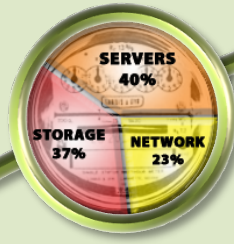
Resource Consolidation in Wired IP Networks



Aruna Prem Bianzino

ECO net

With: Jean-Louis Rougier, Dario Rossi, Claude Chaudet, Stefano Moretti, Marco Mellia, Luca Chiaraviglio

TELECOM ParisTech

Politecnico di Torino

Agenda

- The Resource Consolidation Principle
- An Overview of Classical Centralized Solutions
- On Device Criticality: The G-Game
- The ECOnet Project

Green for ICT

ICT represents a strong contribution to the environmental impact of human activities, and with a very high increasing rate:

Same footprint of the airplane transports,

... but with higher growing rate.

=

Remark: our work focuses on "energy aware" ICT

Gas emission is complex to quantify (type of energy, ...)

Economical arguments (reduce energy cost)

Green for ICT: A Hot Topic

Many works have been initiated in the last years: in ***Data Centers***, in ***Peripherals***, and in ***Networks***

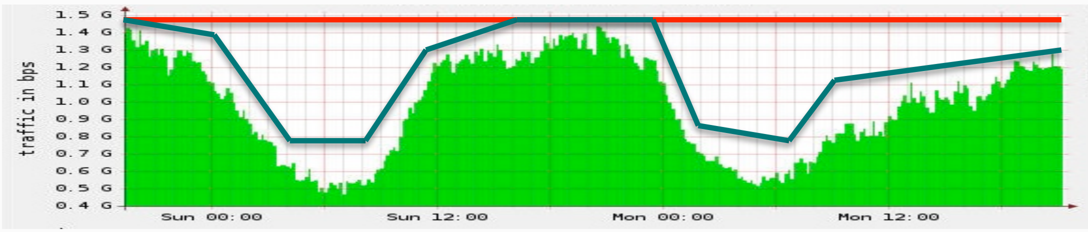
- In *wireless networks*, not completely a “new” subject:
  - Battery constraints in wireless mesh/sensor networks
  - Interferences (power control)
  - Important savings
- In wired networks:
  - Still some interesting opportunities
  - Depend on topologies

The Resource Consolidation Principle

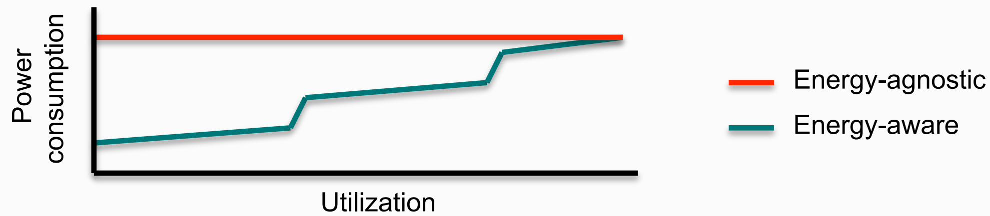
Energy Saving Opportunities

Facts:

- Network systems and devices are over-provisioned
- Traffic fluctuations

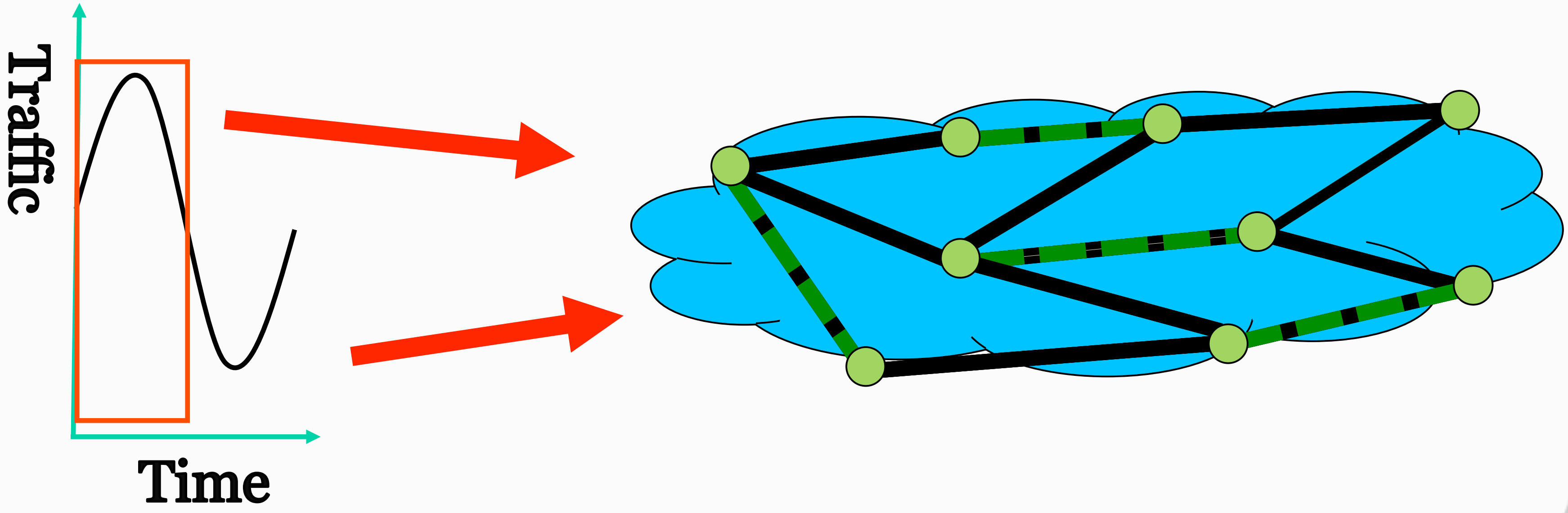


- Today: energy agnostic equipments
- How to reach proportionality (energy/utilization)

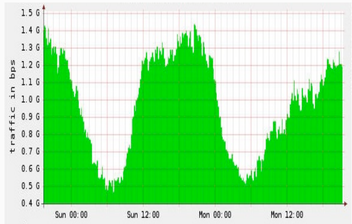


Resource Consolidation

Popular practice in other domains (e.g., Data Centers) Can be applied to obtain a non-agnostic network behavior



Centralized Approaches: Scenario



- Core to Metro IP Networks
  - Stable/Predictable traffic requests
- Central control unit:
  - Knows the network topology and TM
  - Forces the network devices' configuration
- A network configuration is possible for each traffic condition

Centralized Approaches

- Formulation as an optimization problem (MILP) [1]:
  - Minimize the network power consumption
  - QoS constraints (e.g., max link load)
- Greedy heuristics:
  - Consider devices one by one
  - Switching off a device if not affecting the network working state
  - Following a ranking based on power consumption (MP [2]), workload (LF [2]), connectivity, topology specific, etc.

On Device Criticality

Classical Approaches: Open Points

- A solution purely optimizing the energy consumption does not take into account the system robustness
- There is no control on which network elements are switched off
- Definition of a criticality index for the network devices to drive the resource consolidation process

Definition of a trade off between

Energy-saving and Traffic Engineering

A Game-Theoretical Approach: the Idea

- Modeling the communication network as a cooperative TU-game
- Each node is a player
- Every coalition is a network configuration:
  - Nodes in the coalition -> ON
  - Other nodes -> OFF (or failures)
- The amount of delivered traffic is the revenue of the coalition

A Game-Theoretical Approach

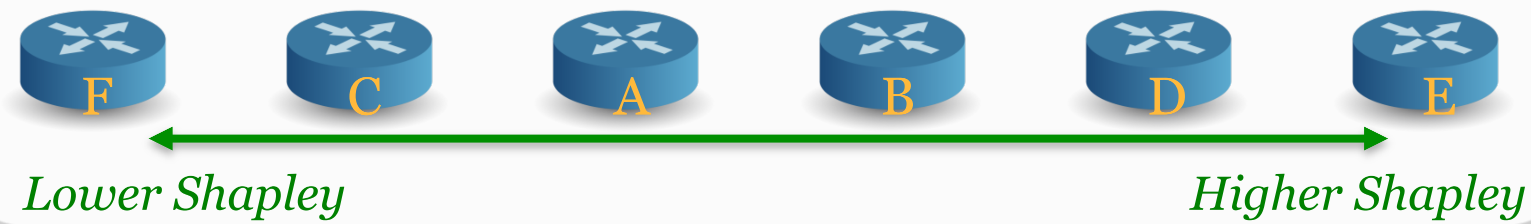
The final game is the composition of two games:

- A Traffic Game (A-Priori) over a full-mesh network (allows all coalitions, accounts only for the Traffic Requests)
- A Topology Game (A-Posteriori), which is the restriction of the first over the network graph, and accounts for the Topology

The two games may be decomposed, exploiting the network structure (paths), making the problem tractable

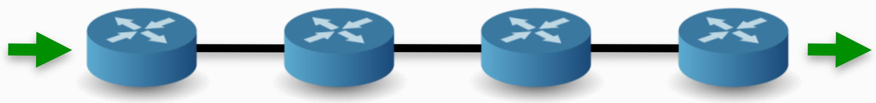
The Shapley Ranking

- The Shapley value defines a rank among players (on the basis of the *amount of traffic* that nodes contribute to carry, and of their *criticality* while composing the coalition)
- Nodes are progressively switched off (if the all traffic requests are still satisfied, with eventual maximum load constraints)

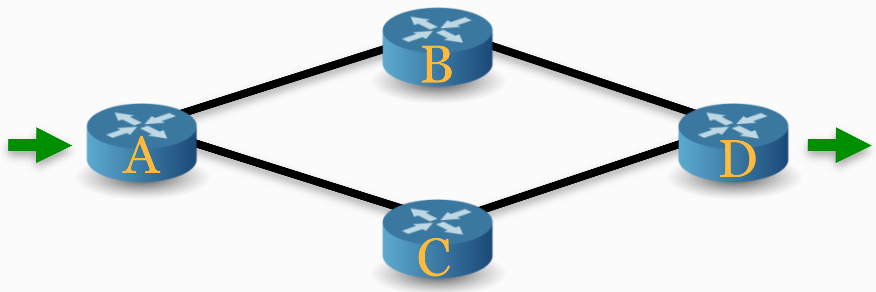


The Shapley Value: an Example

The Shapley Ranking: Toy Cases

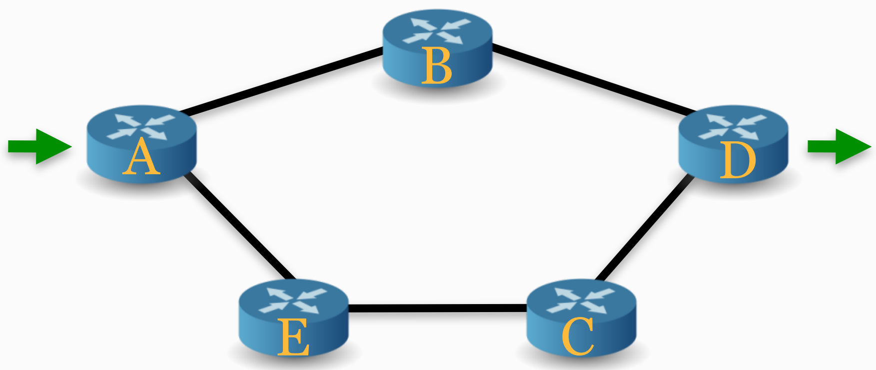


Every node has the same Shapley value



A, D -> 5/12

C, B -> 1/12



A, D -> 23/60

B -> 8/60

C, E -> 3/60

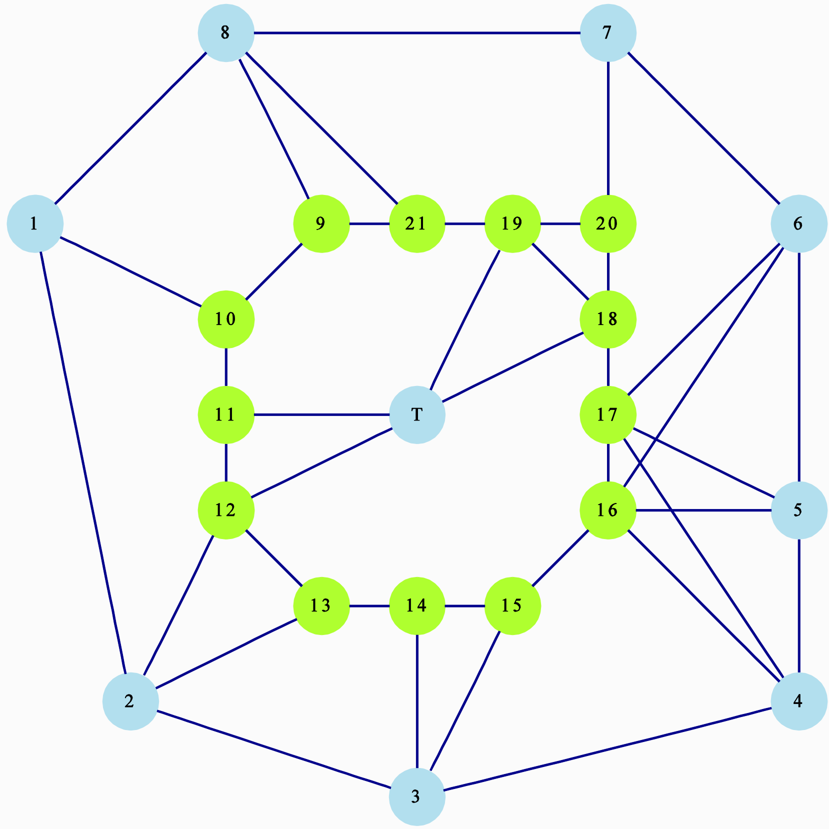
Other Possible Rankings

- Other criticality indexes are present in the literature, but all of them only account for the network topology
- The G-Game considering an uniform TM matches quite well these indexes, but there is low correlation when taking into account the Traffic

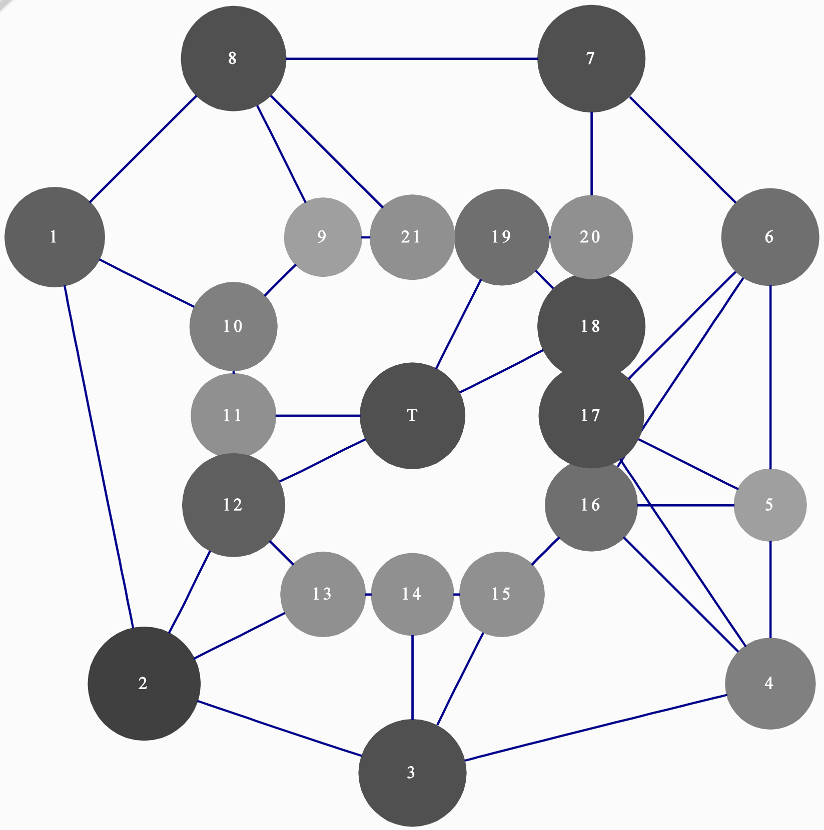
	G-Game (U-TM)	*Betweenness*	*Degree*	*Closeness*	*Eigen*	*G-Game*	*Load*
G-Game (U-TM)	1						
Betweenness	0.9688	1					
Degree	0.4594	0.5321	1				
Closeness	0.8729	0.9057	0.6216	1			
Eigen	-0.0073	0.0792	0.7335	0.1787	1		
G-Game	0.4085	0.4286	0.2527	0.5132	-0.0220	1	
Load	0.4251	0.4868	0.4762	0.6046	0.1911	0.5583	1

A Real Case Study: The Network Scenario

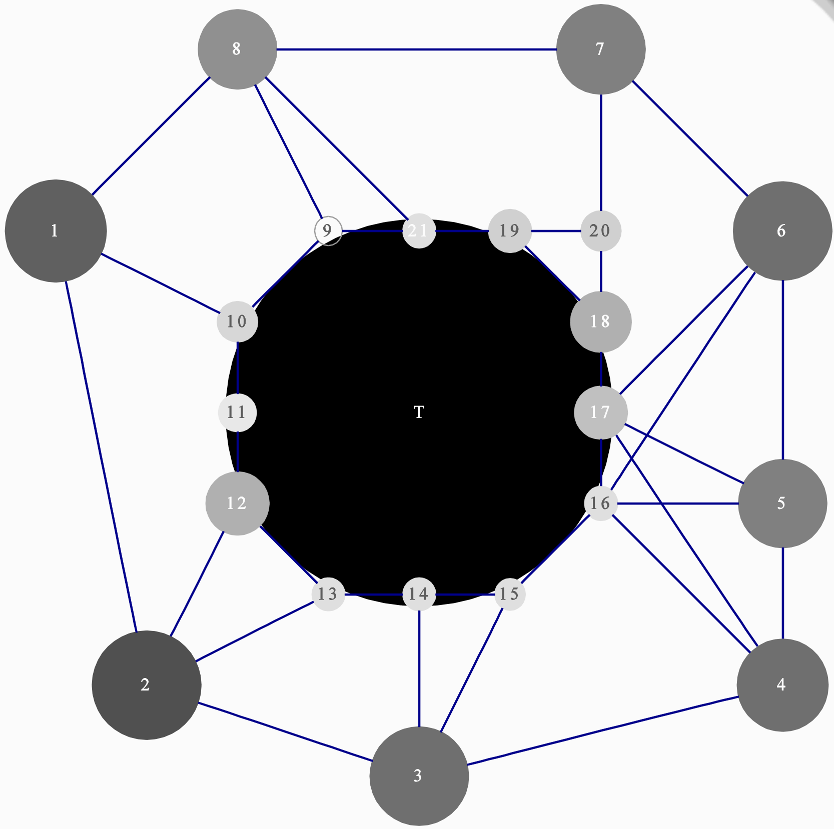
- TIGER2 Network (typical access/metro network)
- *Access* nodes (traffic Sources and Destinations)
- *Core* nodes (only traffic transport)



A Real Case Study: Different Rankings

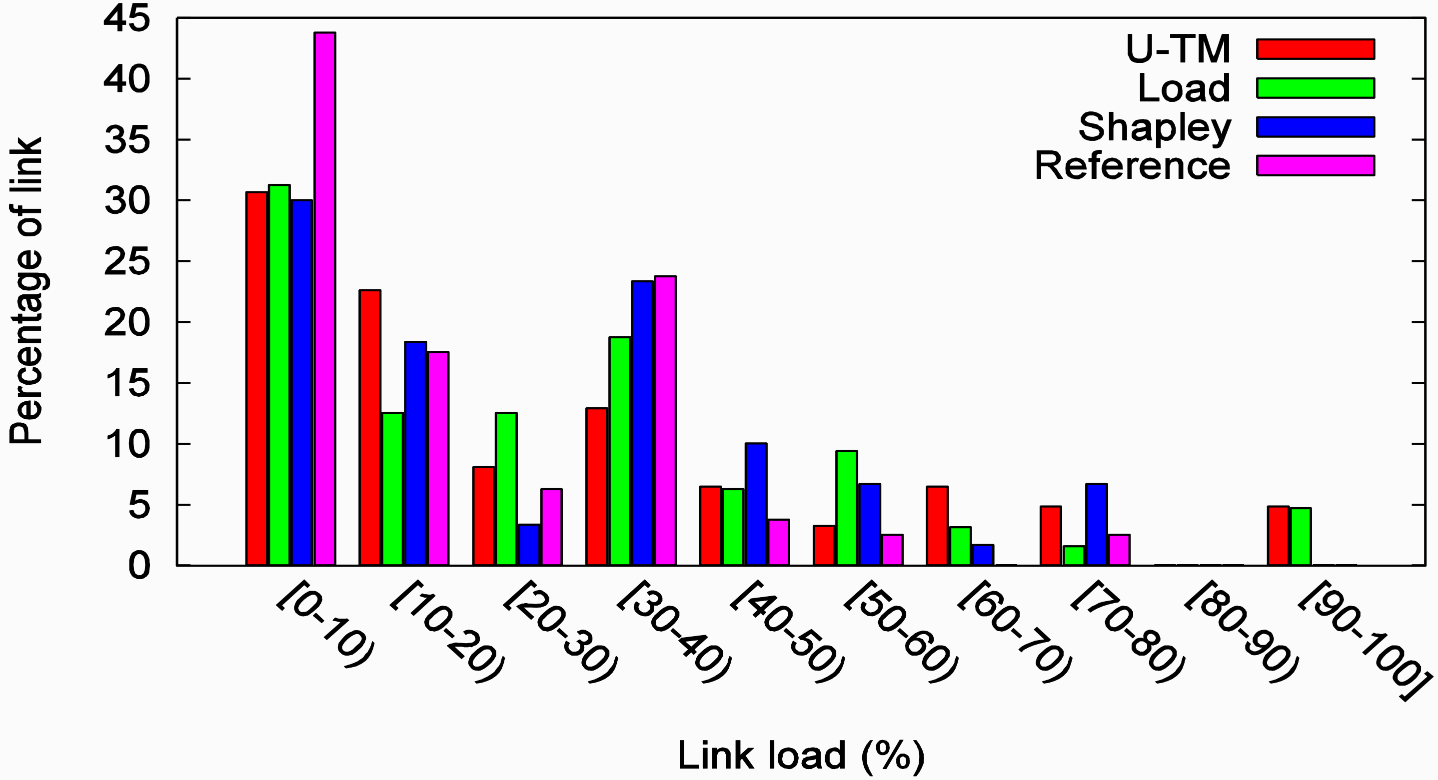


Only Topology



G-Game

A Real Case Study: Results (ii)



Order	**Shapley**	**U-TM**	**Load**
Energy saving (%)	17.05	13.43	16.27
Weighted avg path length	2.99	3.40	3.25

The G-Game: Conclusions

- The G-Game [3] allows taking into account:
  - The network topology
  - The traffic condition
  - The "degraded" scenarios (some devices off)
- Better trade off between energy saving and QoS than previously proposed and classical rankings
- Similarly, the L-Game has been designed to define a criticality ranking among links, with similar results on resource consolidation

Distributing Centralizes Solutions

Toward Distributed Solutions

- Lower algorithm complexity
- Not relying on a central control point
- Not requiring:
  - Strict synchronization among devices
  - Knowledge of the TM and routing paths
- Easier configuration and management
- Automatic reaction to changes (traffic congestions, failures)

Recalling MP and LF

- A solution is computed at every change in the traffic conditions
- Devices are ordered by increasing link load (LF), or decreasing power consumption (MP)
- Sequential switch off attempts are performed for each device once:
  - No disconnections/overloads caused -> switch the device off
  - Else -> keep it on

Distributing MP and LF: Scenario

Assumptions:

- An IP network is considered,
- Each node runs OSPF/IS-IS protocol to share:
  - link load
  - link energy cost
  - network topology
  - and to provide coarse synchronization

Difference from previous work:

- No knowledge of Traffic Matrix, nor routing paths
- No centralized decisions

Goal:

- Automatic reaction to changes in the traffic, to congestion and faults

A Distributed Approach: DLF and DMP

A simple algorithm is run at every node, on the basis of the shared knowledge [4]

At random time intervals, switch off attempts are executed by nodes, coordinated by the LSA state

- Target: the least loaded (*DLF*) link
- Target: most power consuming (*DMP*) link

Nodes select the same target link

The two nodes responsible for such link manage the switch off attempt

A Distributed Approach: **Switch OFF**

Responsible nodes attempt to switch off the target link, subject to:

- An immediate connectivity check
- A cross-check to verify that no congestion has been caused (done via the next LSA)

If fails, link is turned immediately back on

A *tabu list* is kept to blacklist the links that cannot be switched off

- Size of *tabu list* impacts the algorithm evolution

A Distributed Approach: **Switch ON**

In case of traffic congestion on some link, nodes choose a link to be switched back on:

- The last switched off link (*lastSleep*)
- The closest link to the overload (*distance*)

DLF and DMP, the Pseudo-Code

```
Distributed Choice
Input: (i,j)*, lastLSACriticalState
  if lastLSACriticalState == OK:
    if connectivityCheck(x_(i,j)* = 0) == OK:
      x_(i,j)* = 0
      to_be_verified.append(x_(i,j)* = 0)
    else
      tabu.append((i,j)*)
      if length(tabu) > maxLength:
        removeOlder(tabu)
  else:
    ij = selectLink(offLinks)
    x_ij = 1
    offLinks.remove(ij)
```

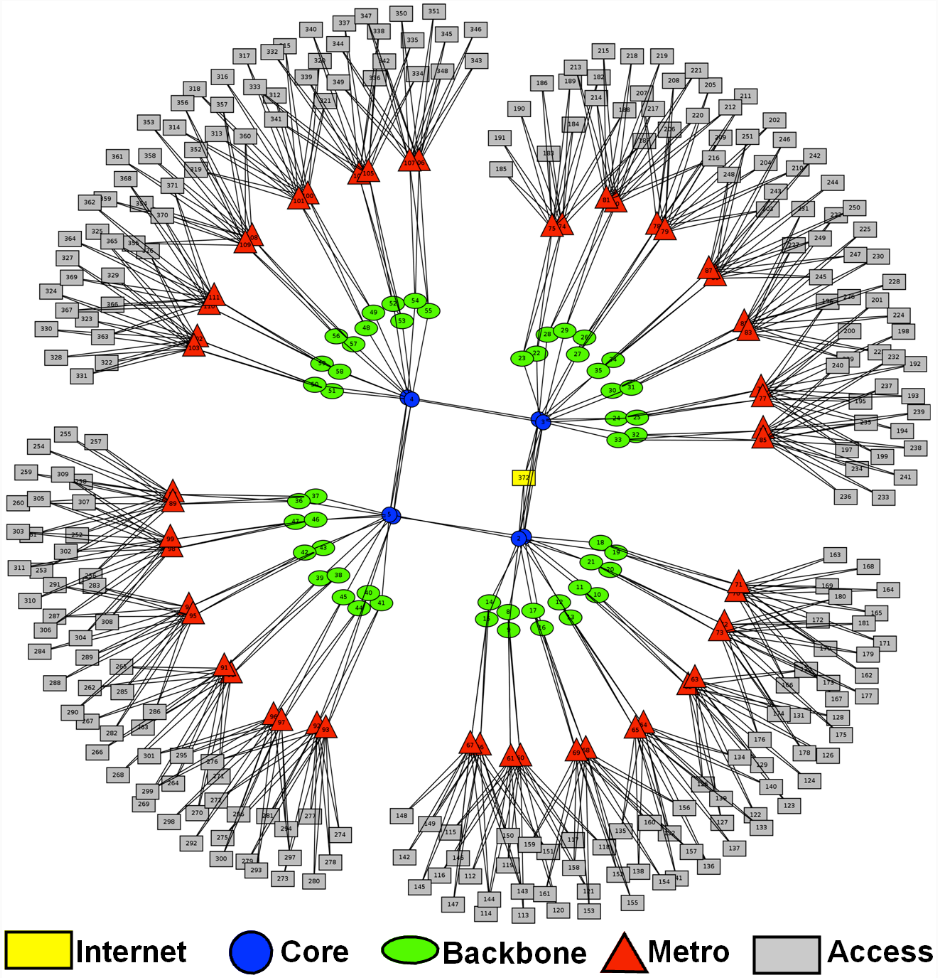
Switch off attempt

Switch on

Tabu list

```
LSA receipt
Input: LSACriticalState
  while length (to_be_verified) > 0:
    ij = removeOlder(to_be_verified)
    if LSACriticalState == KO:
      tabu.append(ij)
      if length(tabu) > maxLength:
        removeOlder(tabu)
      x_ij = 1
    else:
      offLinks.append(ij)
```

The Simulation Scenarios



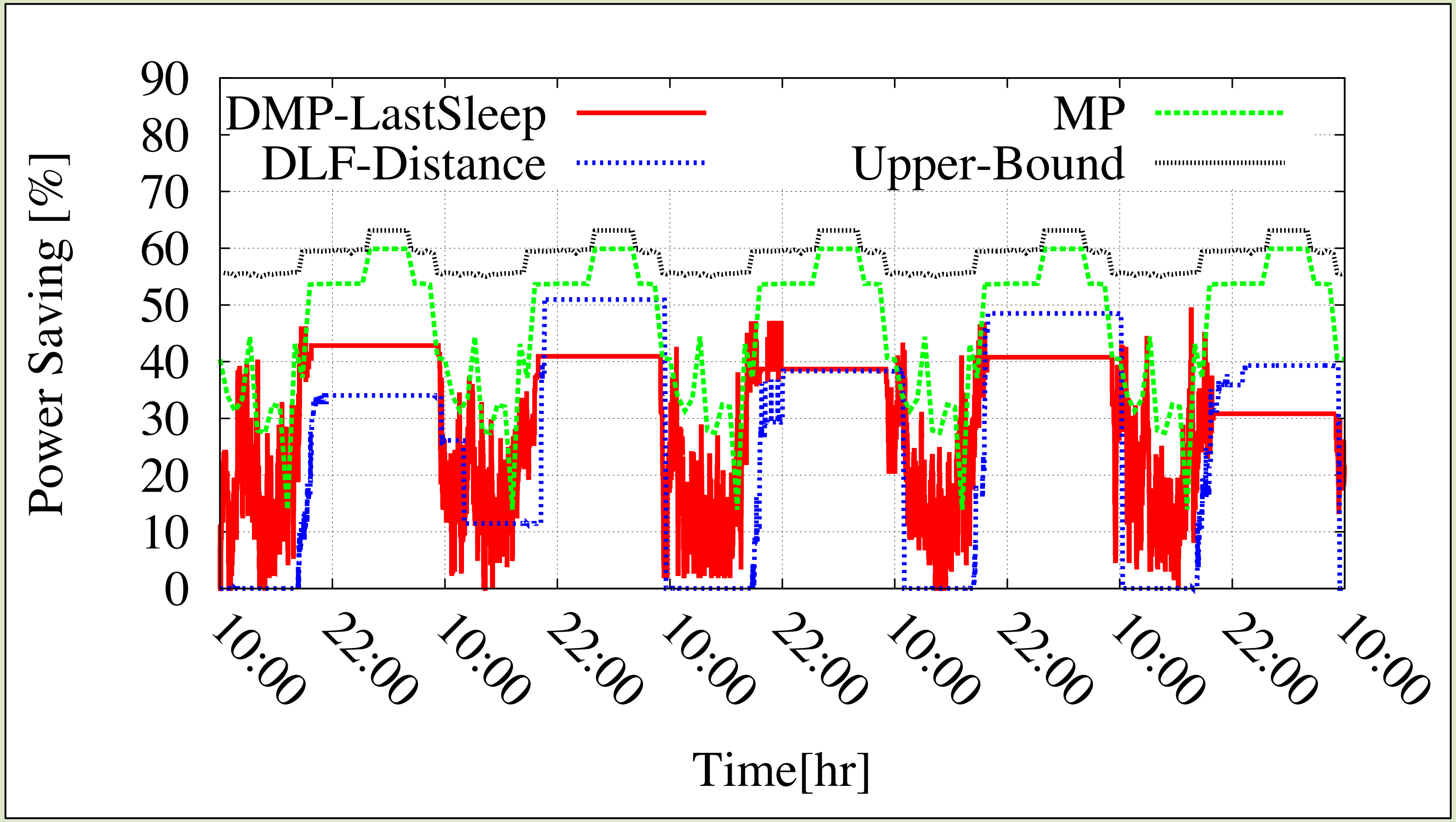
- National-wide ISP in Italy
- Hierarchical structure
- Actual traffic profile

Parameter	Symbol	Value
Inter-choice interval	Δ_C	20 sec
Number of Nodes	N	373
Number of Links	L	718
TM change interval	Δ_{TM}	48 min
Maximum Link Load	Φ	50%

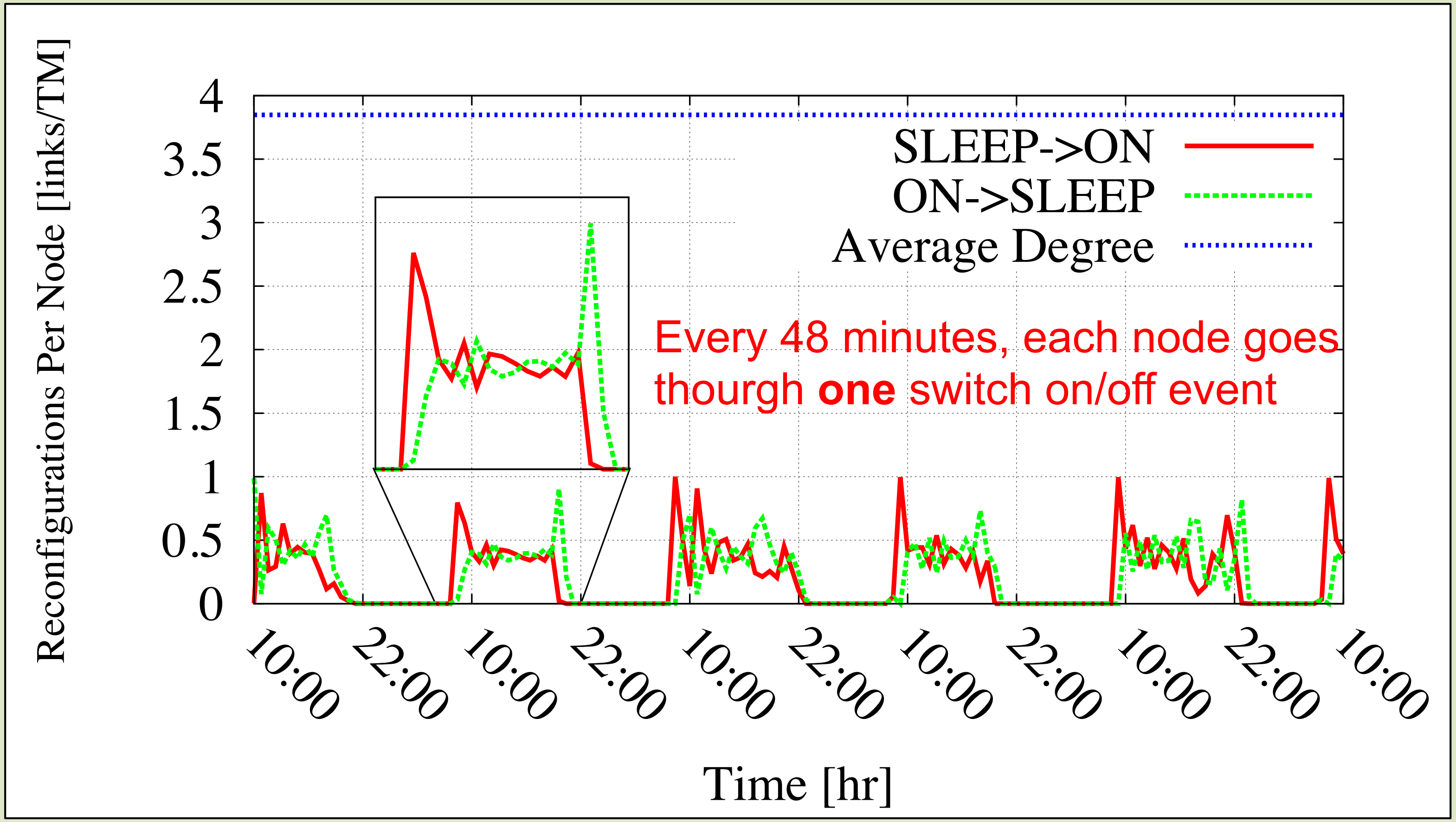
Simulation Parameters

Parameter	Symbol	Value
Inter-choice interval	Δ_C	20 sec
Inter-LSA interval	Δ_{LSA}	10 sec
Tabu-list length	maxLength	70 links
Number of Nodes	N	373
Number of links	L	718
TM change interval	Δ_{TM}	48 min
Maximum Link Load	Φ	50%

Simulation Results: Temporal Evolution



Simulation Results: Temporal Evolution



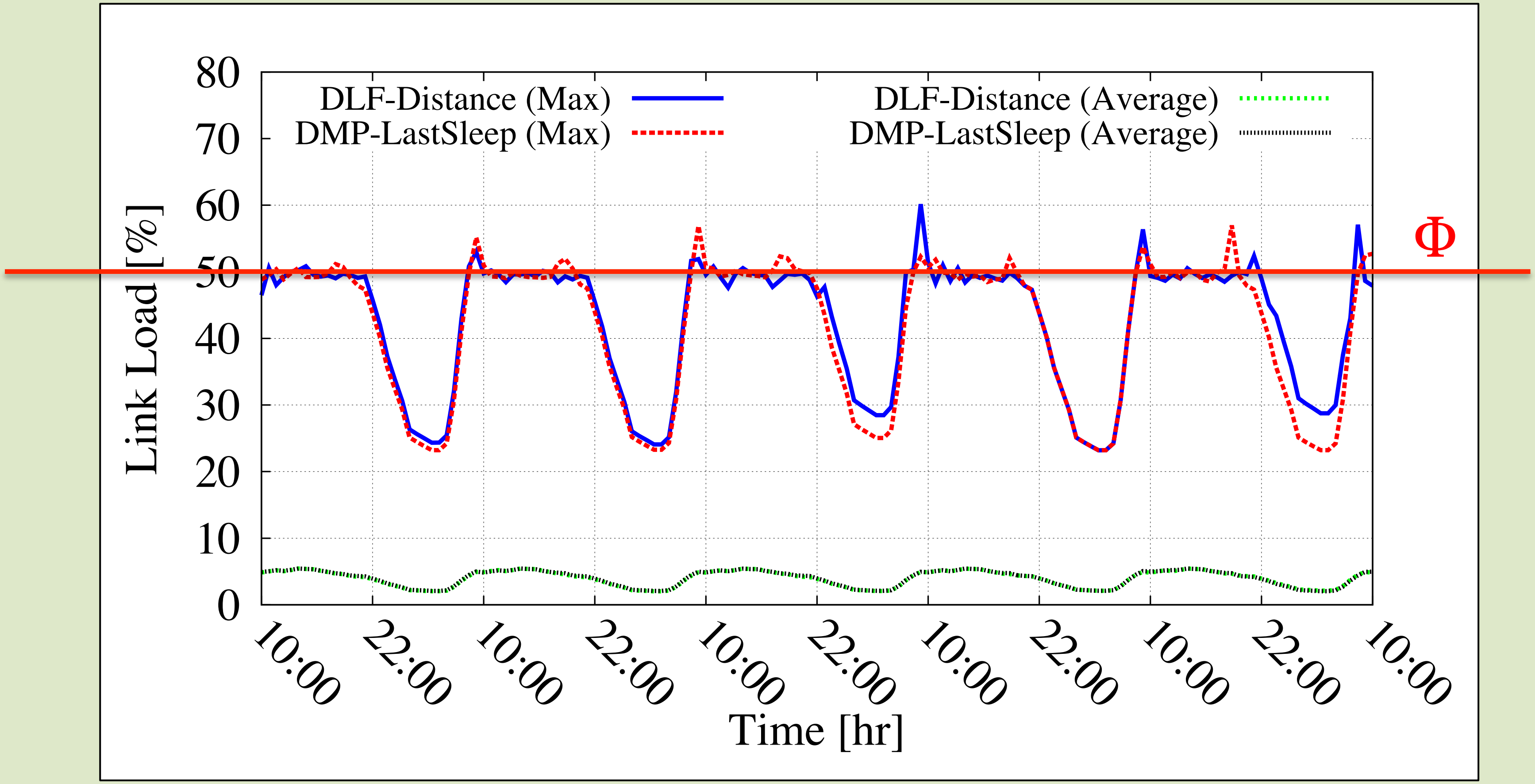
Performance evaluation

The considered performance indicators are:

- *Energy saving*
- Number of *unaccepted choices*
- Network *overload* (ξ):
  - Fraction of traffic exceeding the link overload threshold (Φ) versus total traffic

Simulations average one week of evolution

Performance evaluation: Overload

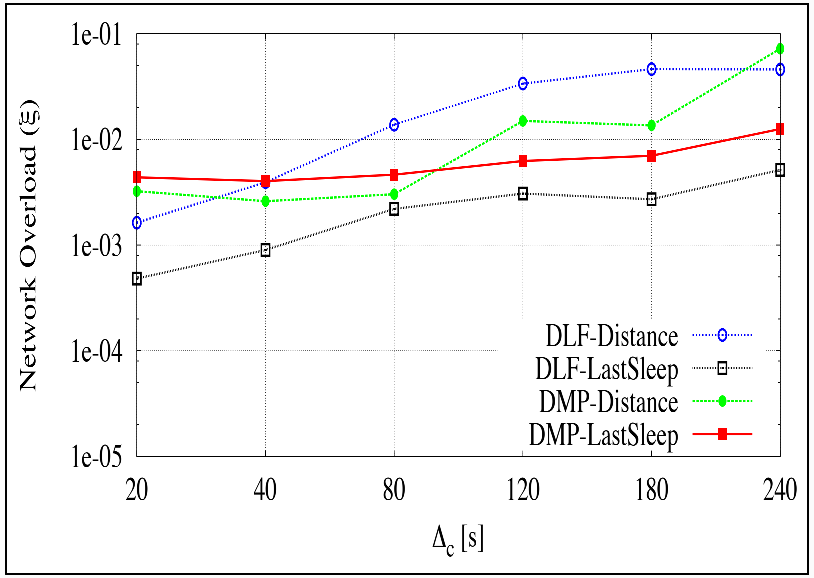


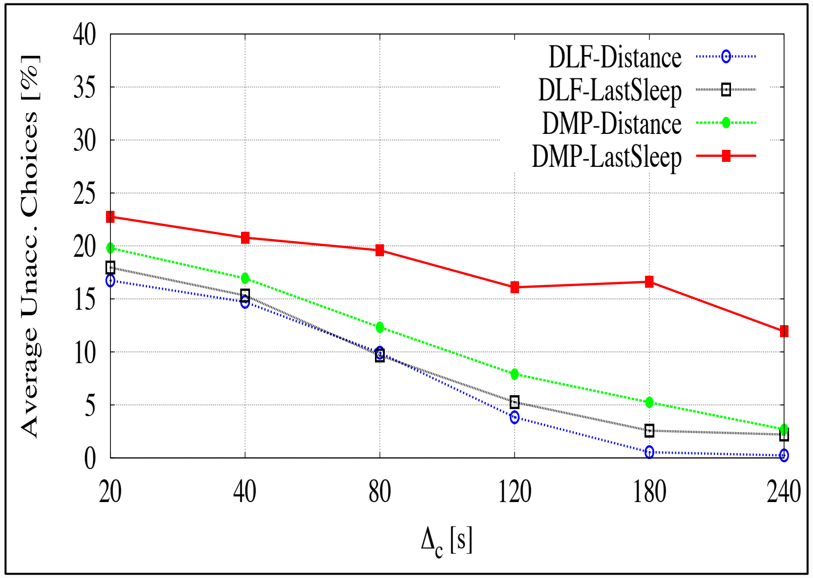
Performance evaluation (ii)

Algorithm	Saving [%]	Unacc. Choices [%]	Overload (ξ)
Upper-Bound[1]	58.56	Centralized	
MP [2]	46.28		
DMP-distance	32.35	20	3.23e-3
DMP-lastSleep	30.30	23	4.37e-3
DLF-distance	25.45	17	1.63e-3
DLF-lastSleep	19.66	18	4.81e-4

E.g., 20 violations per hour, each lasting 20 seconds, with load of 5.5Gbps over 10Gbps links, result in an overload of 0.001

Sensitivity Analysis





- *Impact of inter-choice interval (Δ_C): higher choice frequency helps the network following traffic changes*

Sensitivity Analysis (ii)

- *Size of the tabu list*: limited impact. Exploiting memory is beneficial up to a certain limit, where it starts contrasting exploration (~10% of L, in the considered scenario).

- *LSA frequency*: the algorithm is robust with respect to this parameter. Too low frequency of LSA slightly deteriorate the QoS performance.

GRiDA: a Green Distributed Algorithm

A Distributed Approach: Scenario

The same scenario of the previous solution is assumed:

- An IP network is considered
- Each node runs a link state routing protocol

But nodes share only information about:

- network state (normal/congested)
- network topology

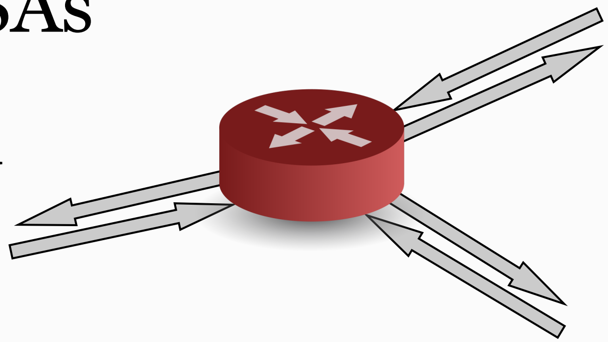
Difference from previous solution:

- Nodes take independent and selfish decisions
- Decisions are based on local information

A Distributed Approach: GRiDA (i)

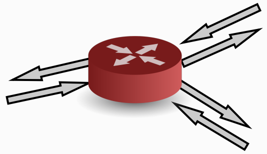
A simple algorithm is run at every node [5]:

- nodes take *independent* decisions, optimizing a *selfish* utility function, at *random time* intervals
- decisions are based on (i) *load* and (ii) *energy cost* of incident links, and (iii) the *network state* reported by periodic LSAs
- No synchronization required
- Based on Q-learning



A Distributed Approach: GRiDA (ii)

The utility function:



$$min_K\ U(K,S) = c(K) + p(K,S)$$

Where:

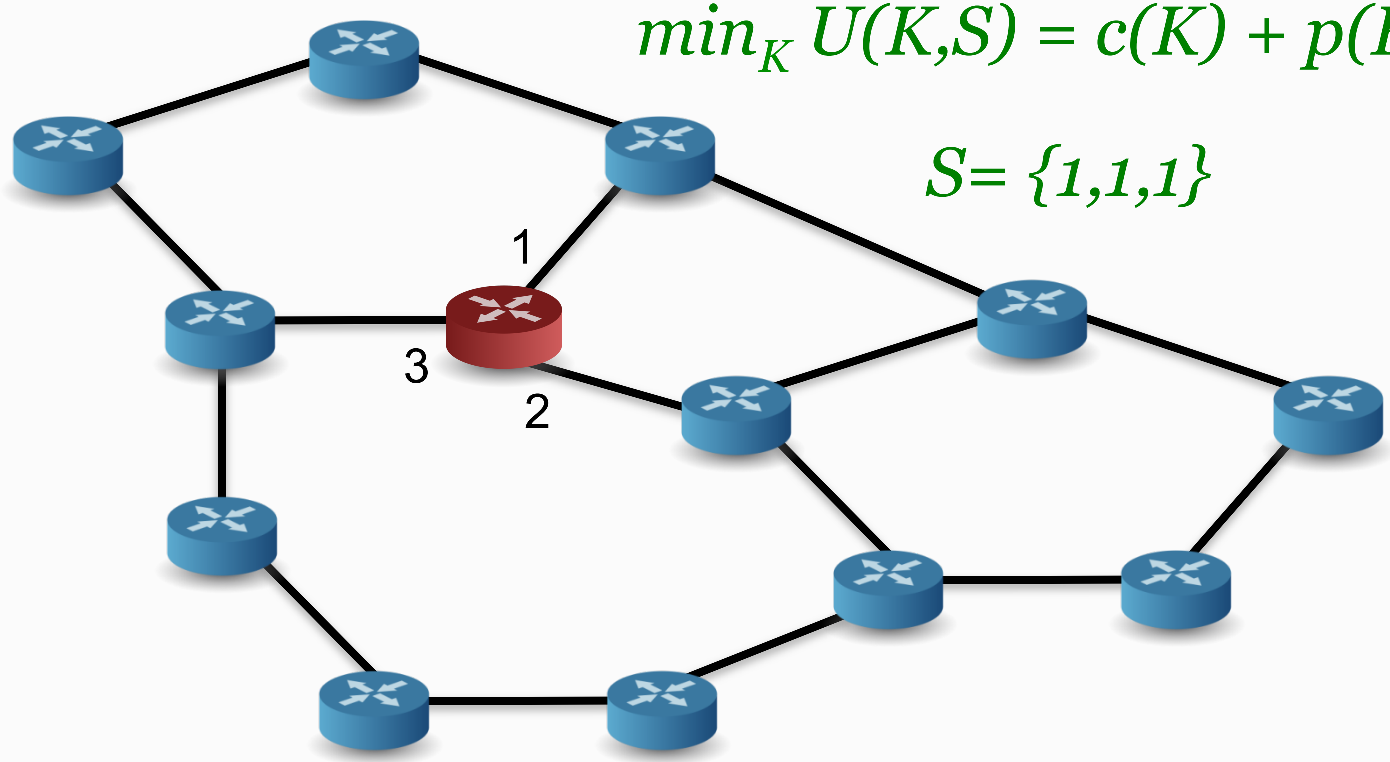
- K = node configuration $[k_1, \ldots, k_d]$, $k_i \in \{0,1\}$
- S = node status $[s_1, \ldots, s_d]$, $s_i \in \{off, normal\}$
- $c(K)$ = Energy cost of the configuration: $\sum_{i=1}^{d} k_i c_i$
- $p(K,S)$ = cost associated to the configuration, on the basis of the status and the history

GRiDA: a Toy Case

Choice:
network state = OK

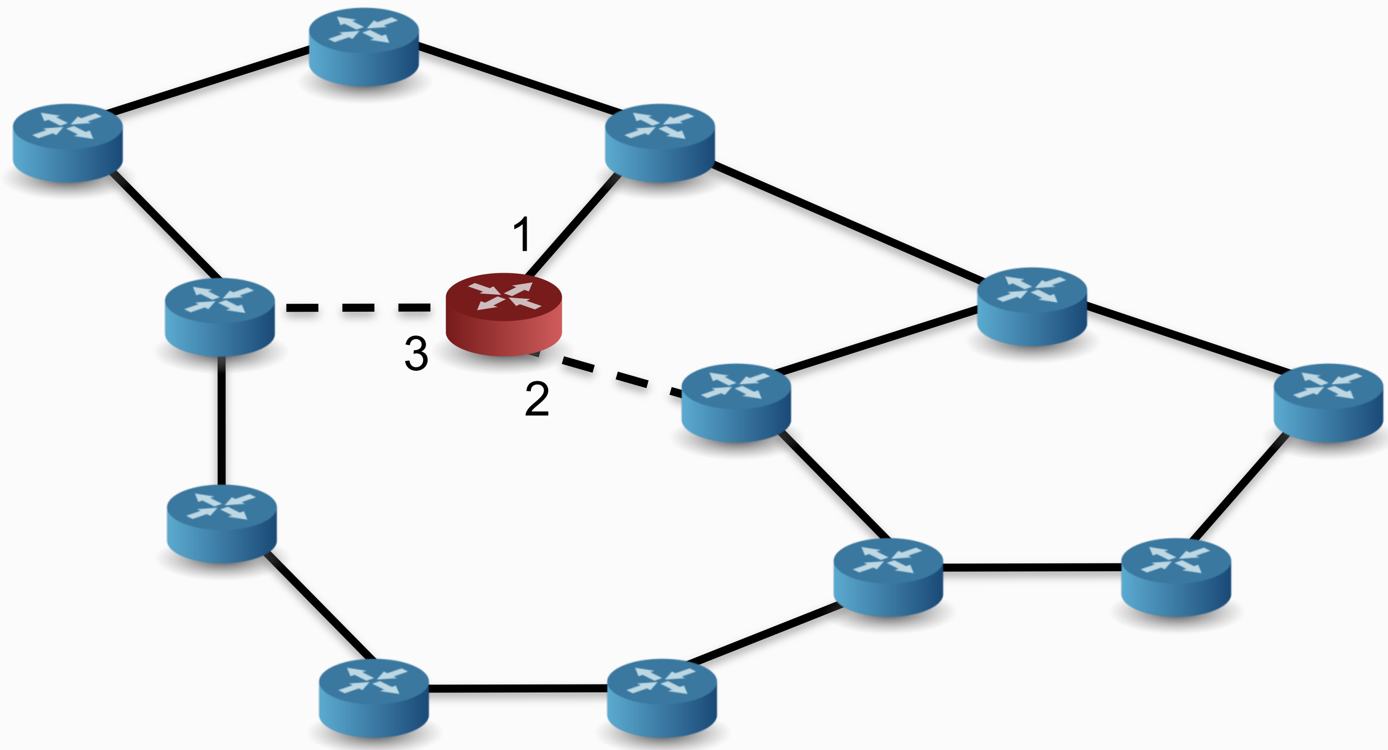
$$\min_K U(K,S) = c(K) + p(K,S)$$

$S = \{1,1,1\}$



GRiDA: a Toy Case

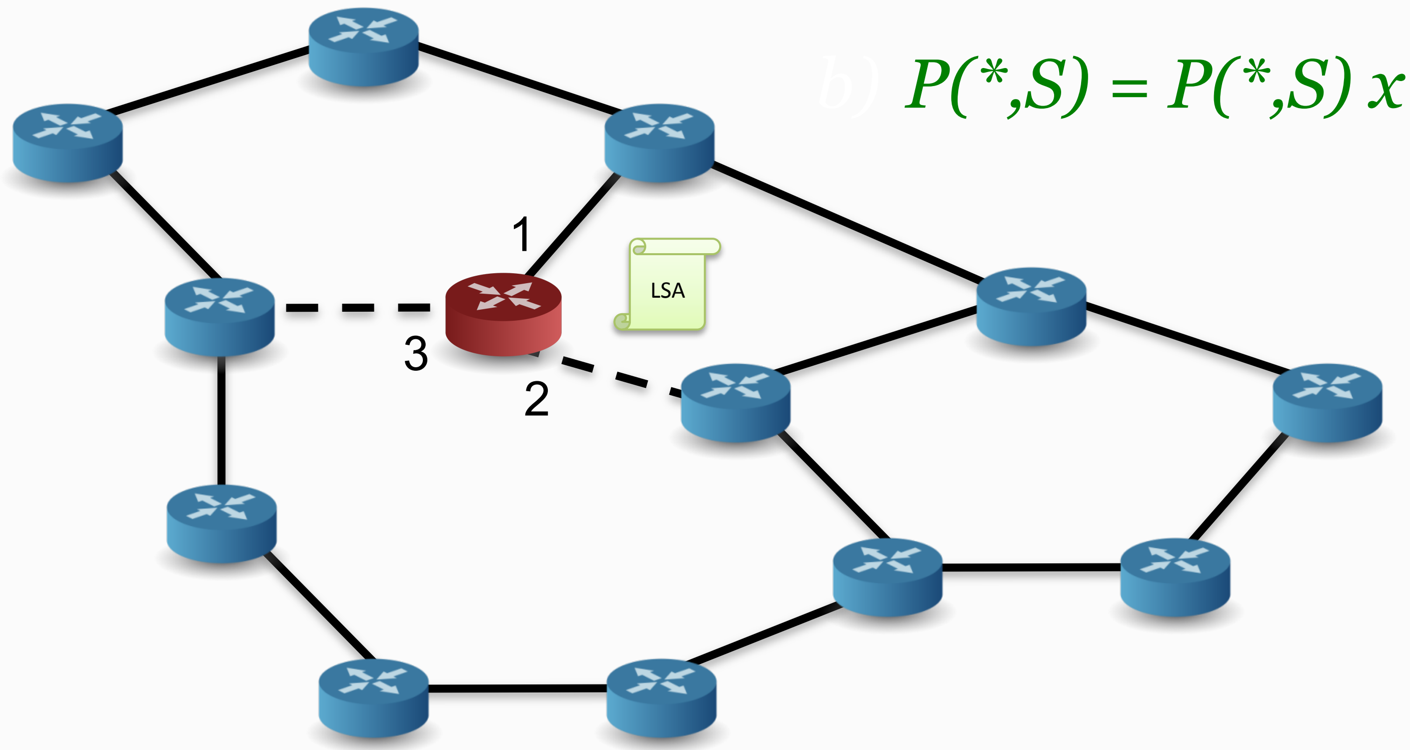
Choice: K={1,0,0}



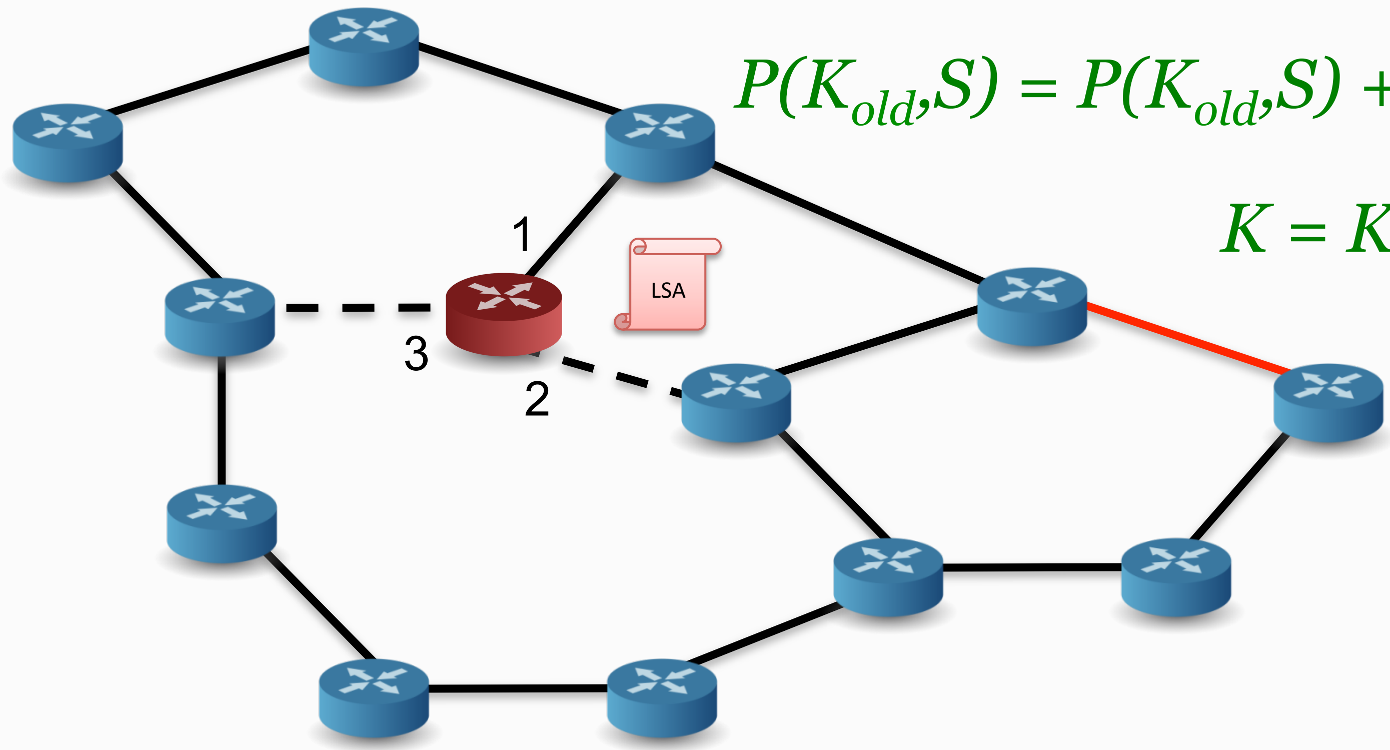
GRiDA: a Toy Case

a) LSA= OK:

b) $P(*,S) = P(*,S) \times \delta$



GRiDA: a Toy Case



a) b) LSA= KO:

$$P(K_{old},S) = P(K_{old},S) + \beta$$

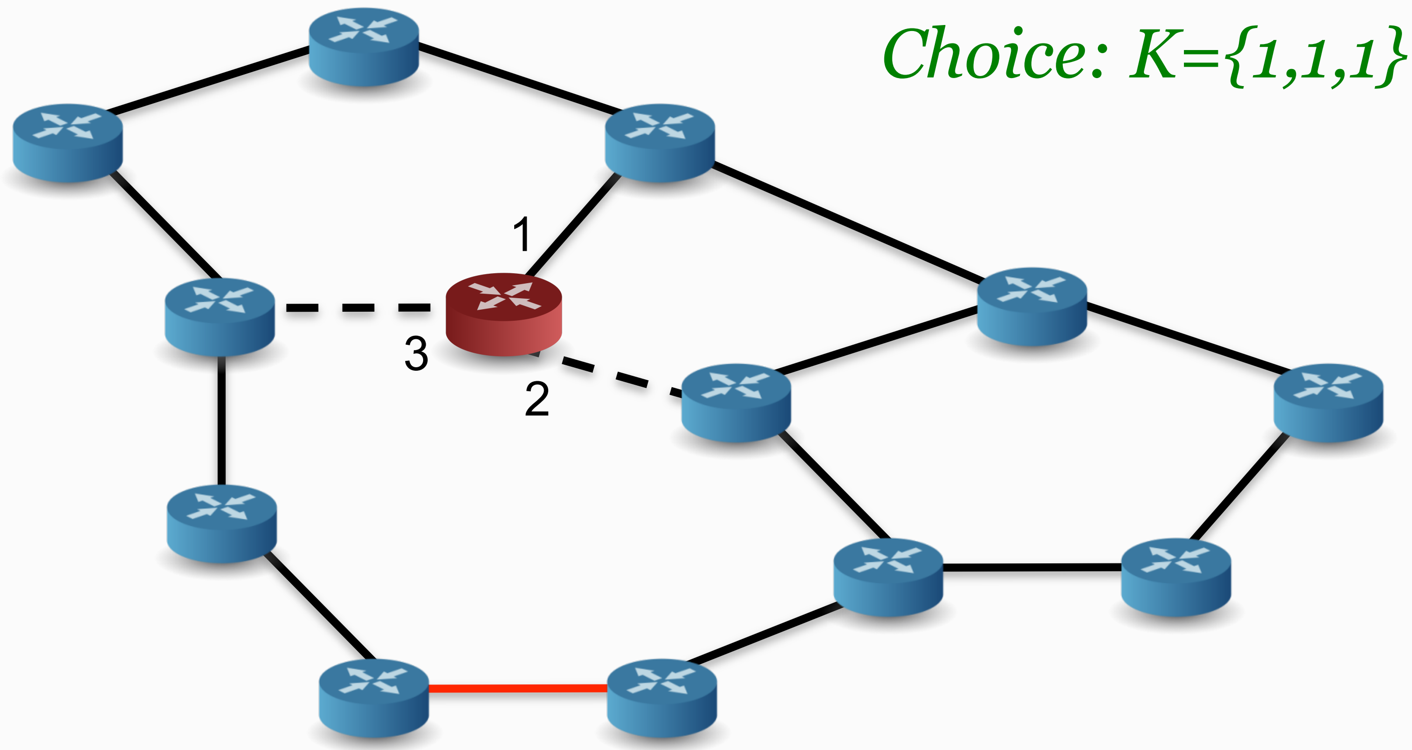
$$K = K_{old}$$

GRiDA: a Toy Case (ii)

Choice:
network state = *KO*

S= {1,0,0}

Choice: K={1,1,1}



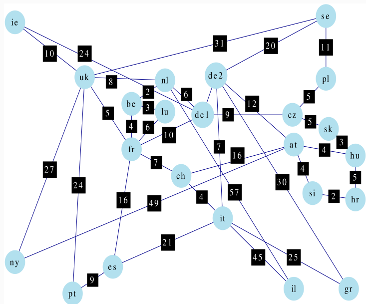
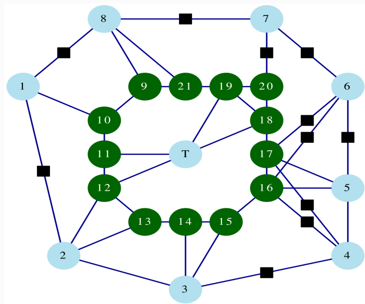
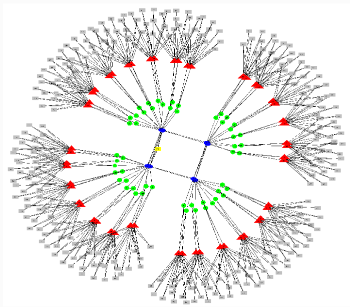
GRiDA: the Decision Process

- If the network is congested/disconnected → the node is forced to the all-on configuration

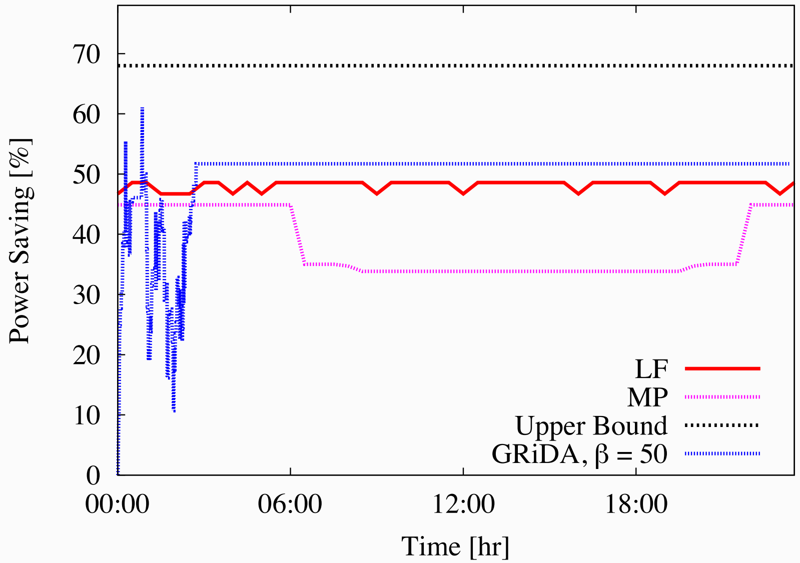
- Else → it enters the configuration K minimizing U(K,S)
- If the subsequent LSA reports congestion or disconnection → go back to previous configuration and increment p(K,S) by $\beta > 0$ (additive increase)
- Else → decrement p(*,S) by $\delta < 1$ (multiplicative decrease)

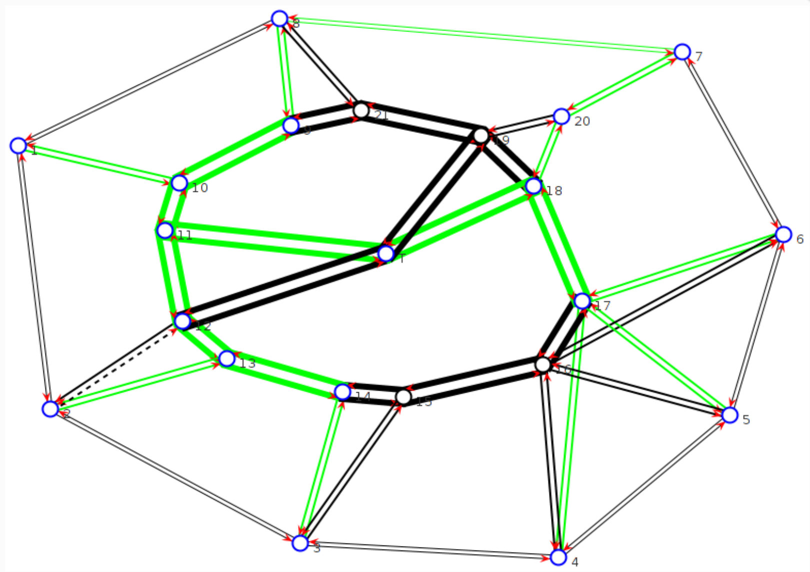
GRiDA: Simulation Scenarios

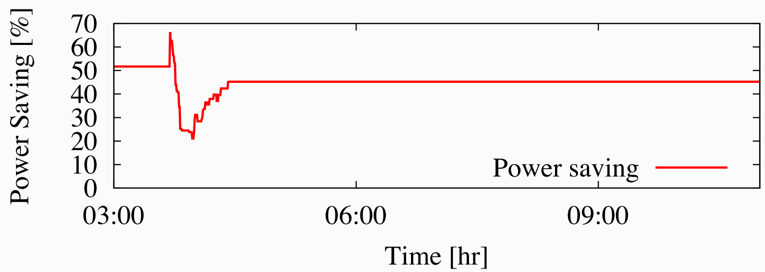
parameter	Geant	ISP 1 (metro)	ISP 2 (national)
Δ_{LSA} (s)	5	5	2
Δ_{TM} (min)	30	30	48
$\Delta_{c,\ Max}$ (s)	25	25	9
N	23	22	112 + 260
β	50,0	50,0	100,0
Φ [%]	70,0	70,0	50,0
Choices/node/ TM	6,2	6,3	3,2

GRiDA: Convergence

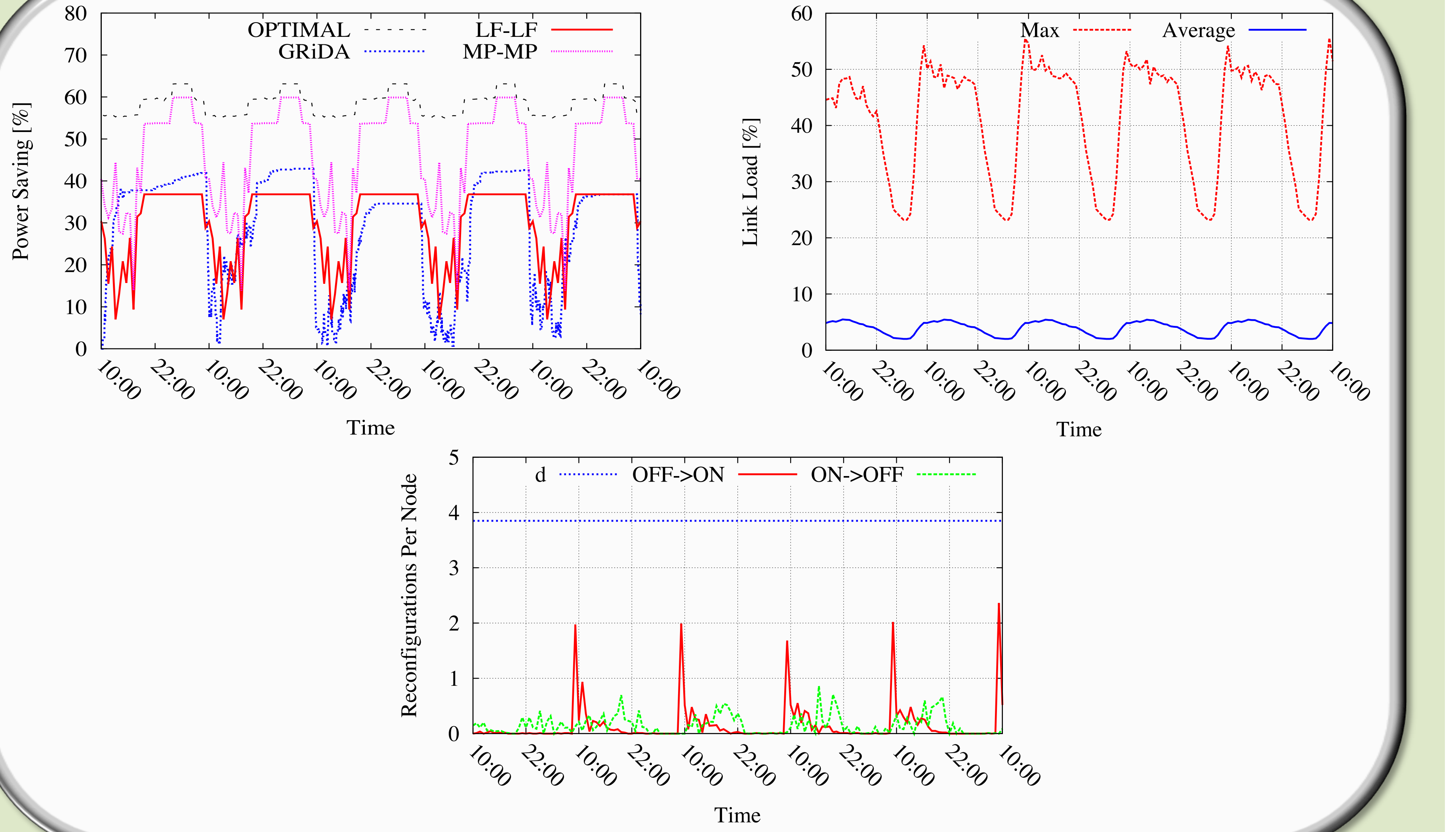




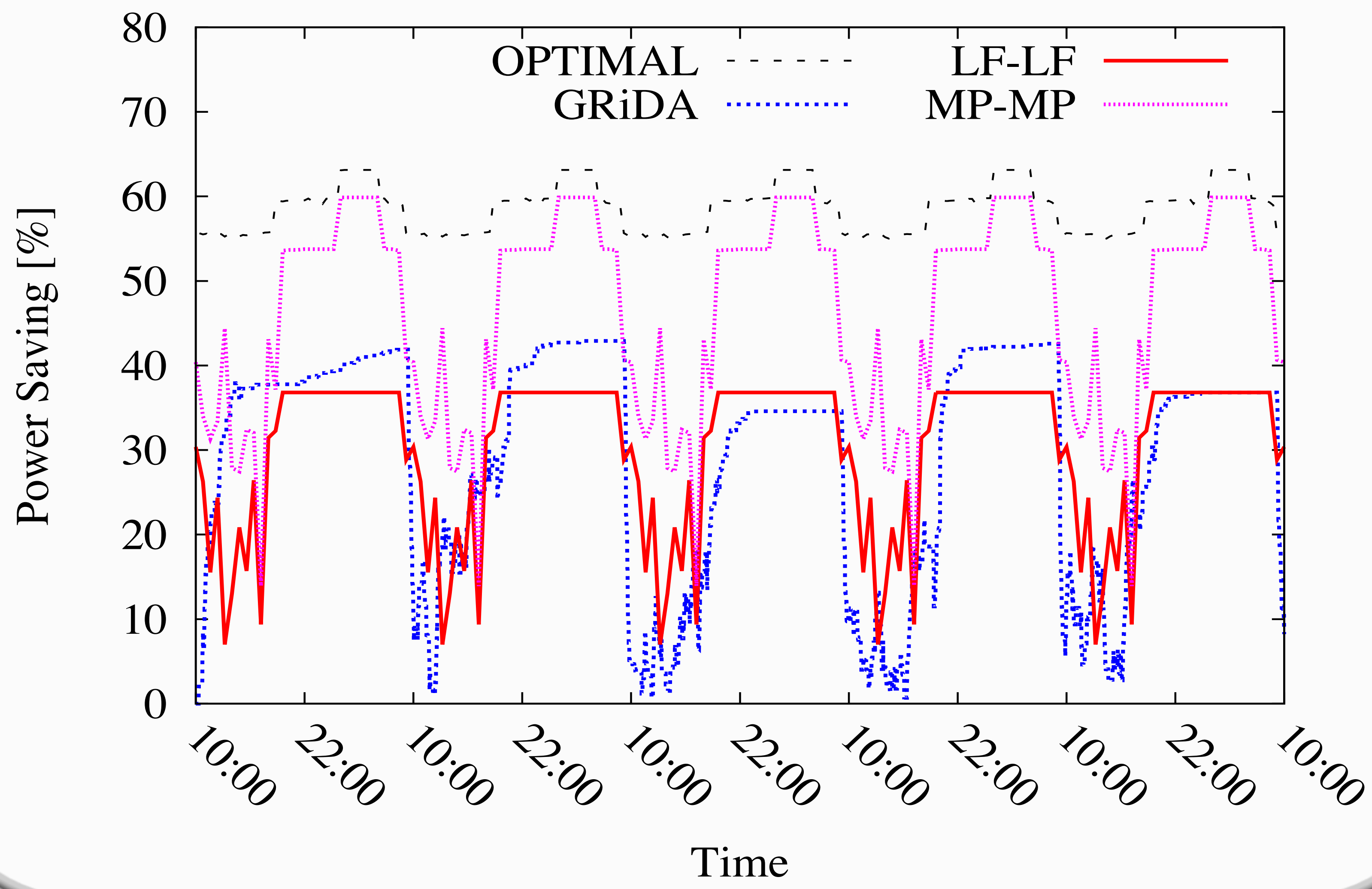


Convergence to a stable solution, in normal working state and in case of a node failure

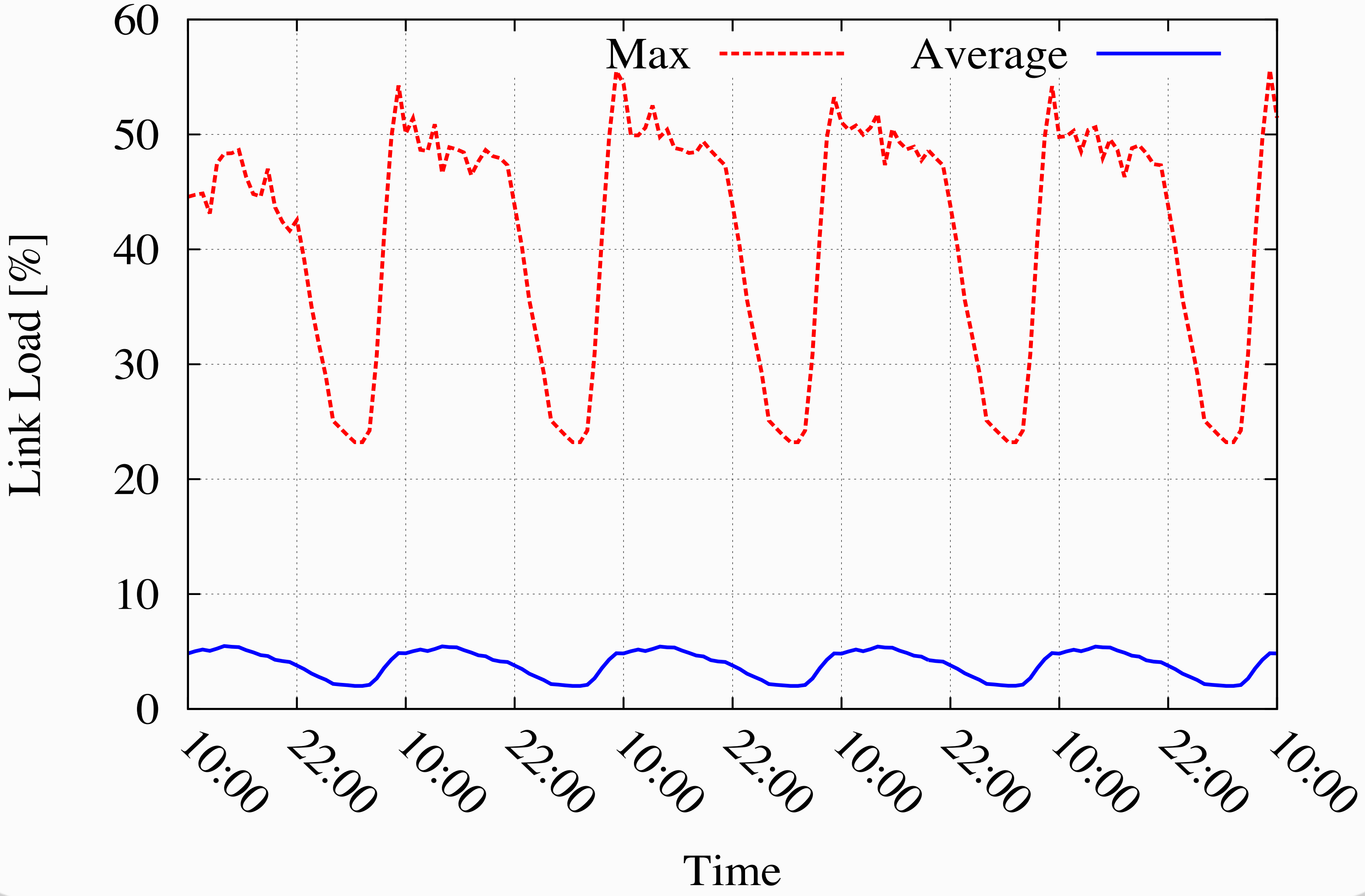
GRiDA: Results on Large Networks



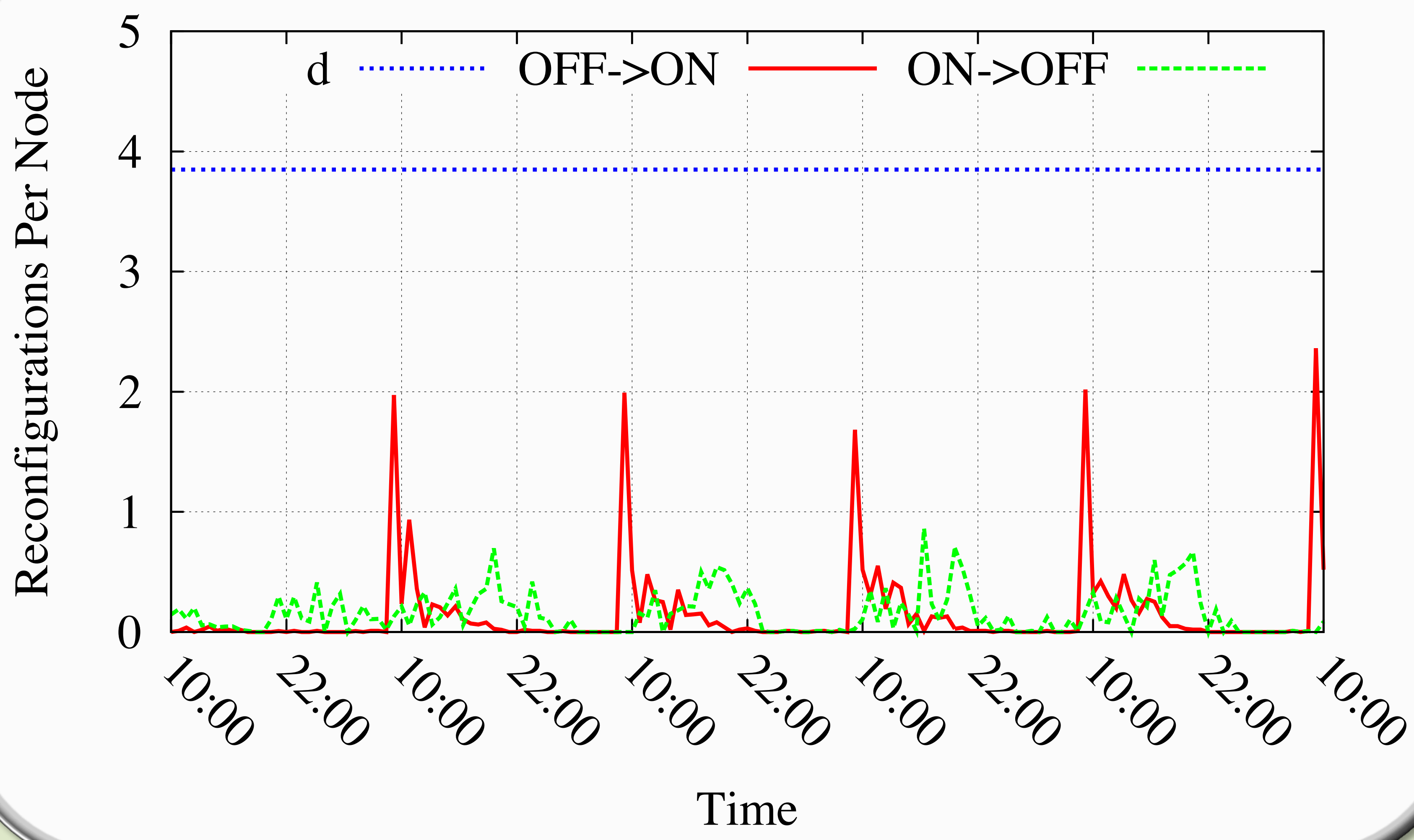
GRiDA: Results on Large Networks



GRiDA: Results on Large Networks



GRiDA: Results on Large Networks



GRiDA: Wrap-Up

- Distributed mechanism for resource consolidation in ISP networks
- No centralized knowledge needed (Traffic Matrix, routing paths, etc), no synchronization needed
- Really low complexity
- Reaction to fault events and to traffic changes
- Comparable performance with respect to centralized solutions in the literature

Future Work

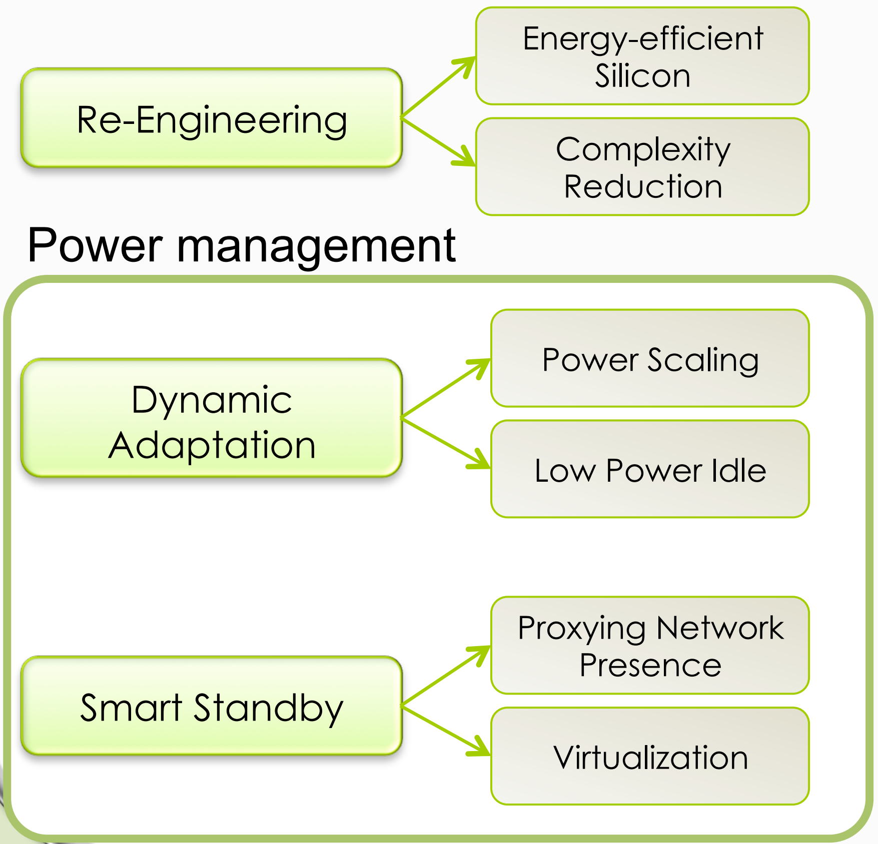
Future Works

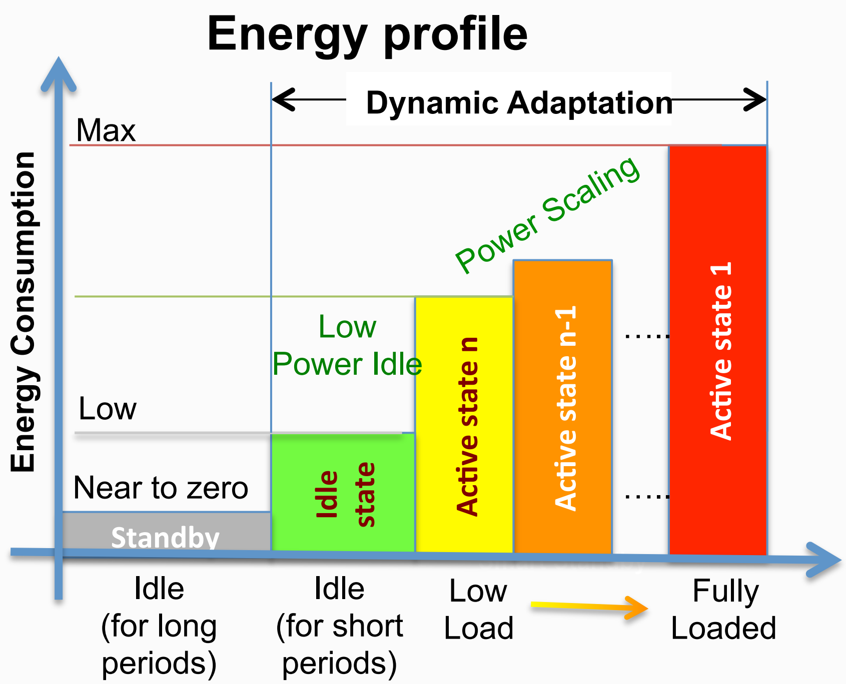
- Test-bed implementation: Based on Quagga and opaque LSA
- Integration with strategies exploiting sleep mode of other devices (e.g., linecards, switching fabrics, full nodes…)
- Theoretical proof of convergence
- Evaluate applicability into other network scenarios (e.g., wireless networks)

The ECOnet Project

The ECOnet Project

Possible Approaches to Green Networking:





The ECOnet Project (i)

low Energy COnsumption NETworks

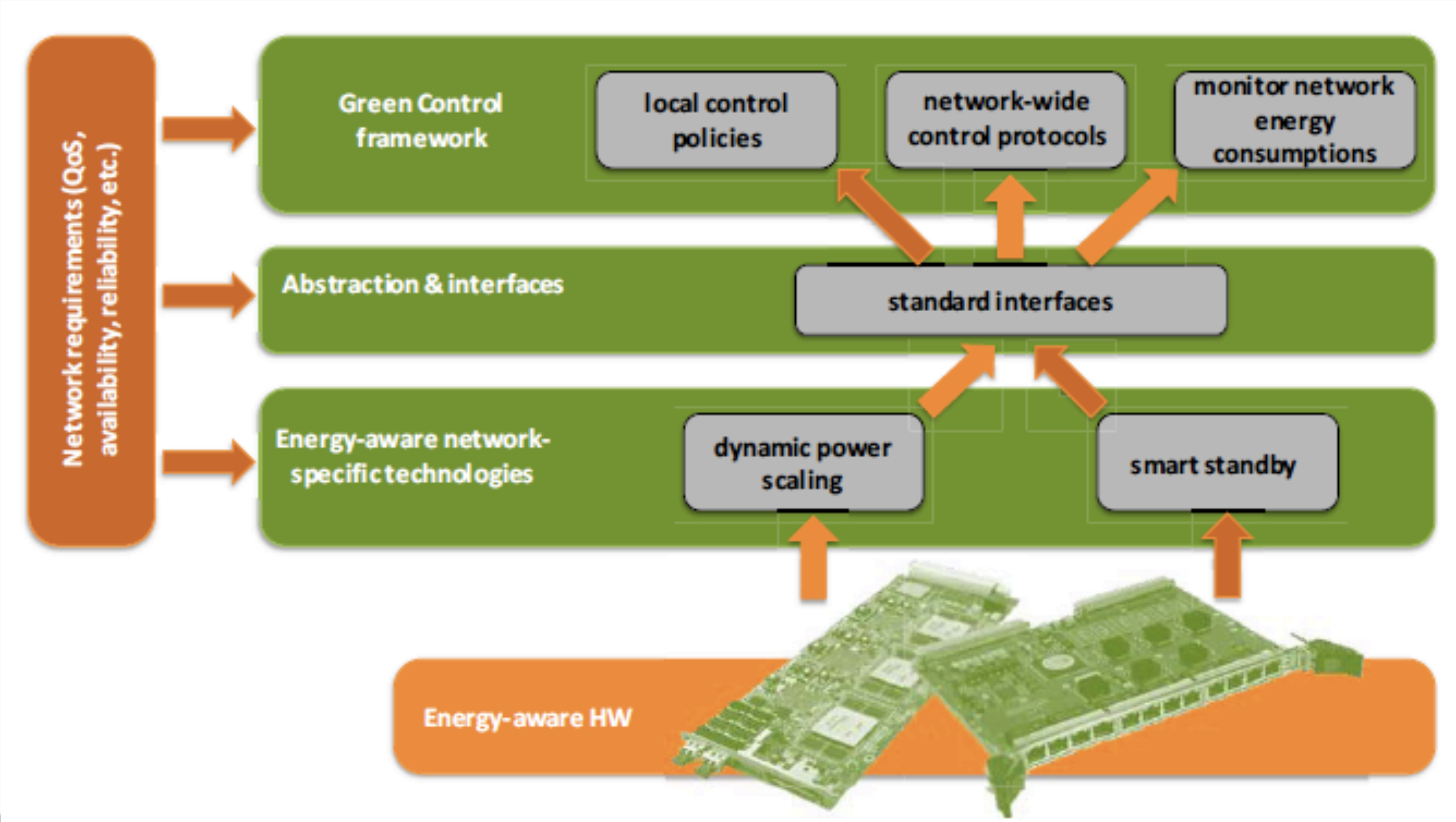
Project data at a glance	
Project Type	FP7 Integrated project
Project coordinator	Prof. Raffaele Bolla (CNIT, c/o University of Genoa)
Project duration	October 2010 – September 2013 (36 months)
Consortium	15 partners from 8 countries and 2 associated American Universities
Project budget	10.5 M€ (6.2 M€ from EU)
Resources	1168 PM (33 full-time persons for three years)
Website	http://www.econet-project.eu

The ECOnet Project (ii)

Participant organisation name	Short name	Country
Consorzio Nazionale Interuniversitario per le Telecomunicazioni – UdR at DIST University of Genoa (Coordinator)	CNIT	Italy
Mellanox Technologies	MLX	Israel
Alcatel Lucent	ALU	Italy
Lantiq	LQDE	Germany
Ericsson Telecomunicazioni S.p.A.	TEI	Italy
Telecom Italia	TELIT	Italy
Greek Research & Technology Network	GRNET	Greece
Research and Academic Computer Network	NASK	Poland
Dublin City University	DCU	Ireland
VTT Technical Research Centre	VTT	Finland
Warsaw University of Technology	WUT	Poland
NetVisor	NVR	Hungary
Ethernity	ETY	Israel
LightComm	LGT	Italy
InfoCom	INFO	Italy
Portland State University	PSU	USA
University of South Florida	USF	USA

The ECOnet Project (iii)

The project approach



Appendix

References

[1] A.P. BIANZINO, C. CHAUDET, F. LARROCA, D. ROSSI, J.-L. ROUGIER, "Energy-Aware Routing: a Reality Check", in *3rd International Workshop on Green Communications (GreenComm3)*, in conjunction with IEEE GLOBECOM 2010, Miami, USA, December 10 2010.

[2] L. CHIARAVIGLIO, M. MELLIA, and F. NERI, "ReducingPowerConsumptioninBackbone Networks," in Proceedings of the IEEE International Conference on Communications (ICC'09), (Dresden, Germany), June 2009.

[3] A.P. BIANZINO, C. CHAUDET, S. MORETTI, D. ROSSI, and J.-L. ROUGIER, "The Green-Game: Striking a Balance between QoS and Energy Saving", in *Proceedings of the 23rd International Teletraffic Congress (ITC 2011)*, San Francisco, CA, USA, September 6-8 2011.

[4] A.P. BIANZINO, L. CHIARAVIGLIO, M. MELLIA, "Distributed Algorithms for Green IP Networks", in *The 1st IEEE INFOCOM Workshop on Green Networking and Smart Grids*, Orlando, FL, USA, March 30 2012.

[5] A.P. BIANZINO, L. CHIARAVIGLIO, M. MELLIA, "GRiDA: a Green Distributed Algorithm for Backbone Networks", in *Proceedings of 2011 IEEE Online Green Communications Conference*, September 26-29 2011.

Further Reading

- D. AWDUCHE, A. CHIU, A. ELWALID, I. WIDJAJA, and X. XIAO, “Overview and Principles of Internet Traffic Engineering.” RFC 3272, May 2002.
- L. A. BARROSO and U. HOLZLE, “The Case for Energy-Proportional Computing,” IEEE Computer, vol. 40, pp. 33 – 37, Dec. 2007.
- A.P. BIANZINO, C. CHAUDET, D. ROSSI, J.-L. ROUGIER, "A Survey of Green Networking Research", in IEEE Communications Surveys & Tutorials, vol. 14, no. 2, pp. 3-20, First Quarter 2012.
- R. BOLLA, R. BRUSCHI, F. DAVOLI, and F. CUCCHIETTI, “Energy Efficiency in the Future Internet: A Survey of Existing Approaches and Trends in Energy-Aware Fixed Network Infrastructures,” IEEE Communication Surveys and Tutorials, 2011.
- M. GUPTA and S. SINGH, “Greening of the Internet,” in ACM SIGCOMM 2003, (Karlsrhue, Germany), August 2003.
- S. LEVY, “Going With the Flow: Google’s Secret Switch to the Next Wave of Networking”, Wired Enterprise, April 2012, http://www.wired.com/wiredenterprise/2012/04/going-with-the-flow-google/
- B. SANSO and H. MELLAH, “On Reliability, Performance and Internet Power Consumption,” in Proceedings of 7th International Workshop on Design of Reliable Communication Networks (DRCN 2009), (Washington, D.C., USA), Oct. 2009.

Thank you!

Further details at: bianzino@enst.fr
perso.telecom-paristech.fr/~bianzino

Overload Definition

- The network overload is defined as the fraction of traffic exceeding the load threshold Φ with respect to the total carried traffic:

$$\xi = \frac{\int_t \sum_{(i,j)\in E} \max(l_{ij}(t) - \Phi, 0)\,dt}{\int_t \sum_{s,d\in V} r^{sd}(t)\,dt}$$

- This is a relative indicator for the network congestion level, averaged over the simulation period, accounting for the *number* of load violations, their *entity*, and their *duration*.

The Energy Model

- We consider only link Energy consumption: ports + amplifiers [1]
- E_{nic}= 50W, E_a = 1kW, for c_{ref} = 10 Gbps
- l_a = 70km, distance between amplifiers

→ for link i: $E_i = int[c_i / c_{ref}](int[l_i/l_a]E_a + 2E_{nic})$

[1] L. Chiaraviglio, M. Mellia, and F. Neri, "Energy-aware backbone networks: a case study," in First Int. Workshop on Green Communications (GreenComm09), 2009.

Shared information

	GRiDA	DMP	DLF
Anomalous state	✔	✔	✔
Link power consumption	✗	✔	✗
Link load	✗	✗*	✔
Link power state	✔	✔	✔