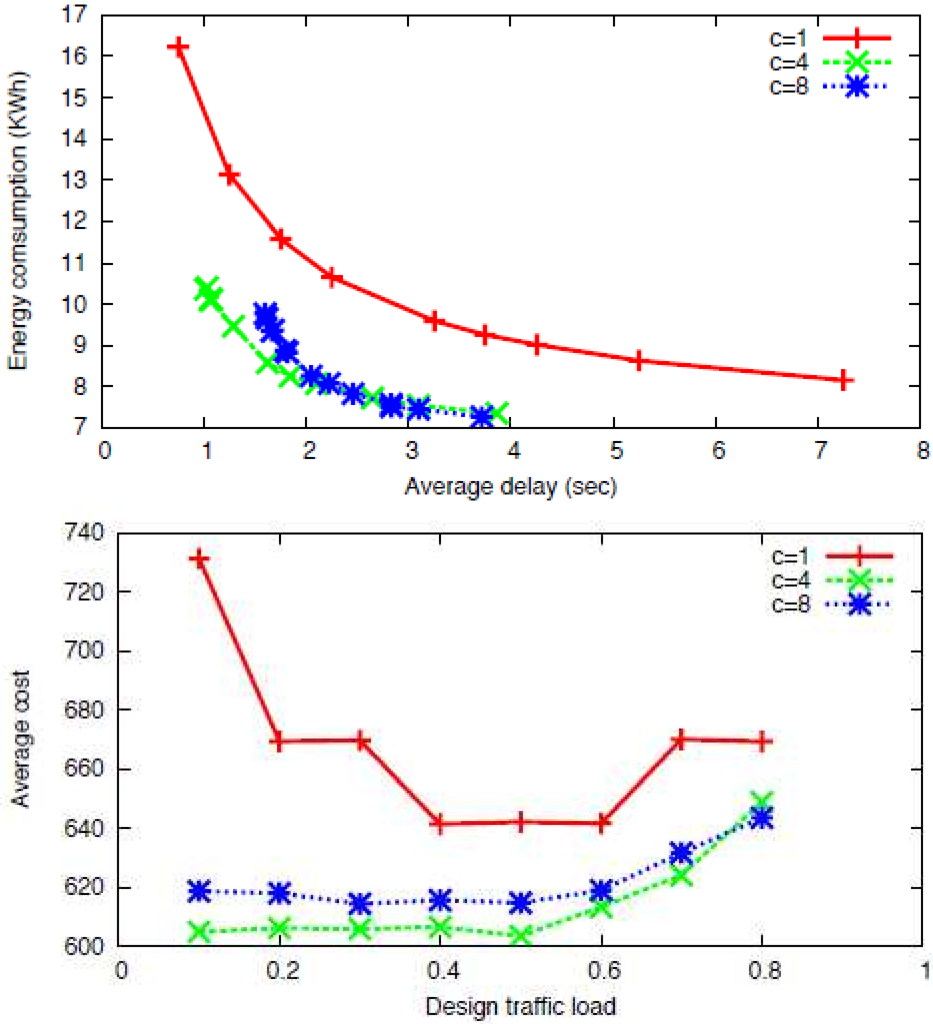


Monotone hysteretic policy

⇩

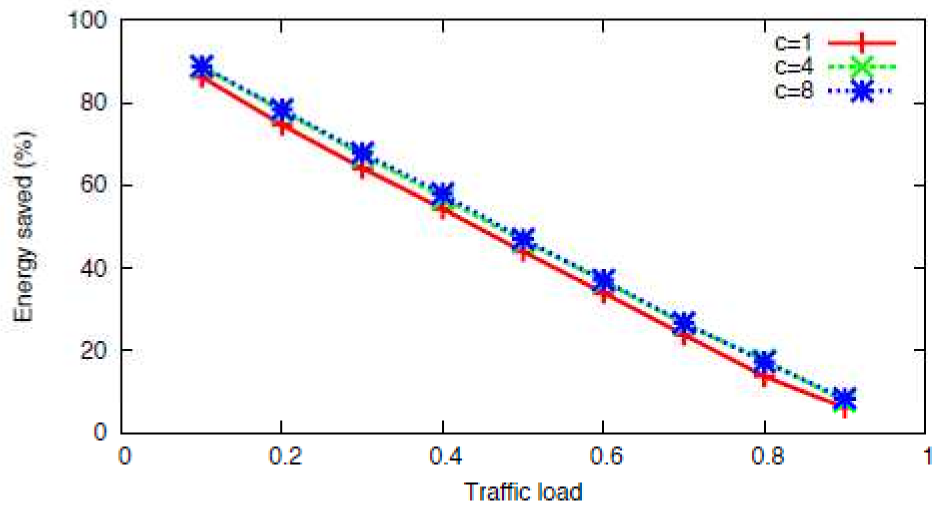
Implemented in EDGE lab experimental testbed

Resource pooling in homogeneous servers



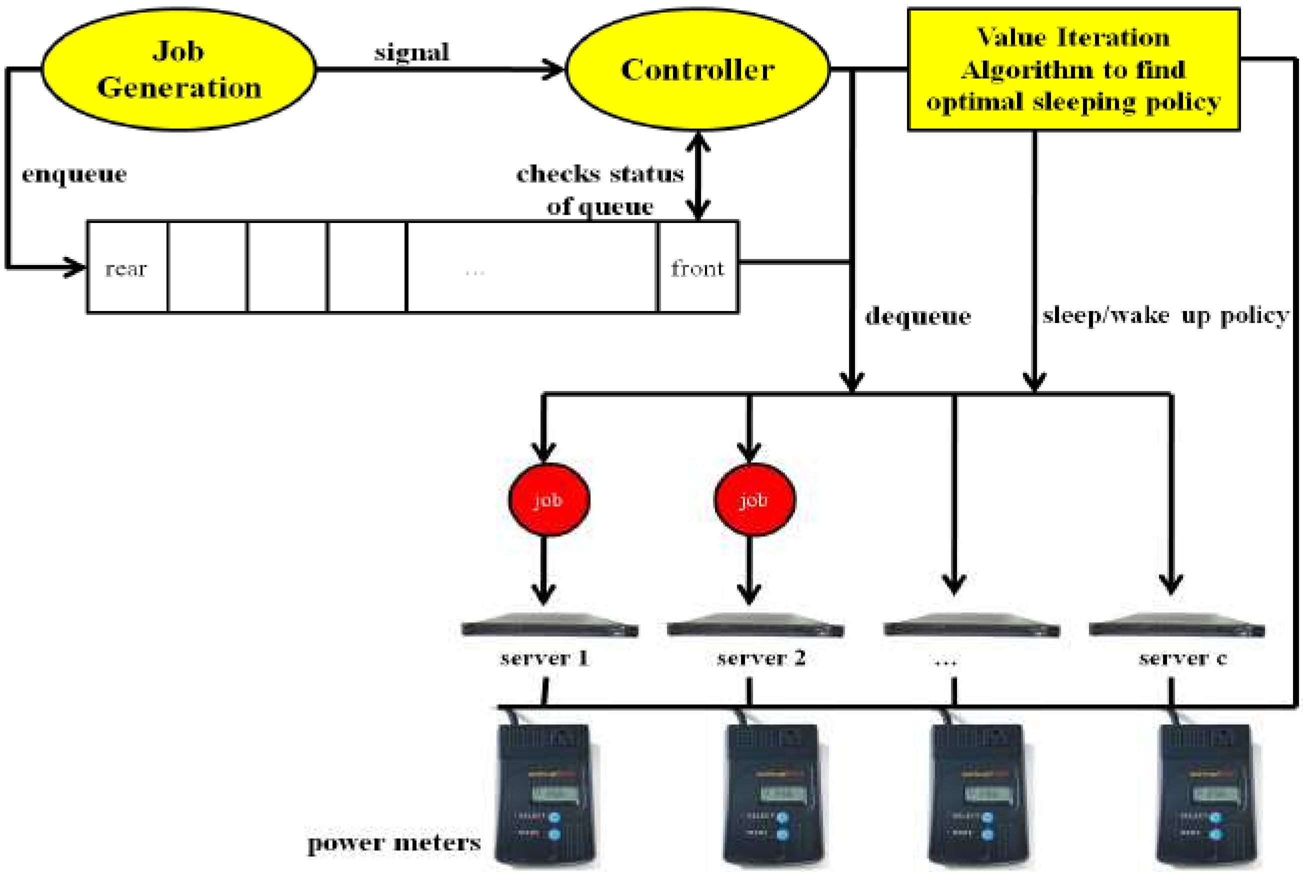
Resource pooling improves both system performance and system robustness against traffic load misestimating.

Energy savings

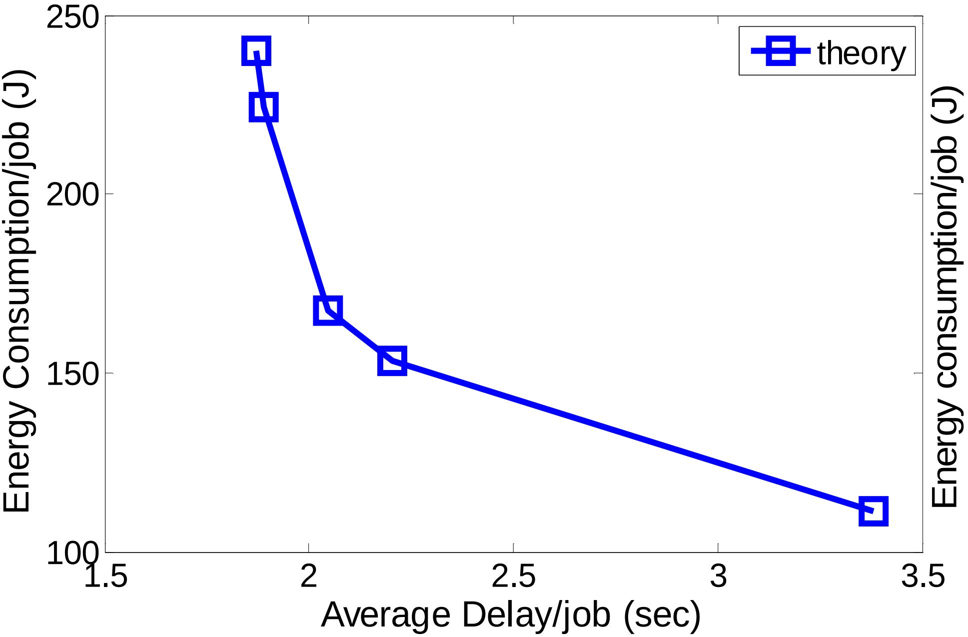


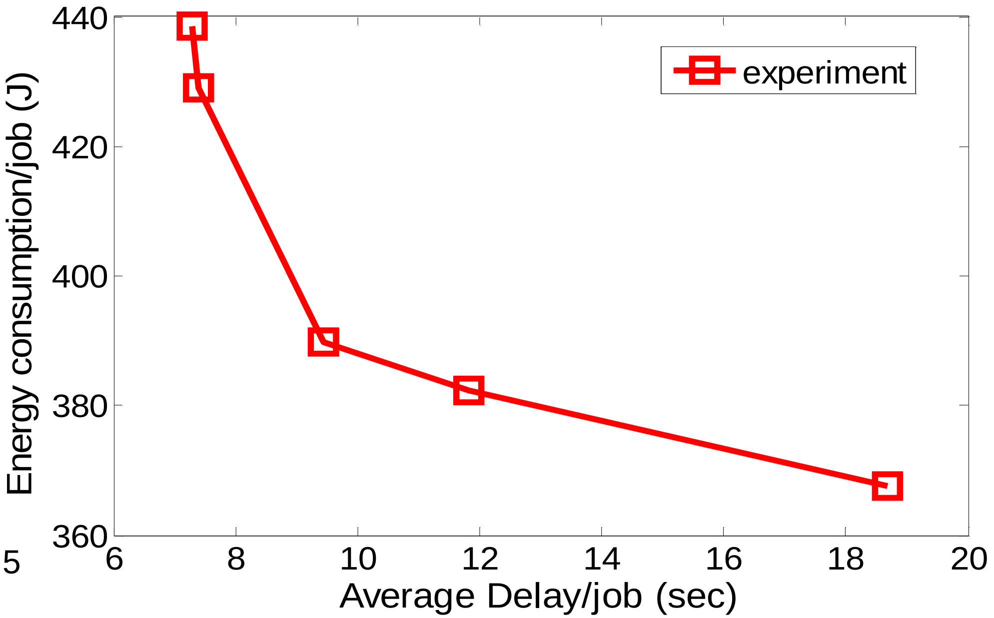
Sleeping results in significant energy savings: Almost 90% for traffic load of 10%.

Experimental testbed



Theory and Practice





Some of the practical issues during implementation:

- Multiple jobs being assigned to the same server.
- Sleeping commands issued at the middle of other commands.
- Waking up/Sleeping sometimes was taking more time than expected.
- Switching cost spikes.

Heterogeneous servers

- Large scale data centers consist of hundreds of heterogeneous servers.

- Suppose we have several different types of processors with multiple processors of the same type.

- Two levels of sleeping decisions

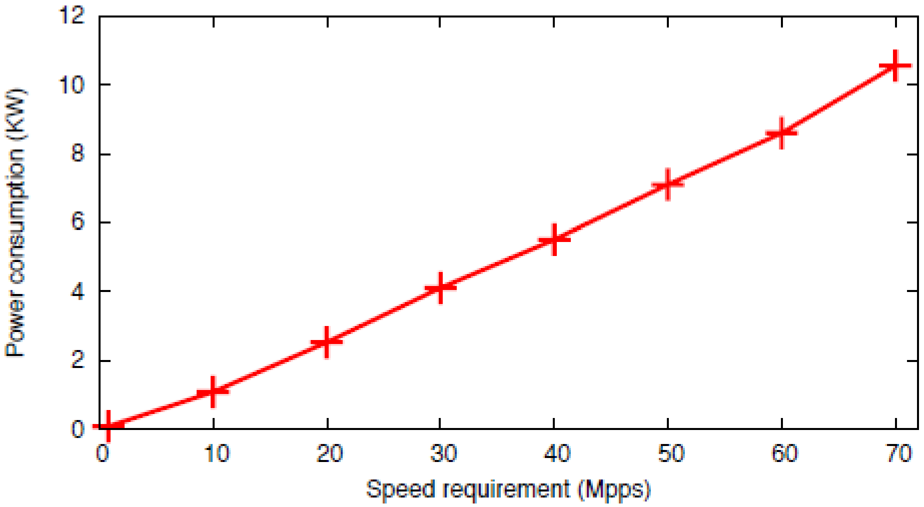
  - Slow timescale: Every time the speed requirements change
  - Fast timescale: Every time there is an arrival/departure

Problem formulation

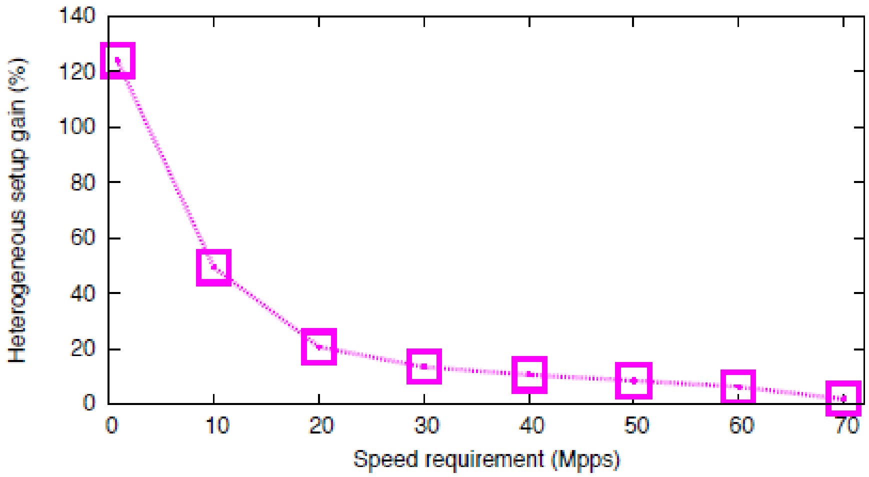
$$\begin{aligned} \min \quad & \sum_{j=1}^{N} P_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^{N} s_j x_j \geq s_0 \\ & 0 \leq x_j \leq b_j, \\ & x_j \text{ integer } , j = 1, 2, \ldots, N \end{aligned}$$

- Solve via greedy algorithm
- Performance within 6% of optimal solution

System performance



Convex power model applicable even when speed scaling is achieved by turning on/off cores/servers



Significantly less power consumption compared to "average" homogeneous setup

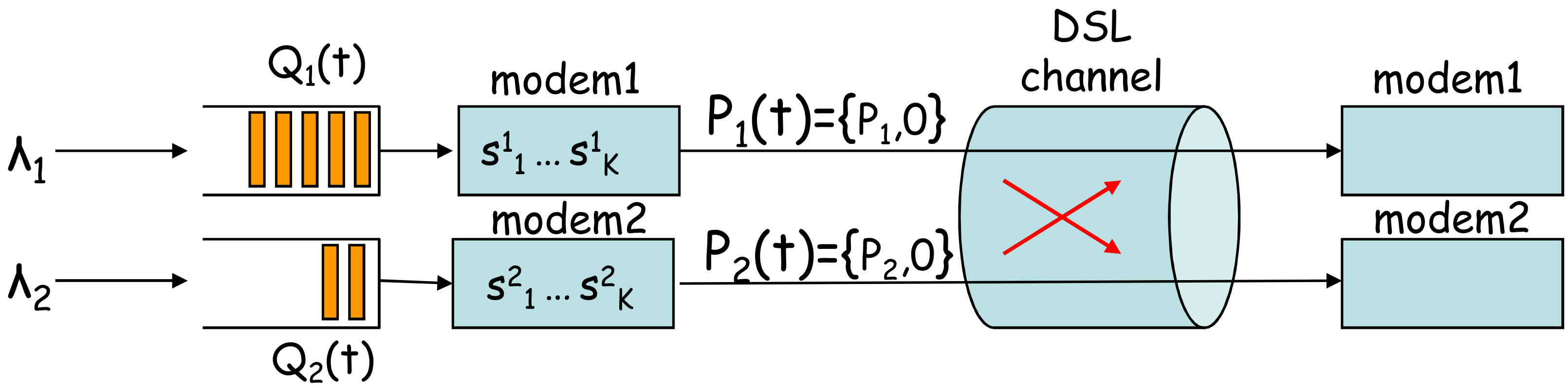
Another application

- DSL is the dominant wireline broadband access.

- L3 sleeping mode already there, however not yet exploited due to stability concerns.
  - Interference limited

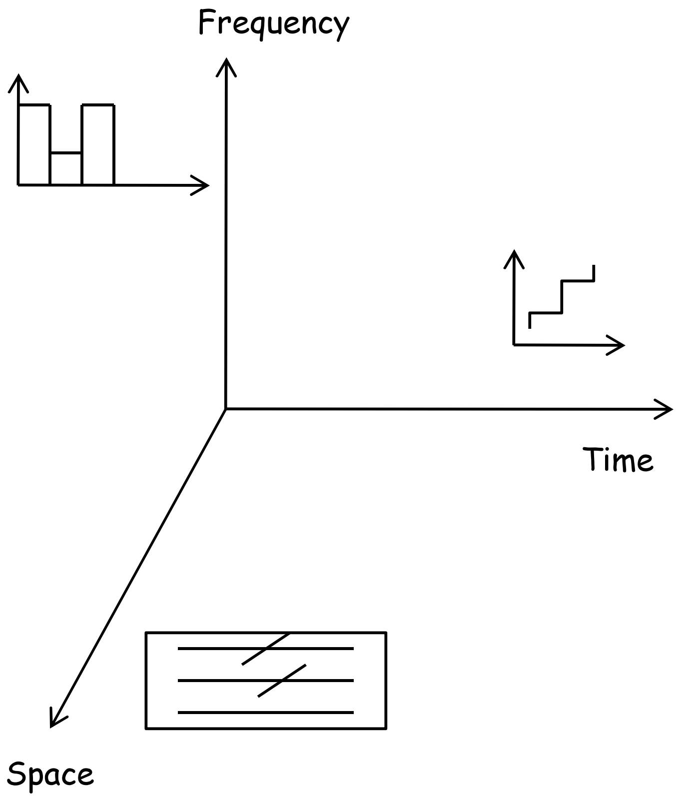
- Application of sleeping theory on DSL while improving stability at the same time.

Sleeping in DSL broadband access



- DSL infrastructure mostly underutilized → sleeping
- Goal: Optimize on-off transmit powers $P_1(t)$ and $P_2(t)$ over time so as to obtain a pareto-optimal energy-delay trade-off for given input arrival statistics

Stability improvement

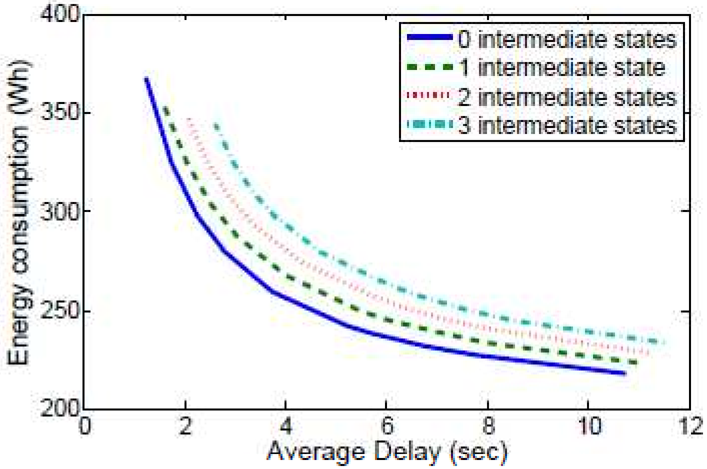


Time: Gradually wake up the modem.

Space: Limit the number of modems that can wake up simulataneously.

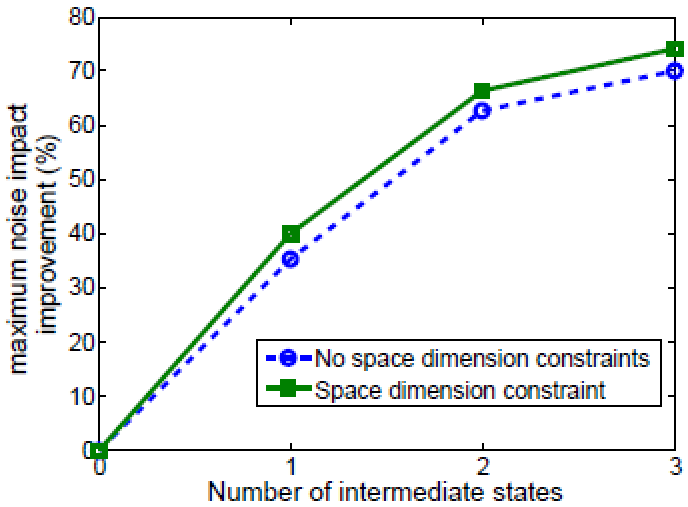
Frequency: Wake up subset of tones.

Energy vs Delay vs Stability tradeoff



Limiting the number of lines that can switch between OFF/ON improves stability.

Gradual wake up via intermediate states between OFF/ON increases delay but improves stability.



Conclusions

- Sleeping results in significant energy savings but requires paying in delay.
- Optimal sleeping policy implemented on experimental testbed.
- Resource pooling improves performance and robustness.
- Heterogeneity further sweetens energy – delay tradeoff.
- Stable sleeping in DSL broadband access is feasible and results in significant energy savings.

Q/A

Contact: Ioannis Kamitsos

kamitsos@princeton.edu