Cluster Deployment and Dynamic Partitionning with Kadeploy and KaVLAN

Emmanuel Jeanvoine, Nicolas Niclausse, Lucas Nussbaum and David Margery

Grid’5000

- Testbed for experiment-driven research on parallel, large-scale or distributed computing and networking (Cloud, High Performance Computing, P2P, Grid)
- 1700 machines (7400 CPU cores) in 26 clusters and 11 sites
- Technologies to support diverse experiments:
  - CPU from one to twelve cores
  - High Performance networks: Infiniband & Myrinet
  - Dedicated 10 Gb inter-site network (RENA TER)
- Key feature: reconfigurable by users
  - Installation of other operating systems on nodes: experiments on any level of the software stack
  - Network isolation: allows the deployment of intrusive or security-sensitive protocols and applications

KaVLAN – network isolation

- Reconfigures switches for the duration of a user experiment to achieve complete level 2 isolation:
  - Avoid network pollution (broadcast, unsolicited connections)
  - Enable users to start their own DHCP servers
  - Experiment on ethernet-based protocols
  - Interconnect nodes with another testbed without compromising the security of Grid’5000
- Relies on 802.1q (VLANs)
- Compatible with many network equipments
  - Can use SNMP, SSH or telnet to connect to switches
  - Supports Cisco, HP, 3Com, Extreme Networks and Brocade
- Controlled with a command-line client or a REST API
- Integrated with the OAR resource manager and Kadeploy
- Several types of VLANs are provided on Grid’5000:

<table>
<thead>
<tr>
<th>Type</th>
<th>Ethernet isolation</th>
<th>IP isolation</th>
<th>Multi-site</th>
<th># of VLAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>3 per site</td>
</tr>
<tr>
<td>routed</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3+3 per site</td>
</tr>
<tr>
<td>global</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1 per site</td>
</tr>
</tbody>
</table>

  global VLAN relies on 802.1ad (Q-in-Q) on the backbone

Kadeploy – scalable cluster deployment tool

- Built on top of PXE, DHCP, TFTP
- Scalable, efficient and reliable:
  - Chain-based and BitTorrent environment broadcast
  - 255 nodes deployed in 7 minutes (including 5 minutes for the two mandatory reboots)
- Flexible: each step can be re-defined to specialize the deployment process for specific needs
- Support of a broad range of systems (Linux, Xen, *BSD, etc.)
- Two user interfaces:
  - Command-line, synchronous interface
  - Asynchronous interface for higher level API (e.g REST API)
- Can be used by system administrators to manage a cluster (kadeploy, kareboot, kaconsole, kapower, kastat), or by end users to deploy their own execution environment (≈ Hardware-as-a-Service Cloud infrastructure)
- Also used outside Grid’5000
  - Debian and RPM packages
- Supported by INRIA (ADT Kadeploy 2011-2013)

Example uses of Kadeploy and KaVLAN

- Experiment on security-sensitive applications (malware) without compromising the security of the rest of the testbed
- Deploy a cluster distribution (OSCAR, Rocks) inside a VLAN, with its own infrastructure
- Run Open-MX (implementation of the MX protocol over Ethernet) over several Grid’5000 sites (tutorial available)
- Create a development environment for Kadeploy, with its own DHCP and TFTP servers

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

https://www.grid5000.fr/mediawiki/index.php/KaVLAN