VM5k and DVMS on Grid'5000

Deploying and Managing Thousands of Virtual Machines on Hundreds of Nodes Distributed Geographically

Jonathan Pastor[1] Laurent Pouilloux[2]

[1]Hemera Phd
ASCOLA - Mines Nantes / Inria

[2]Hemera Engineer
Inria / ENS Lyon

18-06-2014 / Grid'5000 School

Context

Cloud computing usage is becoming very popular.

- Ever-increasing demand $\Rightarrow$ ever-increasing infrastructure size.
- Problems: scalability, reliability, network overhead, energy but also security and juridiction

Proposition: [Greenberg2009]

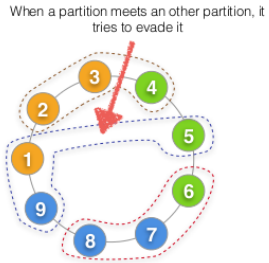
Concept of microdatacenters geographically spread

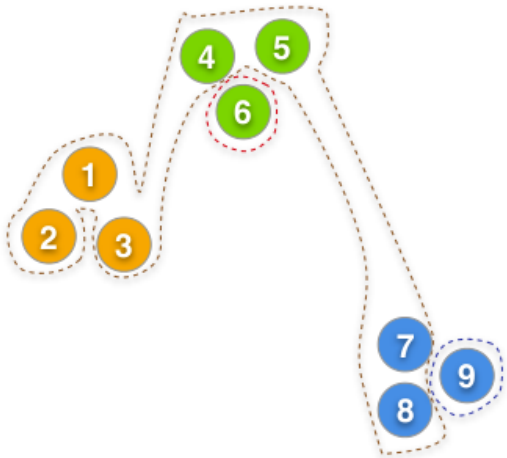
Discovery project

http://beyondtheclouds.github.io/

Decentralise the production of computing ressources

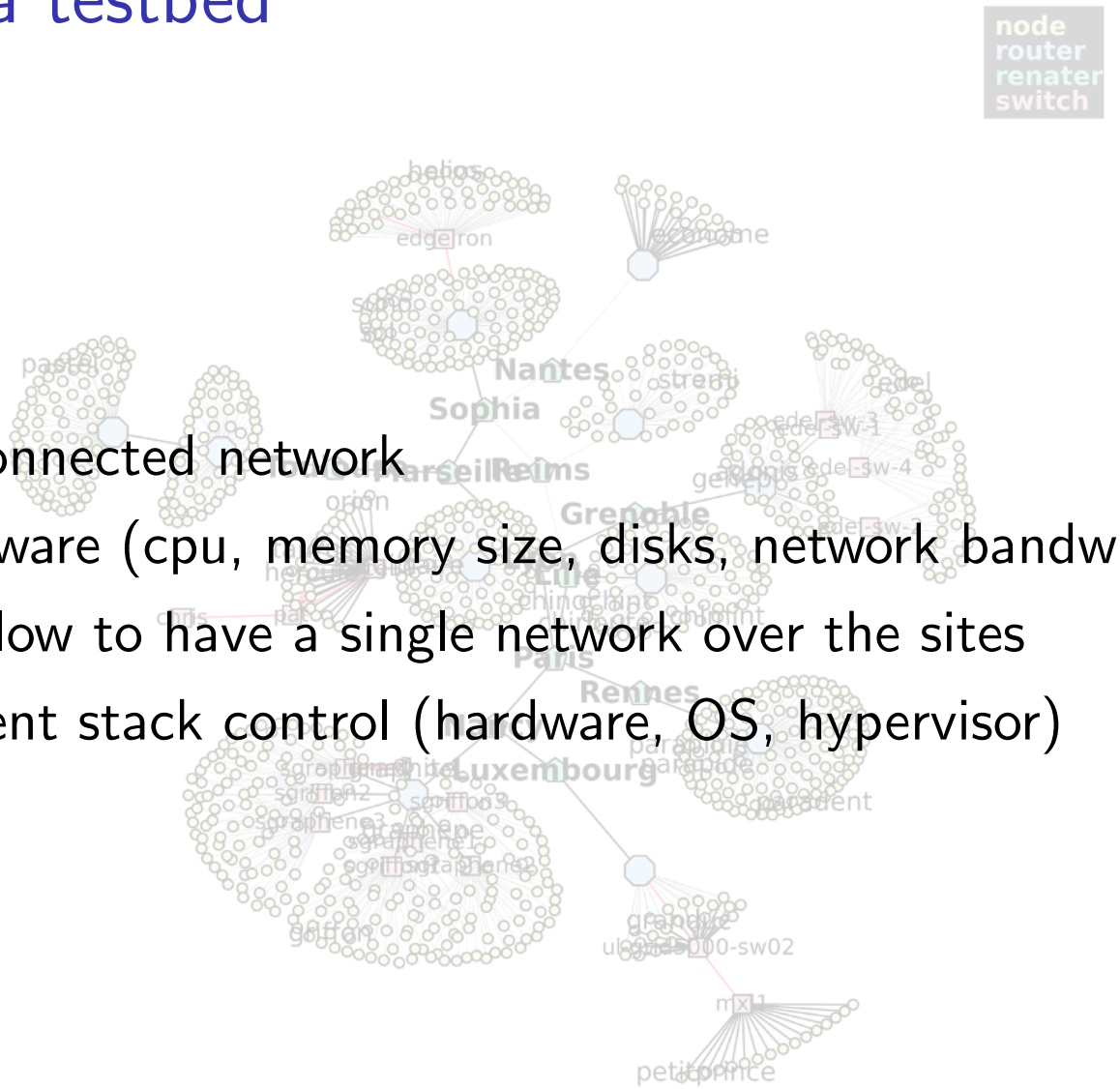
- Chord topology
- taking into account network distance





Evaluating DVMS with Vivaldi (coordinates system)

Grid'5000 as a testbed



- 10Gb interconnected network
- various hardware (cpu, memory size, disks, network bandwidth)
- KaVLAN: allow to have a single network over the sites
- full experiment stack control (hardware, OS, hypervisor)

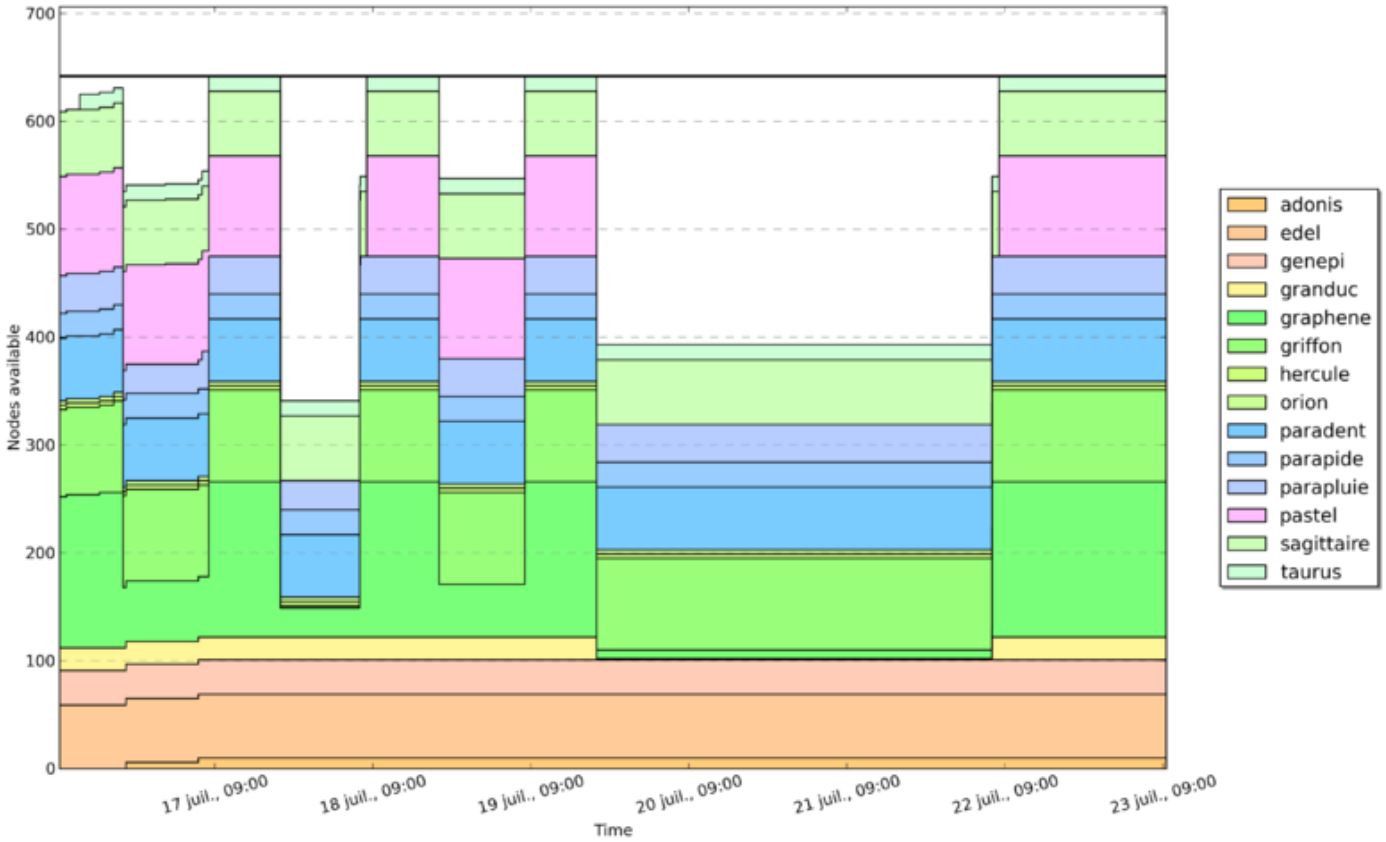
Experimental Workflow

1. reserve many nodes on different sites, with a global-KaVLAN
2. deploy thousants of Virtual Machines
3. initiate stress process on them
4. install DVMS
5. use vivaldi to compute hosts distances
6. generate random stress on the virtual machines
7. live experiment visualization
8. collect results

(F)ind yo(U)r (N)ode on g5(K)

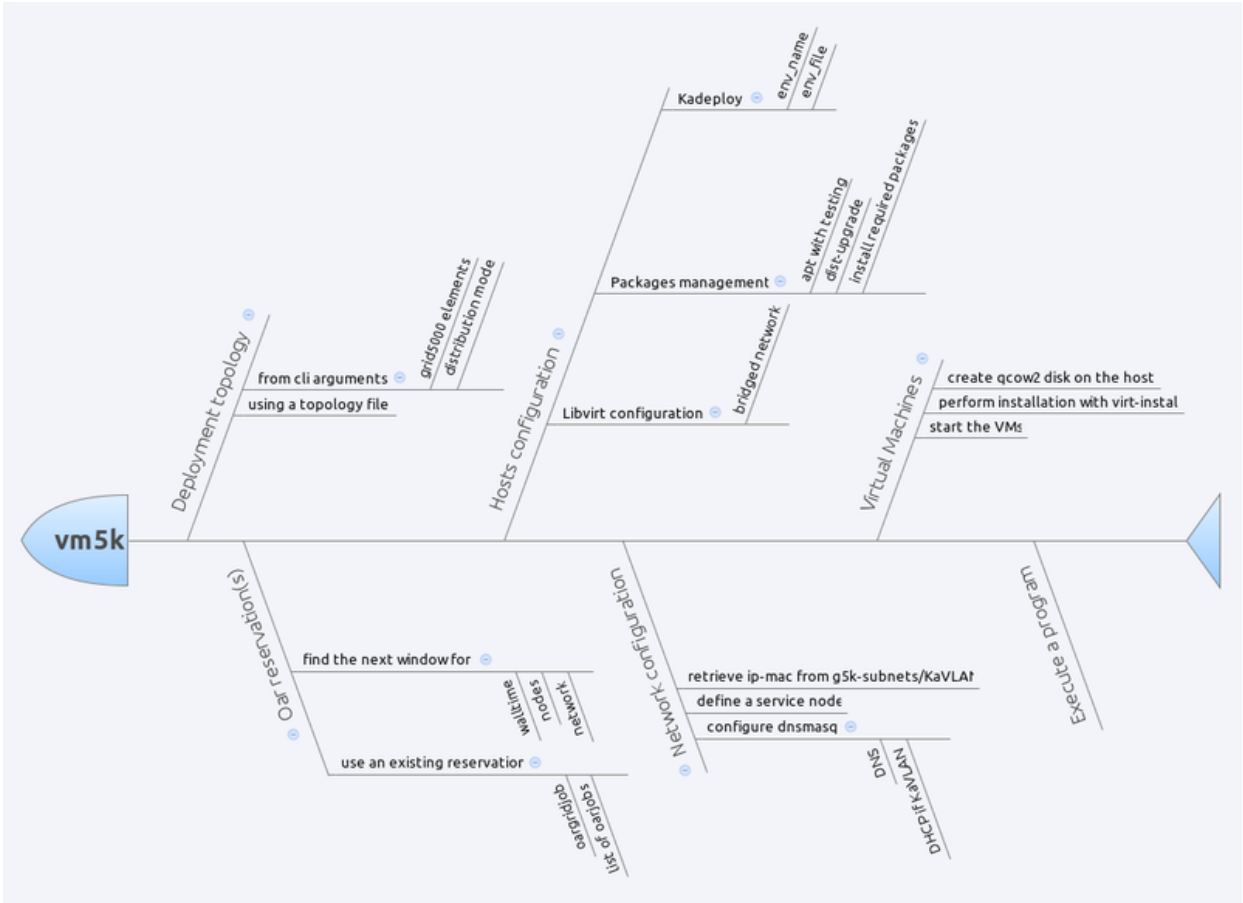
A advanced resources discovery tool for multisite reservation

```
funk -m free -r grid5000:200 -o "-t deploy" -w 12:00:00
  -b helios,sagittaire,nantes,reims,graphite -k -c
```



Automatic Virtual Machines deployment

Moving FLauncher (D. Balouek and F. Quesnel) to vm5k



Tested successfully up to 5 000 VMs on 300 nodes.

Stress initialization

On all running Virtual Machines and using execo

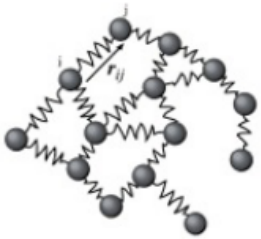
- upload the `memtouch` binary
- start a memtouch process
- set it's cpu usage to 1% using `cpulimit`

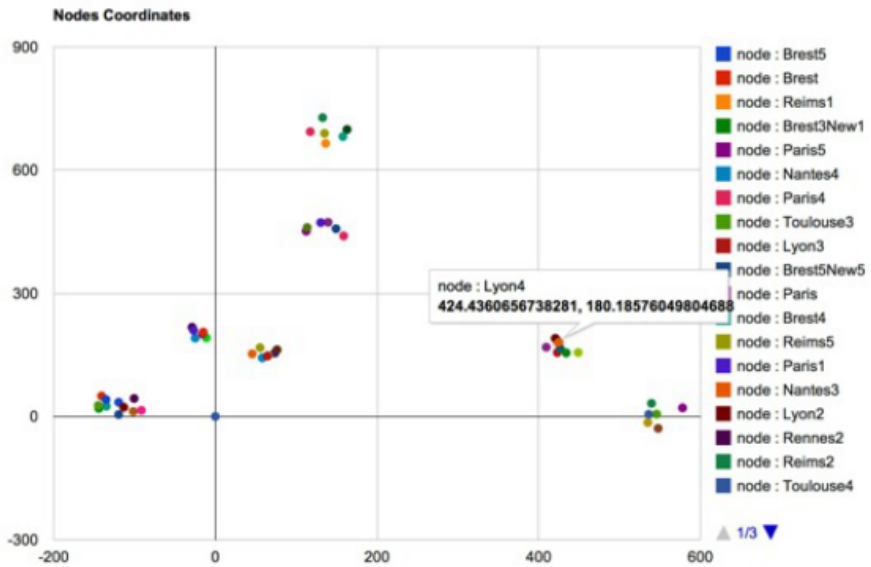
All VMs are ready to be stressed.

Vivaldi

Based on "Spring systems" [Dabek2004]

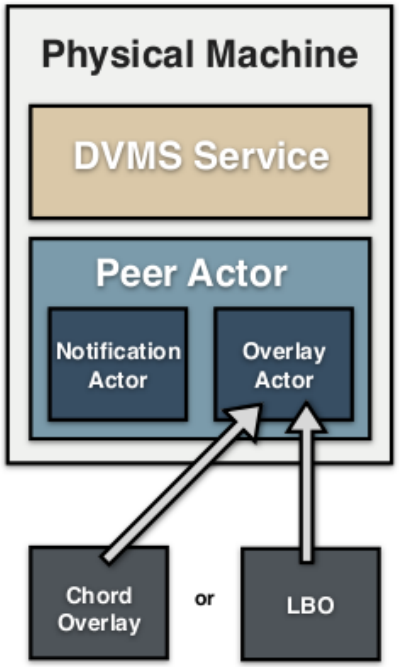
- Contacts are exchanged randomly between nodes
- Latency is measured ⇒ spring tension





DVMS

- DVMS uses the PeerActor model, where services leverage overlay networks.
- Development of a Locality Based Overlay (LBO).
- It uses the Vivaldi coordinate system.
- Through the use of the PeerActor architecture, DVMS collaborate with close neighbours to perform migrations.



Live visualization

`http://localhost:9000`

- infrastructure state (VM position and load)
- distance map from Vivaldi (used by DVMS to determine where to migrate the VMs)
- bonus: node live power usage

Load events generation

We can tune:

- distribution of event load value
- events frequency

```
<?xml version="1.0" encoding="UTF-8"?>
<events>
  <event type="update_cpu_load" time="0.6322858606354911" target="vm-103" location="10.27.216.103" value="30"/>
  <event type="update_cpu_load" time="0.6501133097844356" target="vm-24" location="10.27.216.24" value="50"/>
  <event type="update_cpu_load" time="0.77156495657007" target="vm-59" location="10.27.216.59" value="100"/>
  <event type="update_cpu_load" time="1.001639188886804" target="vm-123" location="10.27.216.123" value="80"/>
  <event type="update_cpu_load" time="1.7390251934205876" target="vm-19" location="10.27.216.19" value="100"/>
  <event type="update_cpu_load" time="2.161792297887189" target="vm-44" location="10.27.216.44" value="80"/>
  <event type="update_cpu_load" time="3.1211590695284515" target="vm-41" location="10.27.216.41" value="100"/>
  <event type="update_cpu_load" time="3.525741904535117" target="vm-61" location="10.27.216.61" value="40"/>
 ...
</events>
```

Use an execo script to set the value of the load using `cpulimit`

Results Analysis

- Vivaldi map
- Migration statistics
- Bonus: fine-grained power consumption for some nodes

Conclusion

Large scale validation of DVMS taking into account node distance

- -almost- fully automatized experiment
- wide usage of Grid'5000 features (API, Kadeploy, KaVLAN, Kwapi)
- real execution up to 5000 Virtual Machines
- demo available on Challenge_DVMS_Live_-_School_2014

Jonathan Pastor, Marin Bertier, Frédéric Desprez, Adrien Lèbre, Flavien Quesnel, and Cédric Tedeschi. *Locality-aware Cooperation for VM Scheduling in Distributed Clouds*. In Euro-Par 2014, Porto, Portugal, August 2014.

Thank your for your attention. Questions ?