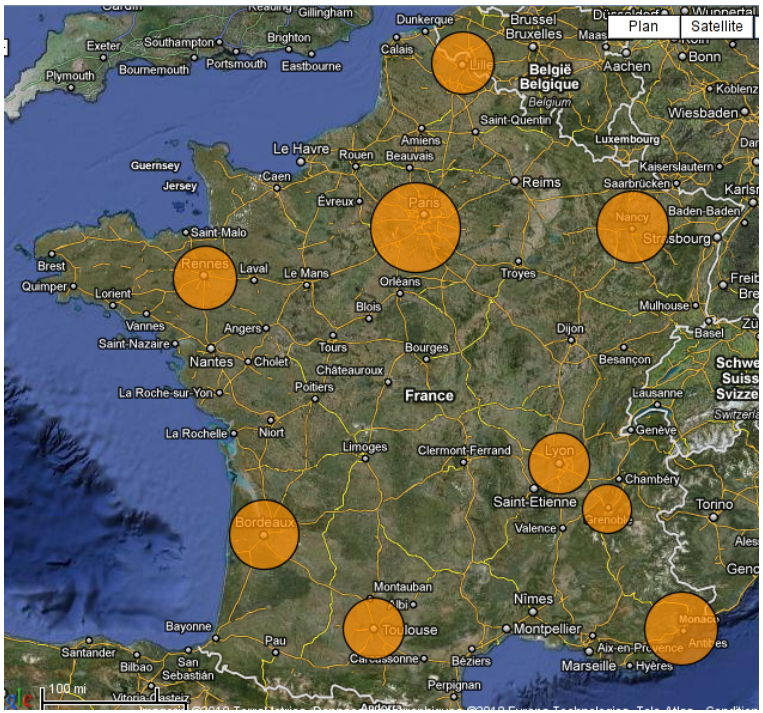
Virtualization on Grid'5000

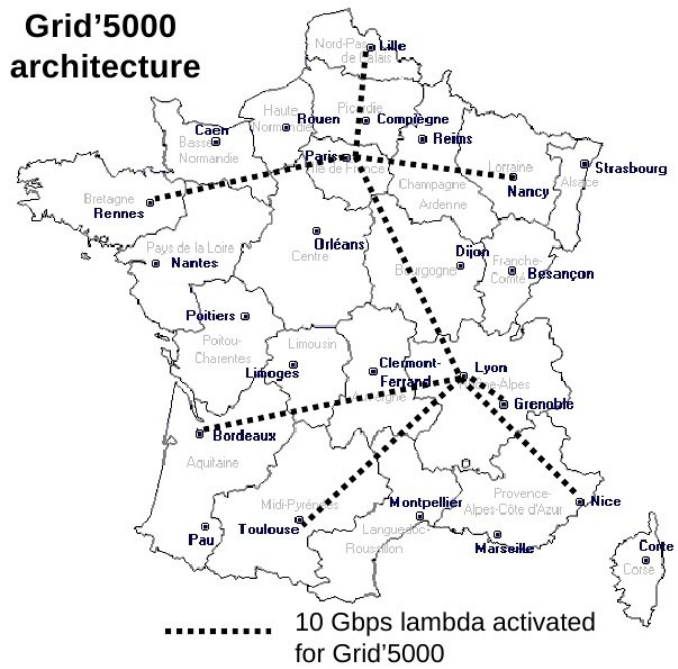
Lucas Nussbaum

LORIA / Nancy-Université

Grid’5000

- French infrastructure for research on large-scale distributed systems
- Funding : INRIA (ADT Aladdin-G5K), CNRS, regions and universities
- 9 sites, 1500 machines, 6000 cores
- Highly reconfigurable, controlable and monitorable
  Experiments on all layers : network, OS, middleware, applications





Virtualization on Grid'5000

- Used for most of the services required to run the platform
  32 dom0, 149 domU managed by the support staff
  Currently migrating to using Puppet

Virtualization on Grid'5000

- Used for most of the services required to run the platform
  32 dom0, 149 domU managed by the support staff
  Currently migrating to using Puppet
- Not used by default on the user nodes
  - Not the right thing for experiments
    Resource sharing ; overhead of the virtualization layer
  - Power to the users : can re-install nodes using KaDeploy
    RIaaS : Real Infrastructure as a Service ?

KaDeploy

`http://kadeploy3.gforge.inria.fr/`

- Fast and scalable deployment tool
  Re-install 300 machines in 10 minutes
- Underlying technologies : PXE, efficient broadcast, Grub
- Usually used to deploy Linux systems
- Can also deploy FreeBSD, Xen, . . .
  dd-based images, *chainloading* of disk partition to boot

KaDeploy

`http://kadeploy3.gforge.inria.fr/`

- Fast and scalable deployment tool
  Re-install 300 machines in 10 minutes
- Underlying technologies : PXE, efficient broadcast, Grub
- Usually used to deploy Linux systems
- Can also deploy FreeBSD, Xen, . . .
  dd-based images, *chainloading* of disk partition to boot

Similar to a cloud/virtualization provisioning tool
but works on real machines !

Choose your virtualization solution and deploy it on Grid'5000

Supporting Virtualization Experiments

- System images
- Network

Supporting Virtualization Experiments : Images

- Pre-built images for Xen
  - Debian Etch – Xen 3.0.3
  - Debian Lenny – Xen 3.2

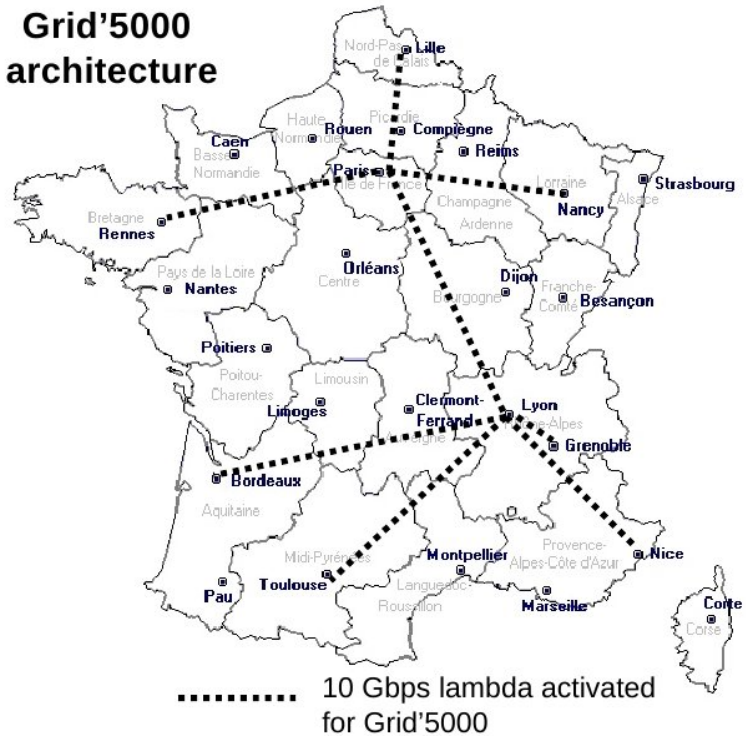
Supporting Virtualization Experiments : Images

- Pre-built images for Xen
  - Debian Etch – Xen 3.0.3
  - Debian Lenny – Xen 3.2
- Work in progress :
  - Newer images
  - More documentation of the image creation process
    - Help users create their own images
    - Required for Xen-based clouds

Supporting Virtualization Experiments : Images

- Pre-built images for Xen
  - Debian Etch – Xen 3.0.3
  - Debian Lenny – Xen 3.2
- Work in progress :
  - Newer images
  - More documentation of the image creation process
    - Help users create their own images
    - Required for Xen-based clouds
- Open question : broadcast of domU environments to nodes
  Chain-based (Kastafior) ? P2P-based (BitTorrent) ? GridFTP-like ?

Supporting Virtualization Experiments : Network



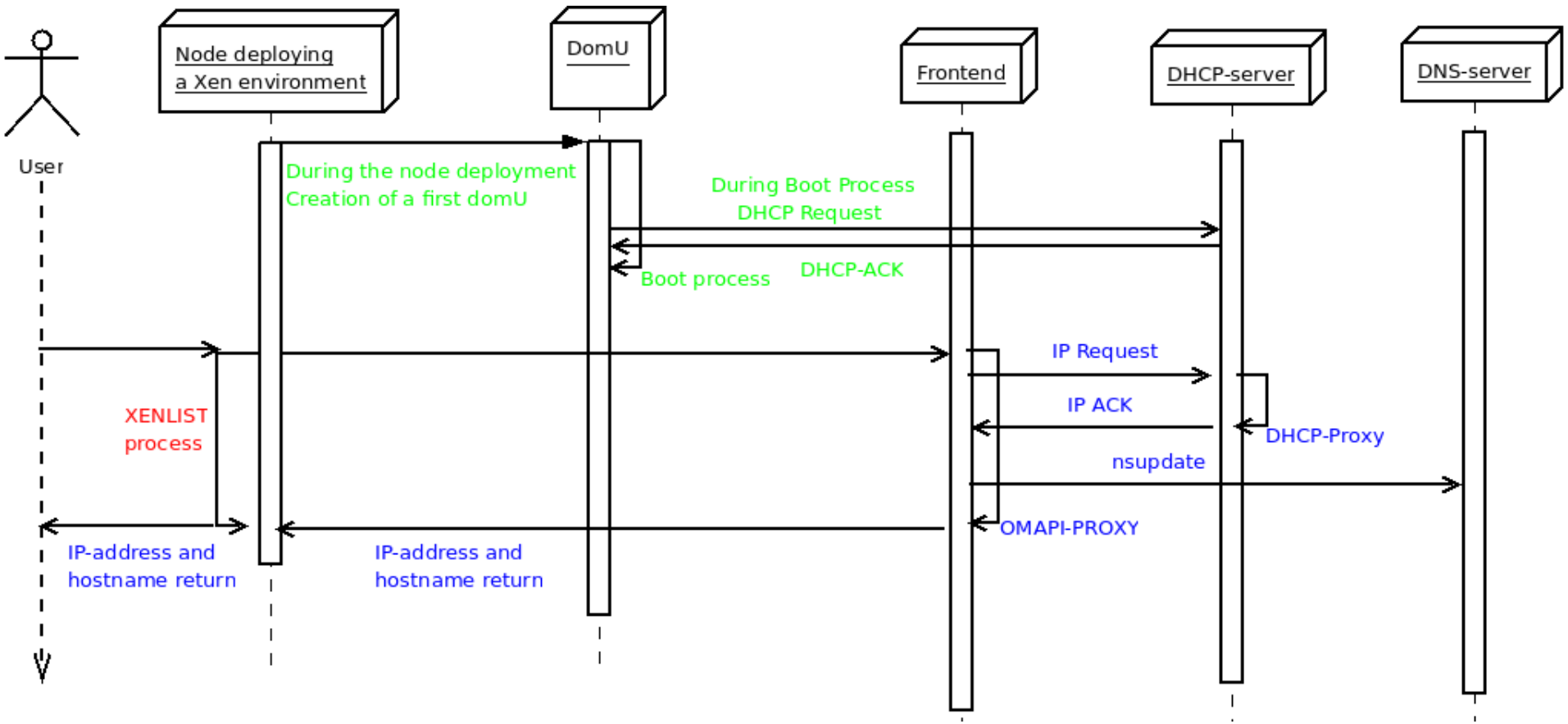
- Grid'5000 uses routing (level 3) between sites
- Routable (inside Grid'5000) IP addresses for virtual machines
  - /14 (262142 addresses) per site
    Example : Lille – 10.136.0.0 - 10.139.255.255

Reservation of virtual adddresses

- Several users might use virtualization simultaneously
  $\Rightarrow$ Need to avoid conflicts
- Need reservation scheme for both IP and MAC addresses
- Several co-existing solutions (on different /16 IP ranges) :
  - *Fingers crossing* : do not use any reservation tool
    User hopes that nobody else is using the addresses
  - In-house address reservation solution

Grid'5000 Address reservation solution

- Work by Cyril Constantin then Philippe Robert
- Uses DHCP to allocate IP addresses
- `xenlist` script on dom0 :
  - Gives allocated IP
  - Configures DNS



Address reservation solution : shortcomings

- Random allocation (the DHCP way)
- Cannot restrict to IP ranges (required by clouds)
- MAC of domU depends on the node
  Corner cases during experiments with migration
- Xen-specific
- Doesn’t integrate with cloud solutions

Address reservation solution : shortcomings

- Random allocation (the DHCP way)
- Cannot restrict to IP ranges (required by clouds)
- MAC of domU depends on the node
  Corner cases during experiments with migration
- Xen-specific
- Doesn’t integrate with cloud solutions

Ongoing discussions to improve the situation

Ideas & feedback welcomed !

Network virtualization

Goal : ensure isolation between experiments

- Run your own DHCP, malicious code, etc

Network virtualization

Goal : ensure isolation between experiments

- Run your own DHCP, malicious code, etc

Inside a Grid'5000 site : KaVLAN

- Dynamic configuration of VLANs on switches and routers

Network virtualization

Goal : ensure isolation between experiments

- Run your own DHCP, malicious code, etc

Inside a Grid'5000 site : KaVLAN

- Dynamic configuration of VLANs on switches and routers

Between Grid'5000 sites : KaVLAN + QinQ

- IEEE 802.1QinQ to propagate VLANs over the inter-site network

Network virtualization

Goal : ensure isolation between experiments

- Run your own DHCP, malicious code, etc

Inside a Grid'5000 site : KaVLAN

- Dynamic configuration of VLANs on switches and routers

Between Grid'5000 sites : KaVLAN + QinQ

- IEEE 802.1QinQ to propagate VLANs over the inter-site network

Both are eagerly waiting for beta-testers !

Other possible uses of virtualization on Grid'5000

Other possible uses of virtualization on Grid'5000

(Due to lack of resources, most of those are not being pursued currently)

Increase the number of available nodes

- Virtualization not suitable for general use
- But during exp. preparation, no need to run on real hardware
- Could help prepare experiments on a large number of nodes without being too disruptive

Increase the number of available nodes

- Virtualization not suitable for general use
- But during exp. preparation, no need to run on real hardware
- Could help prepare experiments on a large number of nodes without being too disruptive

But :

- Open research question (projects : Hipcal/Hipernet, Entropy)
- Still high availability of resources, not really needed

Transform Grid’5000 into Cloud infrastructure

- Cloud’5000 : virtualization everywhere
- "Grid" no longer hype, "Cloud" is
- Easier to *sell* Grid’5000 that way

Transform Grid’5000 into Cloud infrastructure

- Cloud’5000 : virtualization everywhere
- "Grid" no longer hype, "Cloud" is
- Easier to *sell* Grid’5000 that way

But :

- Not the Right Thing to do
  Access to real hardware is invaluable
  .. And required for virtualization research !
- Some (old) nodes do not support hardware virtualization

Provide a Cloud API on Grid'5000

Provide a Cloud API on Grid'5000

- OK, but which one ?

Provide a Cloud API on Grid'5000

- OK, but which one ?
- We already have the Grid'5000 API : `https://api.grid5000.fr/`

Provide a Cloud API on Grid'5000

- OK, but which one ?
- We already have the Grid'5000 API : `https://api.grid5000.fr/`
- Different kind of reservations
  Job scheduling, walltime specified at beginning of job

Provide a Cloud API on Grid'5000

- OK, but which one ?
- We already have the Grid'5000 API :
  `https://api.grid5000.fr/`
- Different kind of reservations
  Job scheduling, walltime specified at beginning of job
- Possible solution : build cloud API over Grid'5000 API

Provide virtual machines for specific experiments

- Some experiments require infrastructure
- Large scale campaigns of best-effort jobs
- Idea : provide virtual machines to users

Provide virtual machines for specific experiments

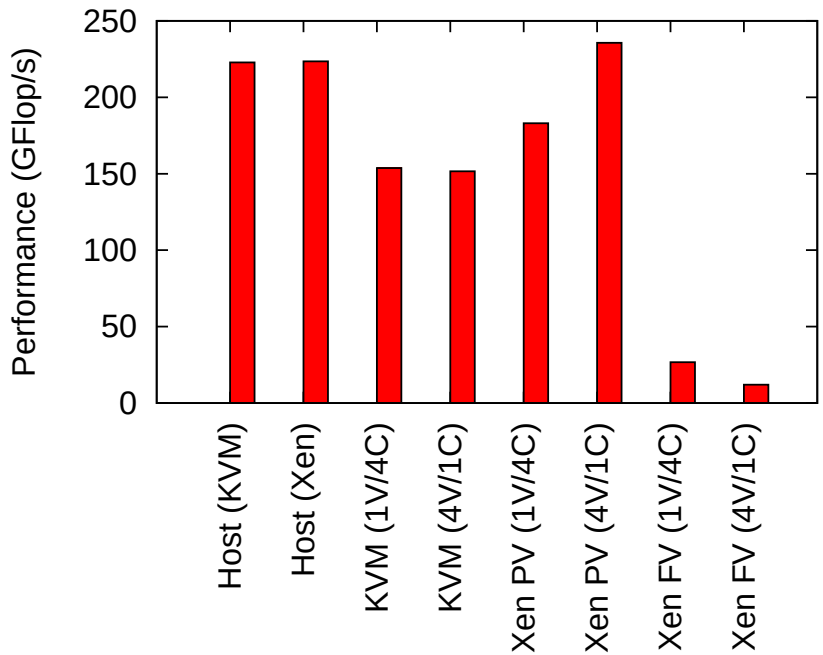
- Some experiments require infrastructure
- Large scale campaigns of best-effort jobs
- Idea : provide virtual machines to users

But :

- Not clear if most needs can be satisfied using VM
  Specific needs : fast storage, etc
- So far allocated manually when needed
  Not clear if infrastructure to automate that necessary

Conclusion

- Grid'5000 : great platform for research on virtualization
- Provides direct access to hardware
- Possible to work, at a large scale, on :
  - Comparative studies between virtualisation solutions
  - Various hypervisor improvements (inter-site migrations)
  - Developing and deploying clouds on hundreds of nodes



Discussion

- Which platform are you using for your virtualization studies ?
- Are you using Grid'5000 ? Did it work for your virtualization work ?
- Not using it ? Why ? Would it be useful for you ? What's missing ?