Energy-Efficient Framework for Networks of Large-Scale Distributed Systems

Anne-Cécile Orgerie and Laurent Lefèvre
ENS de Lyon – INRIA RESO – Université de Lyon – LIP

annececile.orgerie@ens-lyon.fr
laurent.lefevre@inria.fr

May 2011, ISPA 2011, Busan - Korea

Why do we need to be Green?

- "Transmitting data through Internet takes more energy (in bits per Joule) than transmitting data through **wireless** networks."
  Gupta & Singh – *Greening of the Internet* – SIGCOMM 2003
- "By 2015, **routers** will consume 9% of Japan's electricity."
  Michiharu Nakamura (Hitachi) - Nature Photonics Technology Conference 2007

Plan

1. Background

2. HERMES

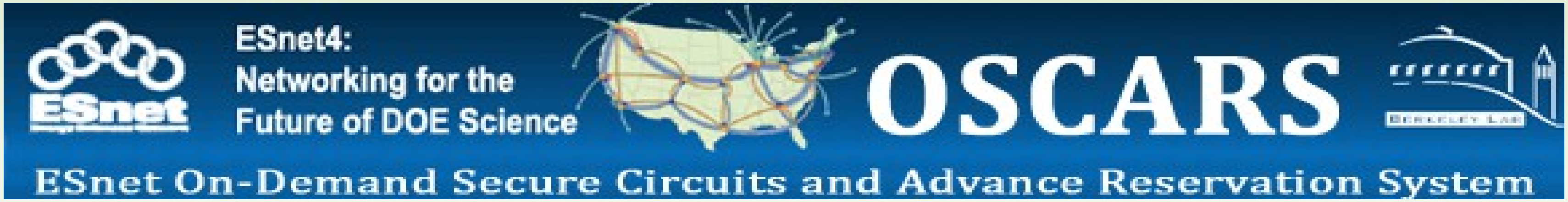
3. Validation

4. Conclusion and Perspectives

Background

Bulk Data Transfers with Advance Reservations in Large-Scale Distributed System Networks

- **BDT (Bulk Data Transfers)** → large volumes of data to transfer, moldable/malleable, deadline
- **ABR (Advance Bandwidth Reservations)** → bandwidth provisioned for the transfer (no resource competition, no congestion)
- **Large-Scale Distributed Systems Networks** → data center, grid, cloud networks



Why dedicated networks are relevant

In **2007**, to distribute the entire collection of **Hubble telescope data** (about 120 Terabytes) to various research institutions, scientists chose to copy these data on hard disks and to send these hard disks via **mail**.
It was faster than using the network.

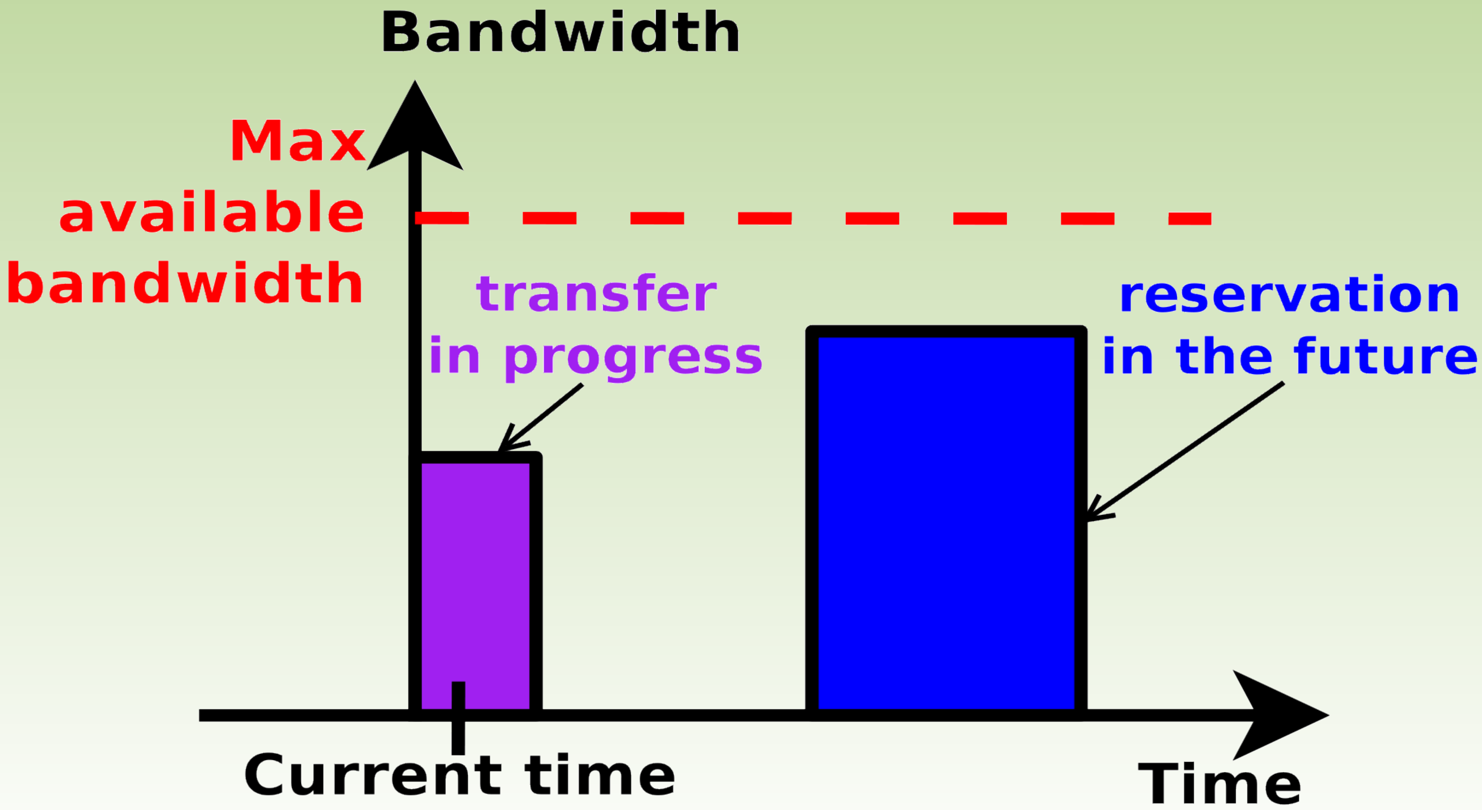
Cyrus Farivar. Google's Next-Gen of Sneakernet. [online] http://www.wired.com/science/discoveries/news/2007/03/73007
,
2007.

The **Large Hadron Collider** (LHC) produces 15 million Gigabytes of data every year.

http://lcg.web.cern.ch/lcg/public/default.htm

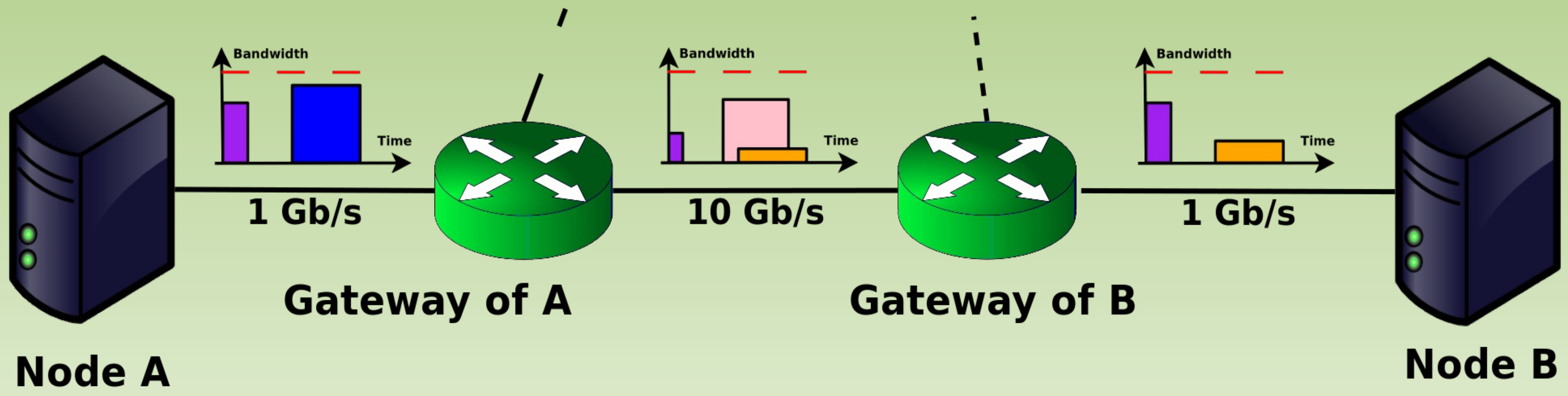
Advance Bandwidth Reservations

- One agenda per port and one per router
- End-to-end reservation (the whole path, at the same time, with identical bandwidth for all the links)



End-to-end reservation

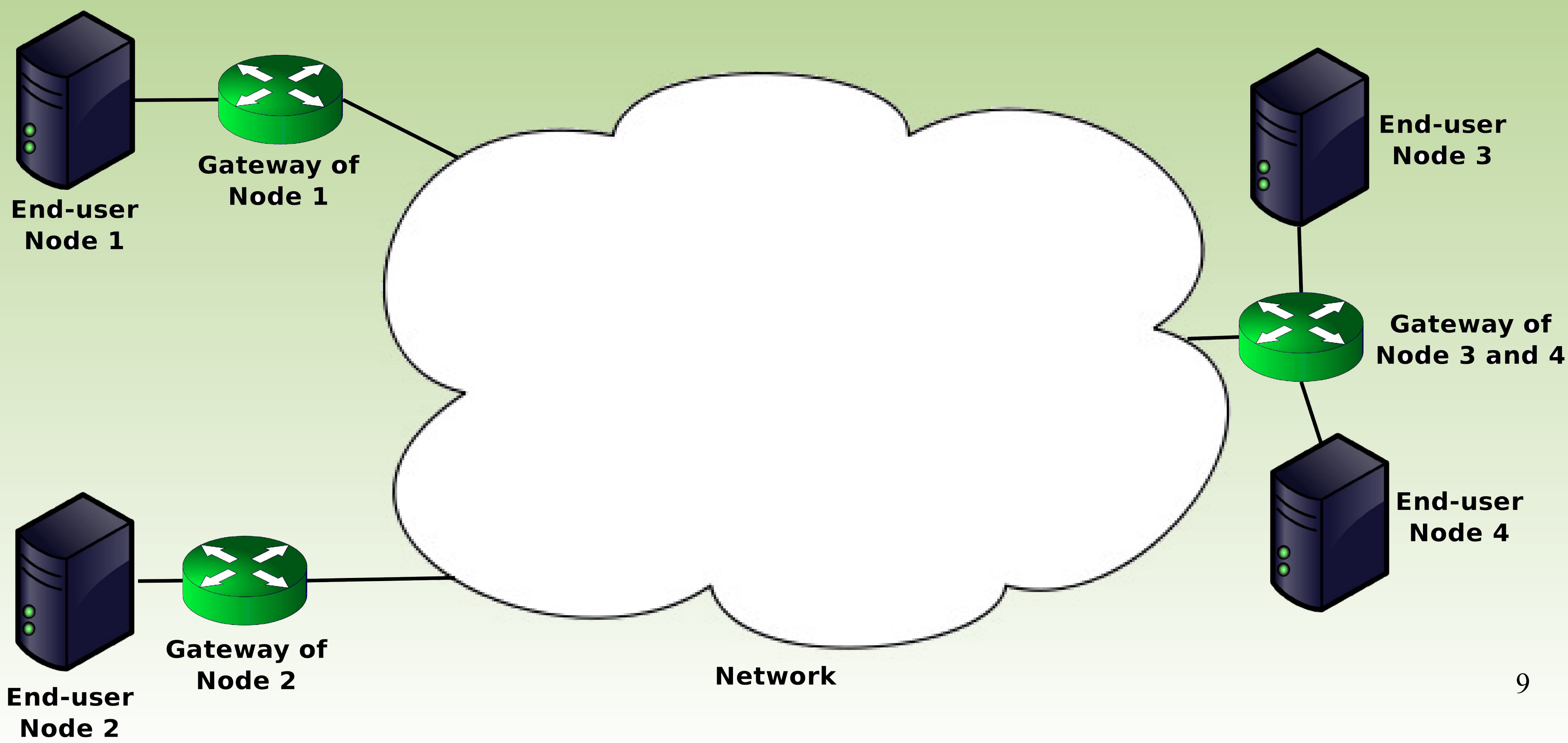
- Scheduling on all the agenda of the path



- Not store-and-forward approach

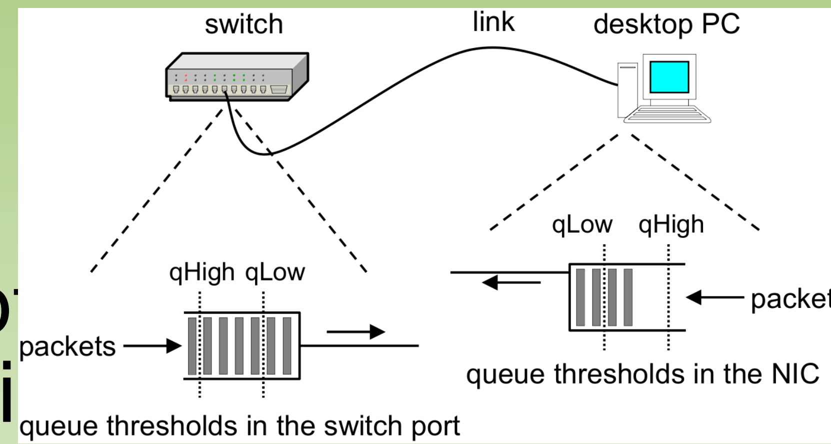
Global architecture & scenario

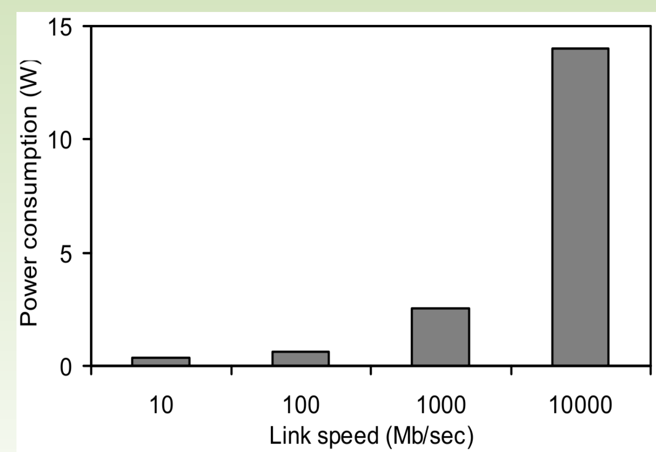
- End users want to send BDT to other end users.
- End users are connected to gateways.



Underlying assumptions

- Routers are ALR-enabled and can be switched off and on.
- Symmetric routing
- End-to-end energy consump is computed using prelimi measurements.





 Goal: **to find a good trade-off between performance (# of granted reservations) and energy**

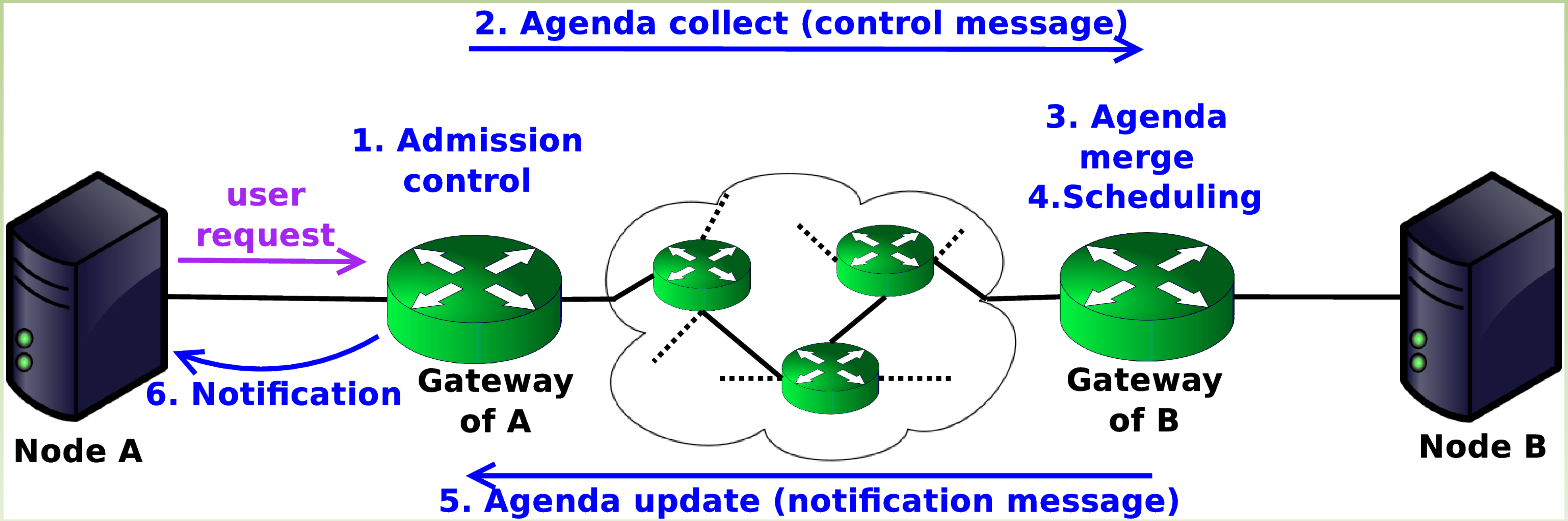
HERMES: High-level Energy-awaRe Model for bandwidth reservation in End-to-end networkS

Main characteristics

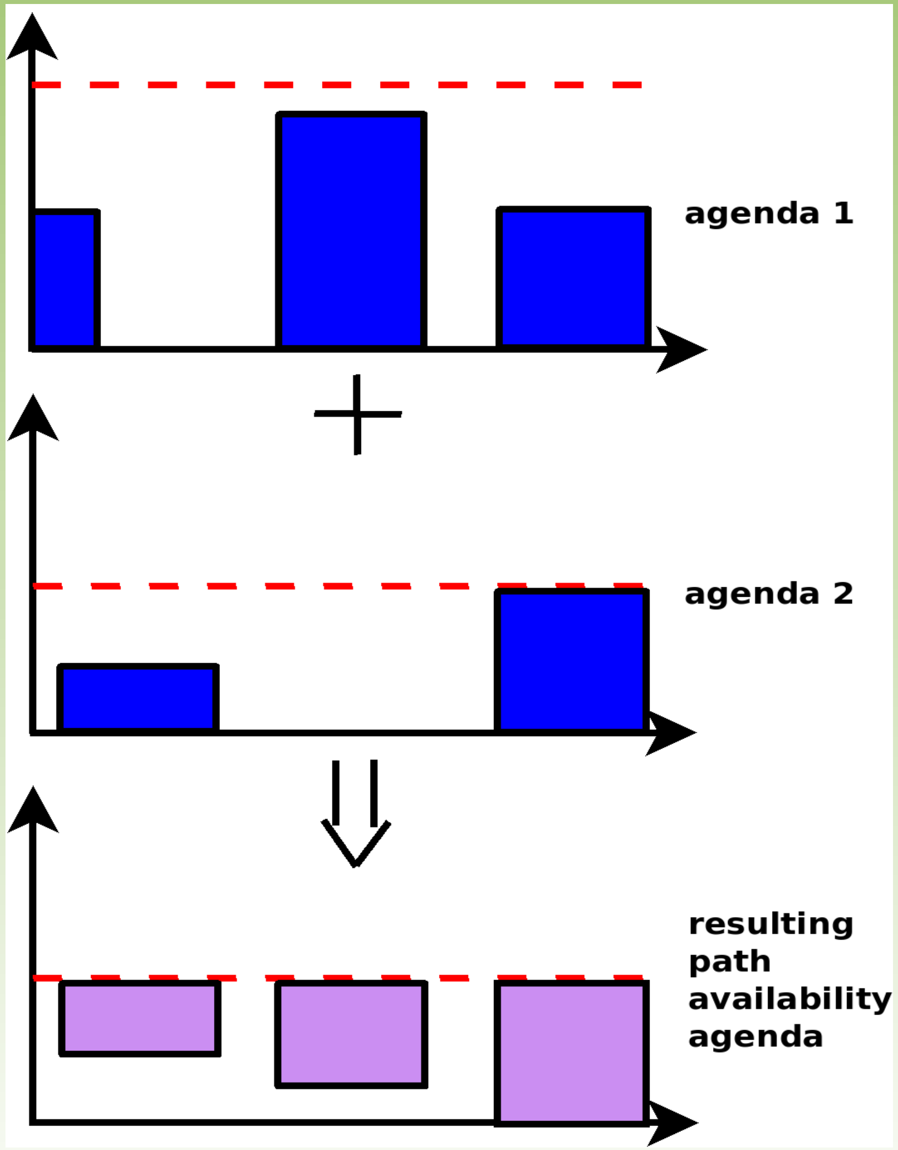
- Switching off unused nodes
- Distributed network management
- Energy-efficient scheduling with reservation aggregation
- Usage prediction to avoid on/off cycles
- Minimization of the management messages
- Usage of DTN (Disruptive-Tolerant Network) for network management purpose

Agenda collect and fusion

- One round-trip aggregated message

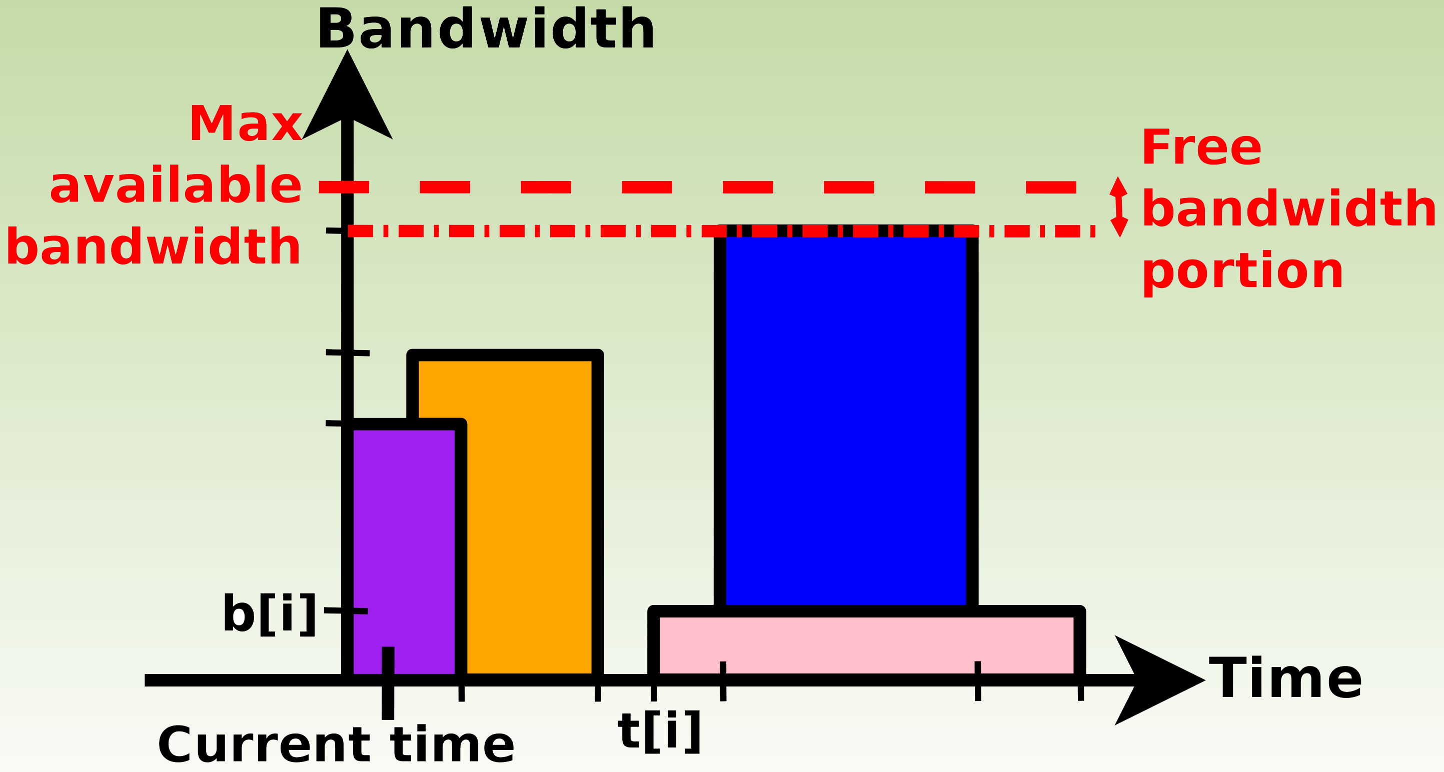


Agenda merge



ABR scheduling

- Try to put the reservation after and before each event, and estimate the energy consumption for each one
- Chose the less energy consuming option



Prediction and switching off

- At the end of a reservation, for each resource:
  - if there is a reservation soon in the agenda
    → stay powered on
  - else
    → predict the next reservation and stay on if it soon, otherwise switch off.

- Prediction using the history.

Network switched off by pieces: Disruption Tolerant Network usage

- Each reservation request has a TTL

  - if TTL = 0 → request to compute now, answer to give as soon as possible

  - otherwise, users can wait for the answer. The request moves forward into the network hop-by-hop waiting for the nodes to wake up. If the TTL is expired, the whole path is awaken.

HERMES Evaluation

Simulation Results

- BoNeS (Bookable Network Simulator)
- Written in Python (6,000 lines)
- Generates random network with the Molloy & Reed method or uses configuration file
- Generates traffic according to statistical laws:

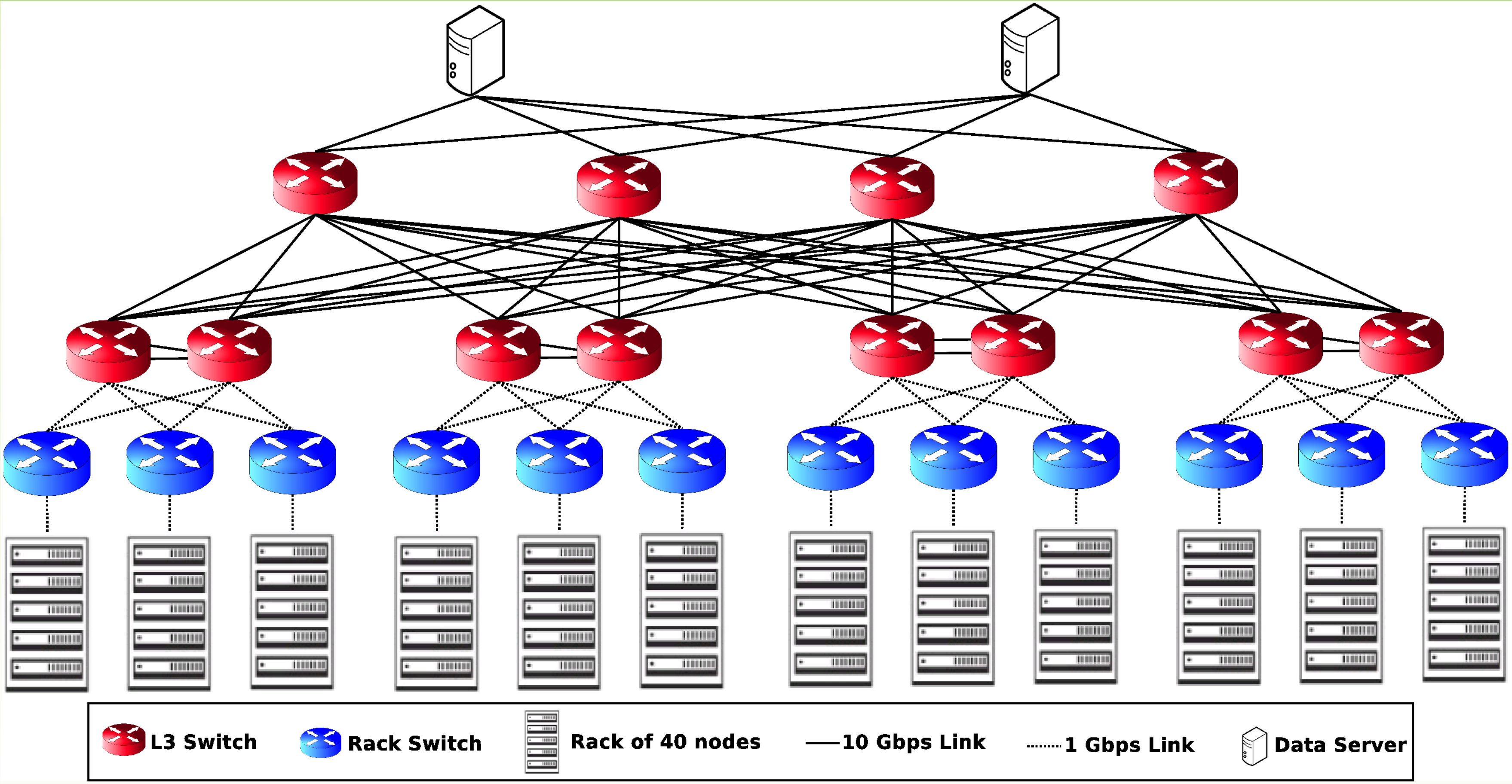
  - submission times (log-normal distribution)
  - data volumes (negative exponential)
  - sources and destinations (equiprobability)
  - deadlines (Poisson distribution)

Comparison with other schedulings

- **First**: the reservation is scheduled at the earliest possible place;
- **First green**: the reservation is aggregated with the first possible reservation already accepted;
- **Last**: the reservation is scheduled at the latest possible place;
- **Last green**: the reservation is aggregated with the latest possible reservation already accepted;
- **Green**: HERMES scheduling;
- **No-off**: first scheduling without any energy management.

Simulated Network

- Typical three-tier fat-tree architecture
- 482 servers, 24 routers, 552 links



Simulations

- All the servers can be sources and destinations.
- Time to boot: 30 s.; time to shutdown: 1 s.
- 1 Gbps per port routers:

Component	State	Power
Chassis	ON	150 W
	OFF	10 W
Port	1 Gbps	5 W
	100 Mbps	3 W
	idle, 10 Mbps	1 W

Results with a 20% workload

- 80 experiments for each value
- One hour period of simulated time for each experiment
- Energy consumption in Wh

Scheduling	First	First green	Last	Last green	Green	No off
Average (Wh)	6 111	6 039	5 684	5 625	5 944	21920
Standard deviation	97	93	76	70	84	371
Accepted volume (Tb)	141.98	141.54	120.24	113.70	141.97	141.98
Cost in Wh per Tb	43.04	42.66	47.27	49.47	41.87	154.39

Results with a 60% workload

- 60%: average occupancy per link

Scheduling	First	First green	Last	Last green	Green	No off
Average (Wh)	7 111	6 973	6 300	6 285	6 590	20 463
Standard deviation	362	335	100	106	305	809
Cost in Wh per Tb	42.18	41.37	40.21	41.25	39.09	121.37

- Compared to current case (no-off), HERMES could save **73%**, and **68%** of the energy consumed depending on the workload (20% or 60%)

Contributions and Perspectives

- Complete and energy-efficient bandwidth provisioning framework for data transfers including scheduling, prediction and on/off algorithms
- Validation of HERMES through simulations
- Perspective: to encourage network equipment manufacturers to design new equipments able to switch on and off and to boot rapidly.

Thank you for your attention!

Questions?

Anne-Cécile Orgerie
annececile.orgerie@ens-lyon.fr

ECOLE NORMALE SUPERIEURE DE LYON RESO INRIA RHÔNE-ALPES LIP UCB CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE UNIVERSITÉ DE LYON