Exploring the virtual infrastructure service concept in Grid'5000

20th ITC Specialist Seminar on Network Virtualization
18.05.2009, Hoi An, Vietnam

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

RESO

Pascale Vicat-Blanc Primet
pascale.primet@ens-lyon.fr

Fabienne Anhalt
fabienne.anhalt@ens-lyon.fr

Guilherme Koslovski
guilherme.koslovski@ens-lyon.fr

Laboratoire de l'Informatique du Parallélisme
École Normale Supérieure de Lyon - France

LIP

ECOLE NORMALE SUPERIEURE DE LYON

UNIVERSITÉ DE LYON

Context: Virtualized Infrastructures

- Virtualized Infrastructure is a concept emerging from Virtual Networks, Virtualization and Infrastructure as a Service waves
- Currently, there are many projects/users seeking a private and customized infrastructure utilization
- Eg. Computer science researchers need access to customized scalable testbeds where they can run reproducible experiments

Key issue:

network performance control in virtualized infrastructures

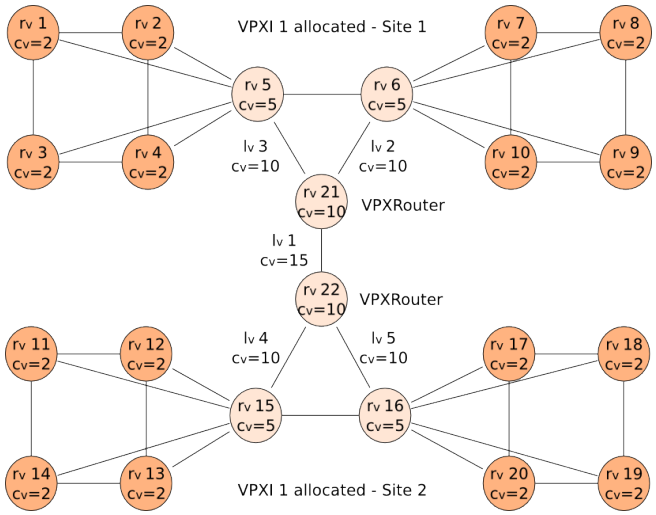
Key points of our proposal

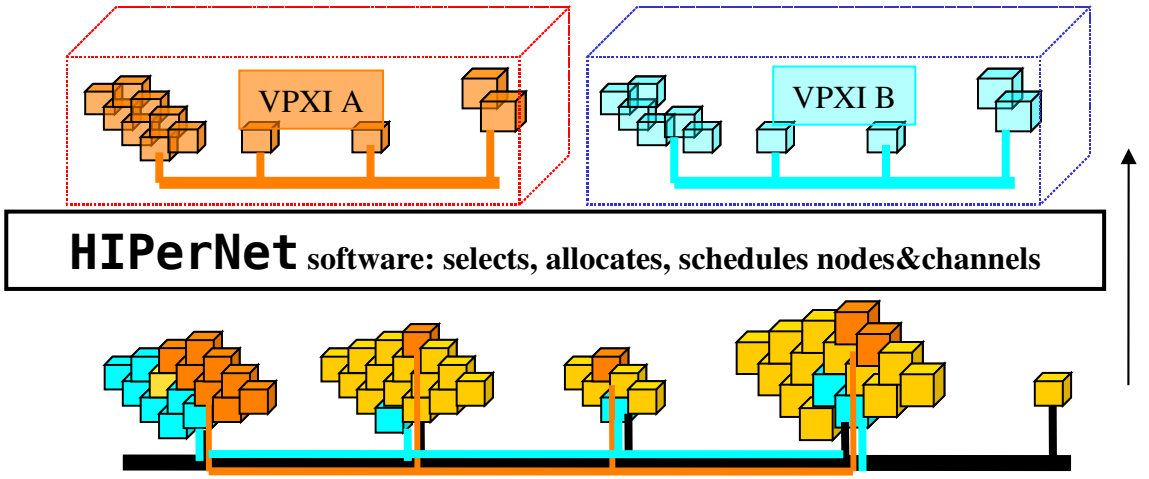
1) Unifying concept (VPN+Vgrid/Cloud)
2) Original description langage
3) Orchestration & configuration software
4) Customizable virtual router
5) Adaptation to reproducible experiments
6) Evaluation in a large scale testbed

V(PX)I : virtual infrastructure concept

A virtual (private execution) infrastructure : **V(PX)I** is defined in **VXDL** as:

- A list of **individual virtual resources** and groups
- A virtual **network topology** with weighted links (rate, latency…)
- The **executing timeline** of the application (for co-scheduling).

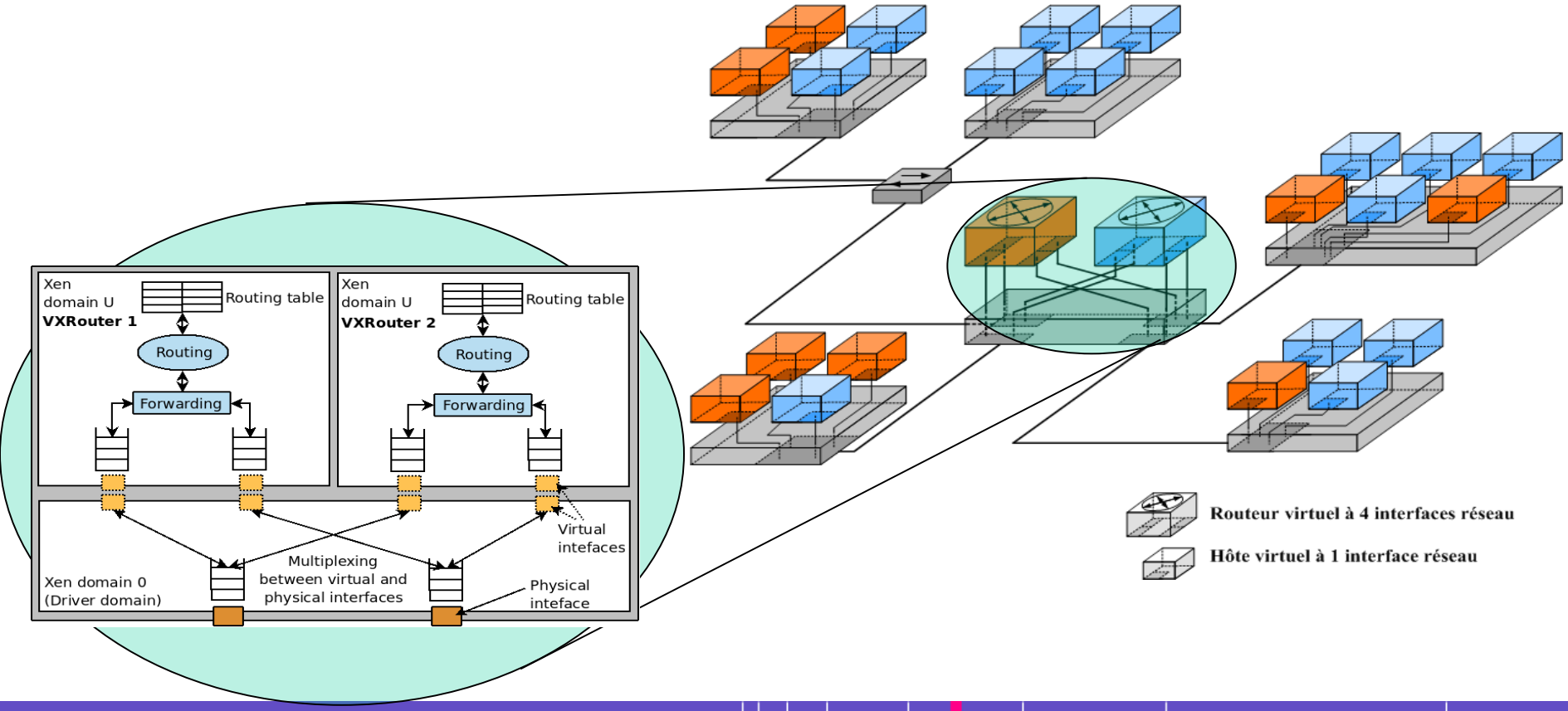




[Koslovski, Vicat-Blanc Primet. Grid05, GridNets08, ICNS09, CCGrid09]

Virtual Infrastructures building blocs

⇒Dynamic link provisioning
⇒Virtual software routers design (VXRouter)

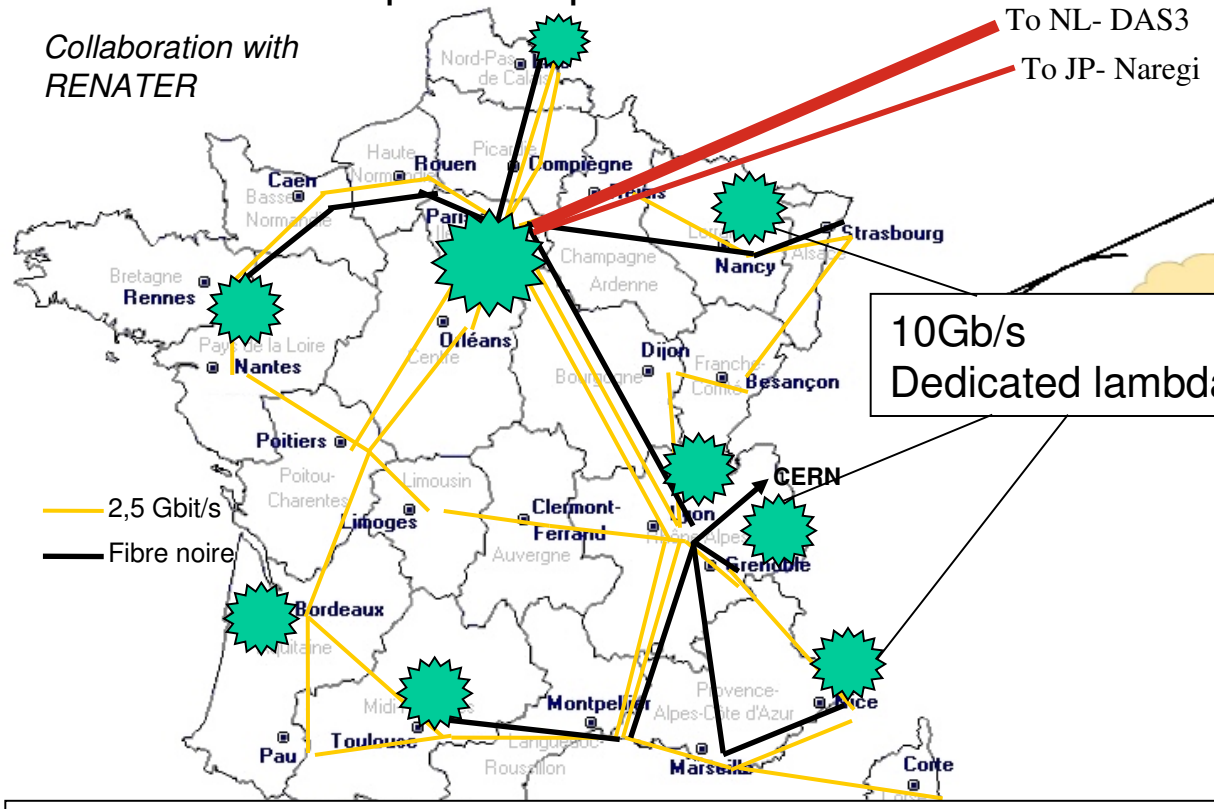


INRIA's experimental Cloud : Grid5000

⇒Grid5000 facility

- Resource mutualisation
- Reservation & reconfiguration features
- Over a private optical network

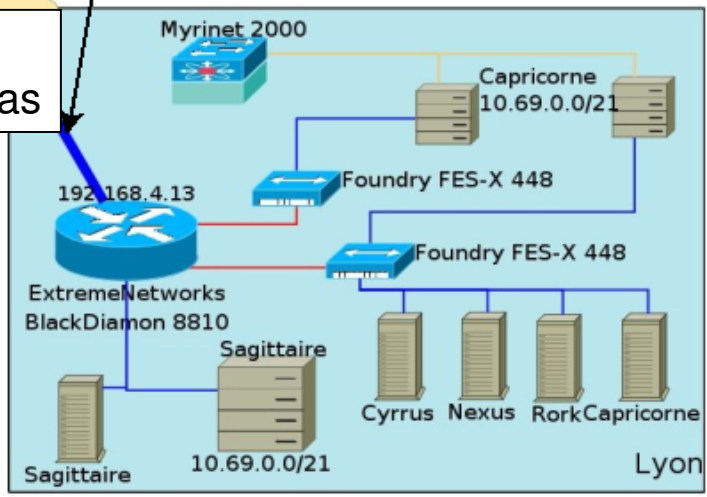
Collaboration with RENATER



Typical Grid5000 usage:

1. Connect to the platform on a site
2. Reserve a set of resources
3. Configure the resources (optional)
4. Run an experiment
5. Grab the results
6. Free the resources

10Gb/s
Dedicated lambdas



Grid5000@Lyon

[F. Cappello & al. Grid5000: a nation wide experimental grid testbed. in International Journal on High Performance Computing Applications, 2006.]

Grid'5000 as an instrument

4 main features:

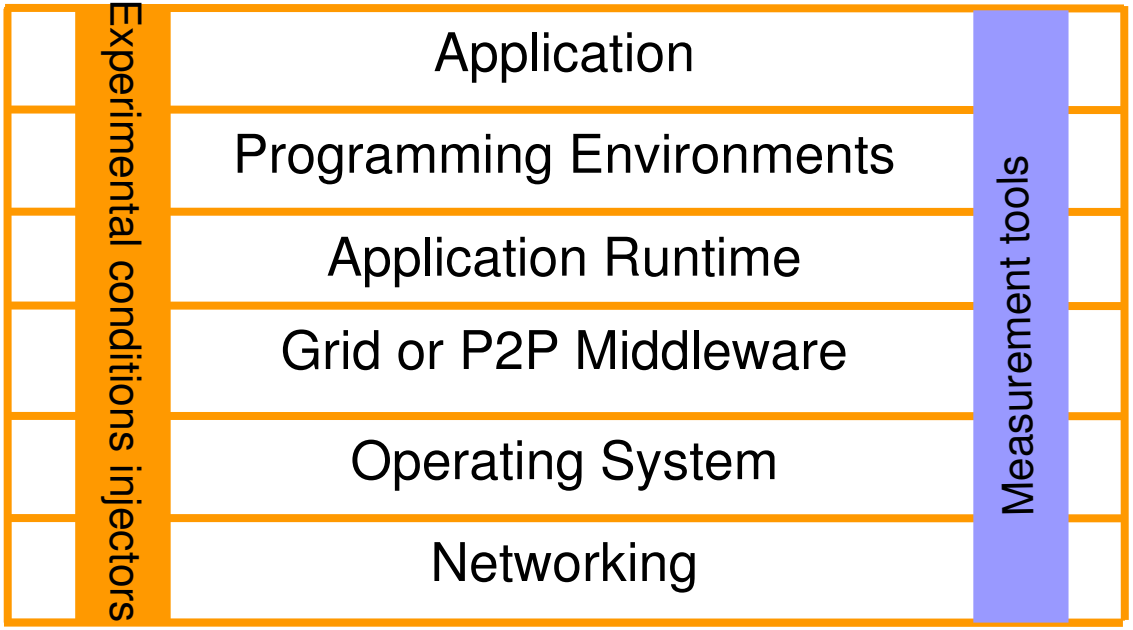
- A high security for Grid'5000 and the Internet, despite the deep reconfiguration feature
  --> Grid'5000 is confined: communications between sites are isolated from the Internet and Vice versa (level2 MPLS, Dedicated lambda).

- A software infrastructure allowing users to access Grid'5000 from any Grid'5000 site and have a simple view of the system
  --> A user has a single account on Grid'5000, Grid'5000 is seen as a cluster of clusters

- A reservation/scheduling tools allowing users to select nodes and schedule experiments
  --> A reservation engine + batch scheduler (1 per site) + OAR Grid (a co-reservation scheduling system)

- A user toolkit to reconfigure the nodes
  --> Software image deployment and node reconfiguration tool

Grid'5000 principle: A highly reconfigurable experimental platform



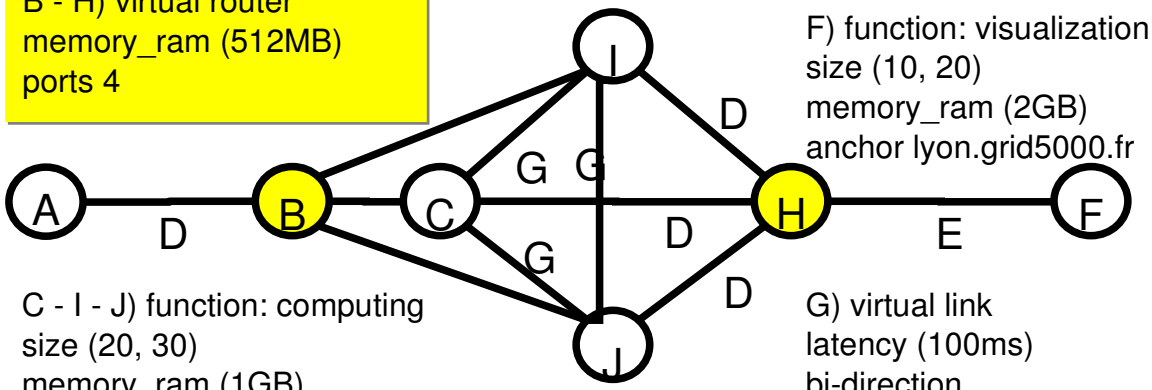
- Let us create, deploy and run their software stack, including the software to test and evalute the environment using measurement tools + experimental conditions injectors
- Grid5000 is a real powerfull & easy to use grid, used every day by 300 researchers

Virtual Infrastructure description

Grid5000 user's define their VPXI in **VXDL**

A) function: storage
size (10, 20)
hd_size (30GB, 60GB)

B - H) virtual router
memory_ram (512MB)
ports 4



E) virtual link
bandwidth (2Gbps, 3Gbps)
uni-direction
between [(H port 4,F)]

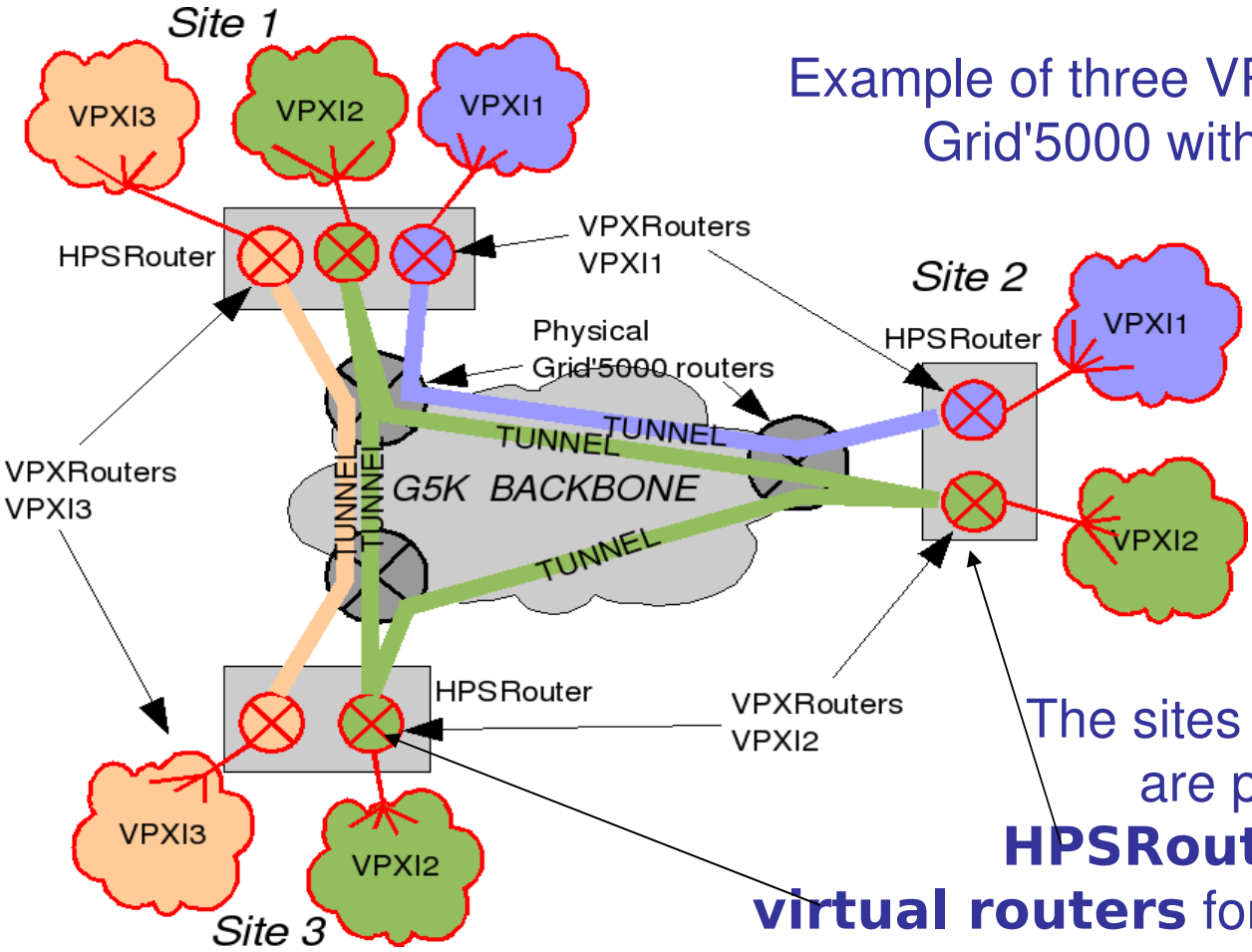
F) function: visualization
size (10, 20)
memory_ram (2GB)
anchor lyon.grid5000.fr

C - I - J) function: computing
size (20, 30)
memory_ram (1GB)

G) virtual link
latency (100ms)
bi-direction
between [(C,I), (I,J), (C,J)]

D) virtual link
bandwidth (1Gbps, 2Gbps)
uni-direction
between [(A, B port 1), (I, H port 1), (J, H port 2), (C, H port 3)]

HIPerNET's VPXI allocation on Grid'5000

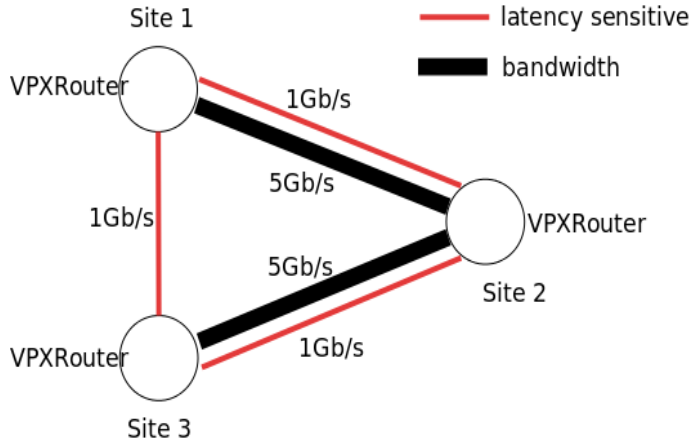


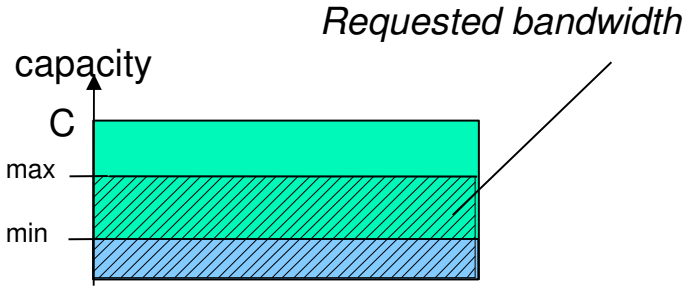
Example of three VPXIs allocated on Grid'5000 with HIPerNET

The sites of Grid'5000 are provided with **HPSRouters** hosting **virtual routers** for each VPXI.

Customized routing and bandwidth allocation VXcontrol

- HIPerNET exposes different bandwidth services:
  - Minimum guaranteed bandwidth:
    - guarantee the user a minimum bandwidth over its virtual links
    - *Cross_traffic < (C-min)*
  - Maximum allowed bandwidth:
    - Forbid the user's traffic to exceed a specified threshold
    - *User_traffic < (Max)*
  - Static bandwidth reservation:
    - guarantee a minimum bandwidth available and forbid to exceed a maximum
    - *Cross_traffic < (C-min) & User_traffic < (Max)*

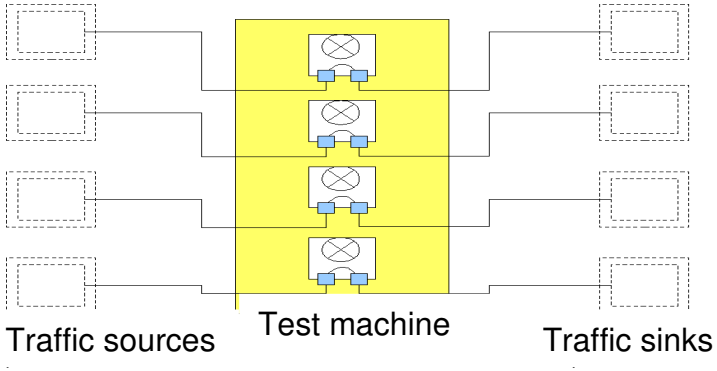




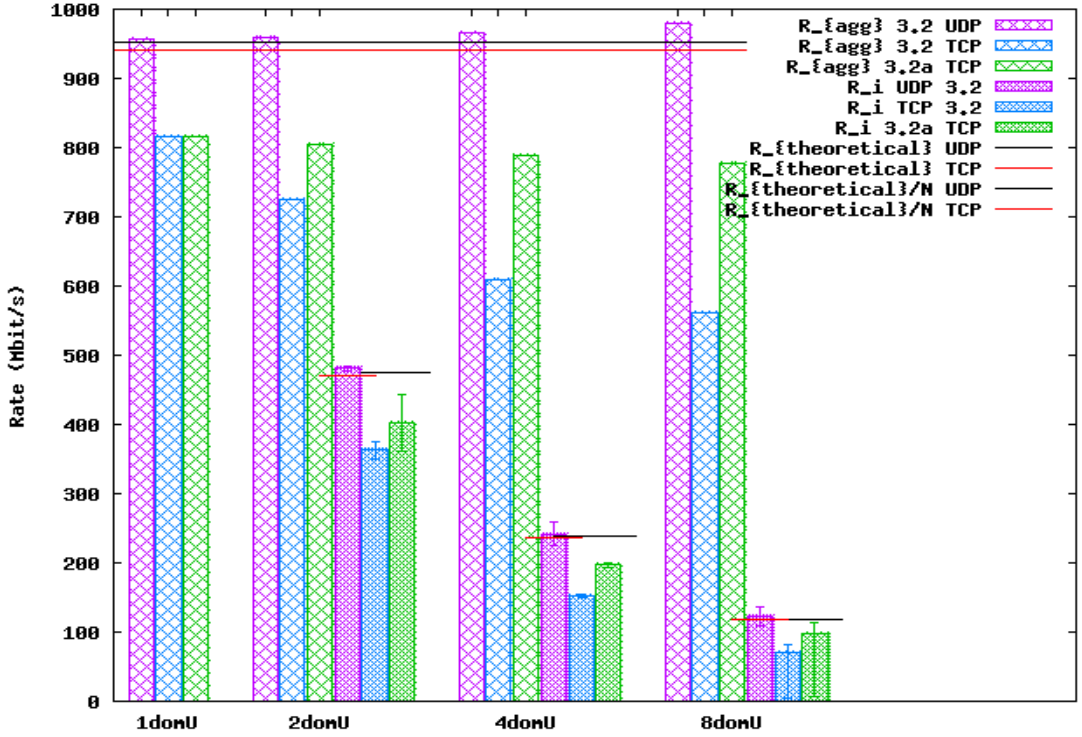
Experimental testplan and setup

- Goal 1: Virtual router performance
  - Xen 3.2 para-virtualization, Linux kernel 2.6.18
  - Netperf benchmark
- Goal 2: Rate control on virtual routers
  - Rate limitation with Linux trafic control tools
    - Token bucket filter (tbf)
    - Precise Software Pacer (PSPacer)
- Test platform
  - IBM eServers with 2 AMD Opteron CPUs, 2 GB of memory
  - 1 Gb/s NICs

Virtual network performance on virtual routers



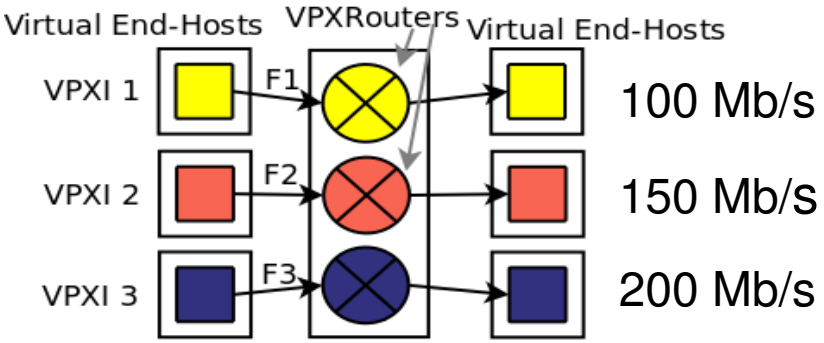
✘ Aggregate throughput low compared to theoretical throughput (941.49Mb/s)
✘ Rather fair bandwidth sharing



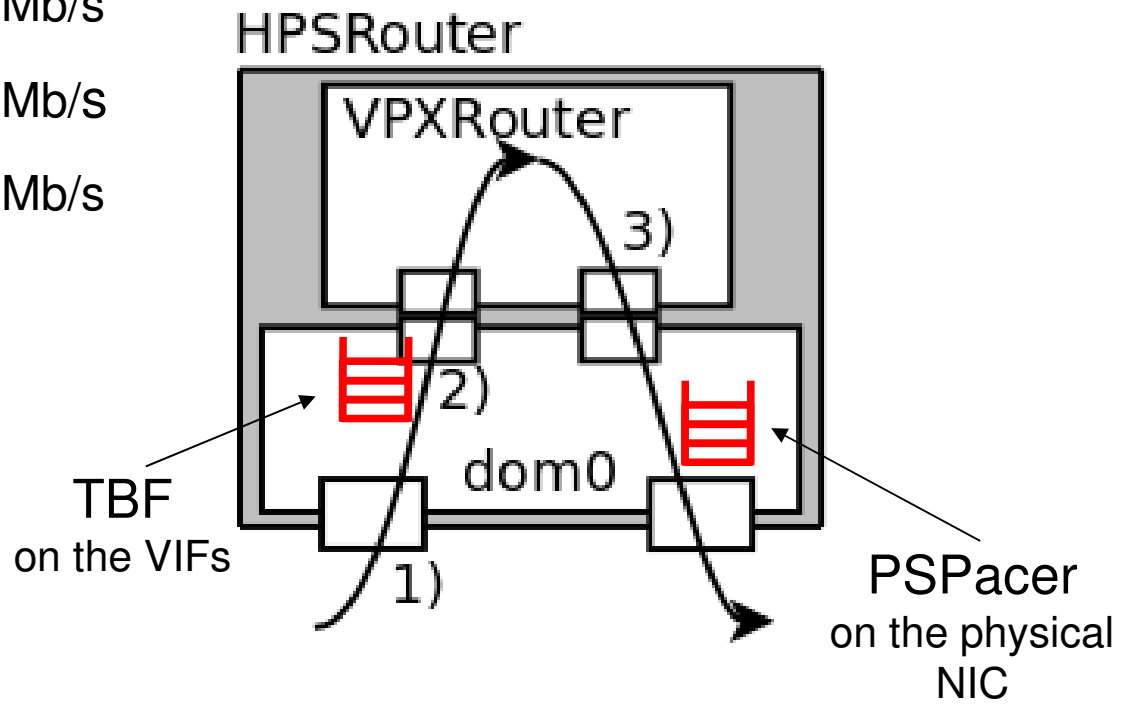
✓ Better throughput giving dom0 more CPU weight (3.2a)
✓ Improvement of scalability and predictability

Rate control (VXcontrol)

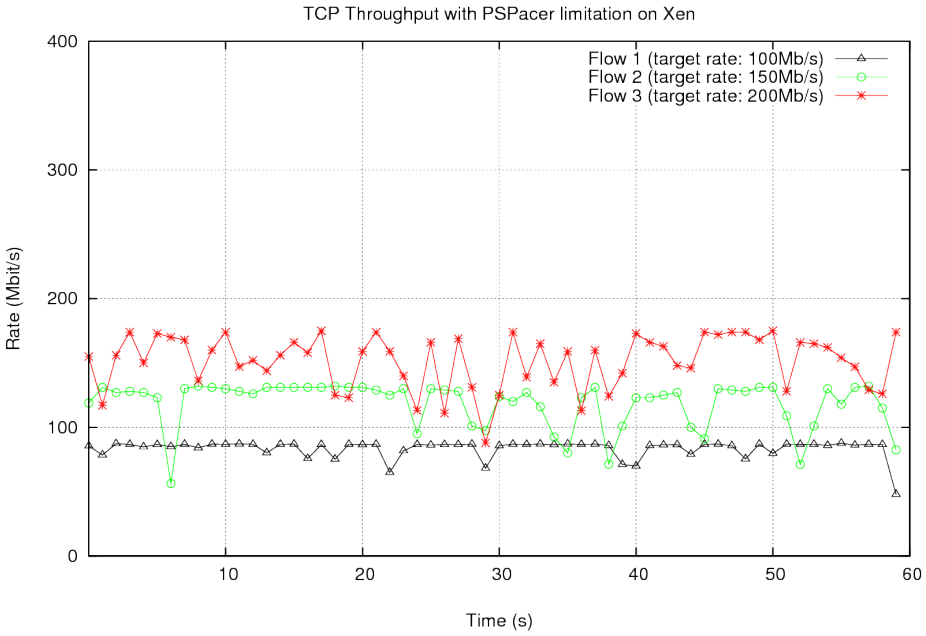
- Sending flows over three VPXIs with different static bandwidth reservations over the virtual routers



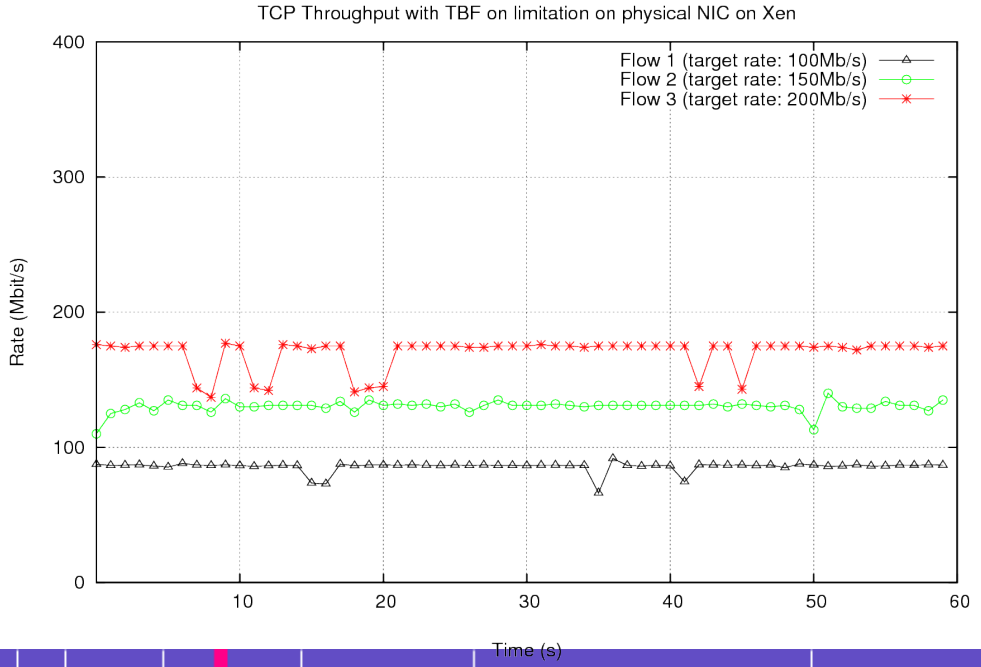
Comparison of two solutions of rate limiting inside the HPSRouter



Rate control: Results



PSPacer rate limitation on outgoing physical NIC (congestion factor = 0.9)



TBF rate limitation on the virtual NICs (congestion factor = 0.9)

Conclusion

Key points of our proposal:

1) Unifying concept:	**VPXI**
2) Original description langage:	**VXDL**
3) Orchestration & configuration software:	**HIPerNET**
4) Customizable virtual router:	**VXRouter**
5) Adaptation to reproducible experiments:	**VXcontrol**

Perspective: exploration of the full customization potential & isolation evaluation in a large scale with Grid5000 users

Thank you for your attention !

Any questions ?

pascale.primet@ens-lyon.fr - fabienne.anhalt@ens-lyon.fr - guilherme.koslovski@ens-lyon.fr